Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative Multi-Cohort Transcriptomic Analysis Reveals Necroptosis–Immune Crosstalk and a Predictive Gene Signature in Breast Cancer

Shehzad Khalil

shehzadkhalil911@gmail.com

The University of Agriculture, Peshawar

Research Article

Keywords: breast cancer, necroptosis, gene signature, machine learning, immune infiltration, SHAP

Posted Date: October 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10836415/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrative Multi-Cohort Transcriptomic Analysis Reveals Necroptosis–Immune Crosstalk and a Predictive Gene Signature in Breast Cancer

Shehzad Khalil[1]*

[1] Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar, Khyber Pakhtunkhwa, Pakistan

Shehzad Khalil: shehzadkhalil911@gmail.com

* **Corresponding author:** Shehzad Khalil (shehzadkhalil911@gmail.com)

Abstract

Background: A type of non-apoptotic, regulated, inflammatory cell death called necroptosis has emerged as a growing player in the regulation of breast cancer progression and anti-tumor immunity, but its transcriptomic signature has not been systematically defined.

Methods: Three discovery cohorts (GSE42568, GSE15852 and GSE29431, totaling 273) were merged with an independent validation cohort (GSE70947, n=296). Differentially expressed genes (DEGs) between tumor and normal tissues were determined after quantile normalization and ComBat batch correction using limma and intersected with 159 curated necroptosis regulators. Penalized logistic regression, support vector machines, and random forests were used to develop predictive models and rank the most important features, using SHAP values. The infiltration of immune cells was quantified using xCell, and the immune-crosstalk between necroptosis and immune cells were analysed by Spearman correlation.

Results: We found 840 DEGs (304 up and 536 down), including seven genes associated with necroptosis (PLA2G4A, STAT5A, PYGL, TLR4, MAPK10, and IL33 all down, while STAT1 was up). A 31-gene consensus signature, which consisted of the most discriminatory genes identified by machine learning, discriminated well in held-out test samples (SVM AUC=0.968, AUPRC=0.987, Brier score=0.049) and showed strong predictive power in the independent validation cohort (AUC=0.842). The most important genes for model prediction were chosen: PCK1, RBP4, TNMD, ITIH5, and PLIN1, through the SHAP analysis. Large-scale immune profiling showed significant enrichment of epithelial cells (0.381 vs 0.231, $p<0.001$), a depletion of adipocytes (0.216 vs 0.442,

p=8.0e-27) and stromal contraction (0.201 vs 0.440, p=8.2e-25) and 148 significant necroptosis-immune associations. There were comparable gene-immune correlation patterns between the discovery cohort and the validation cohort (r=0.757).

Summary and Perspectives: Coordinated inhibition of multiple components in the necroptosis pathway is the general mechanism by which necroptosis is suppressed in breast cancer, and is strongly linked to remodeling of the immune microenvironment. The resulting 31-gene signature is able to discriminate between tumor and normal tissue with high accuracy (SVM AUC 0.968) and is generalizable to an independent cohort (AUC 0.842).

Keywords: breast cancer, necroptosis, gene signature, machine learning, immune infiltration, SHAP

1. Introduction

Worldwide, breast cancer is the most common cancer in women, with approximately 2.3 million new cases diagnosed each year [1]. While good progress has been made in diagnosis in some patients and targeted and endocrine treatment have been expanded for others, a significant proportion of patients, especially those with triple-negative or aggressive luminal disease, are still treatment resistant and are likely to suffer relapse [2, 3]. However, there is now a large amount of evidence that the clinical behaviour of breast cancer is not solely determined by tumour-intrinsic genetic alterations; the contributions of the surrounding microenvironment, such as immune cell infiltration and inflammatory signalling also play an important role in the progression of tumour cells and their response to treatment [4, 5]. Molecular programs that link the regulation of cell death to this microenvironment are therefore central to the need to improve the prognosis and therapies.

One such program is a genetically encoded, caspase-independent pathway for necrotic cell death that uses receptor-interacting protein kinases (RIPK1 and RIPK3) and a pseudokinase called MLKL [6]. Necroptosis is a process that disrupts the plasma membrane and releases damage-associated molecular patterns (DAMPs) and pro-inflammatory cytokines that directly activate innate and adaptive immune system, whereas apoptosis is generally immunologically silent [7]. This dual identity gives necroptosis a context-dependent role in cancer. In some tumor settings it behaves as a tumor-suppressive mechanism, priming anti-tumor immunity through immunogenic cell death [8]. In others, sustained necroptotic signaling fuels a chronic inflammatory microenvironment that instead favors tumor growth and immune evasion [9, 10].

Within breast cancer specifically, individual components of the necroptosis pathway have been linked to disease biology, although mostly in isolation. RIPK3 expression is reported to be reduced in breast tumors relative to adjacent normal tissue, with lower expression associated with poorer prognosis [11]. STAT1, a principal mediator of interferon signaling, can license necroptotic death under specific conditions [12], whereas STAT5A has instead been associated with luminal differentiation and more favorable outcomes [13]. TLR4 signaling within the tumor microenvironment has been reported to exert opposing effects depending on context, at times promoting anti-tumor immune responses and at other times fostering an inflammatory, pro-tumorigenic phenotype [14].

What is still missing from this literature is a systems-level, transcriptome-wide view of how the necroptosis gene network behaves as a whole in breast cancer, rather than through single-gene case studies. Equally underexplored is the relationship between necroptosis pathway expression and the composition of the immune microenvironment: while both necroptosis biology and immune infiltration have separately been studied in breast cancer, the two have rarely been examined together in a systematic, cohort-integrated manner, and no study to date has tested

whether necroptosis-immune association patterns generalize across independent patient cohorts.

We addressed these gaps by integrating four independent gene expression datasets spanning a combined 569 samples with a curated list of 159 necroptosis-related genes. Specifically, we aimed to (1) identify which necroptosis pathway genes are differentially expressed between breast tumor and normal tissue, (2) build and validate a predictive gene signature using three complementary machine learning approaches together with SHAP-based interpretability, and (3) map the necroptosis-associated immune landscape and determine whether the resulting gene-immune association patterns replicate across independent cohorts. The overall analytical workflow of the study, from multi-cohort transcriptomic integration to machine-learning modeling, immune profiling, and external validation, is summarized in **Figure 1**.

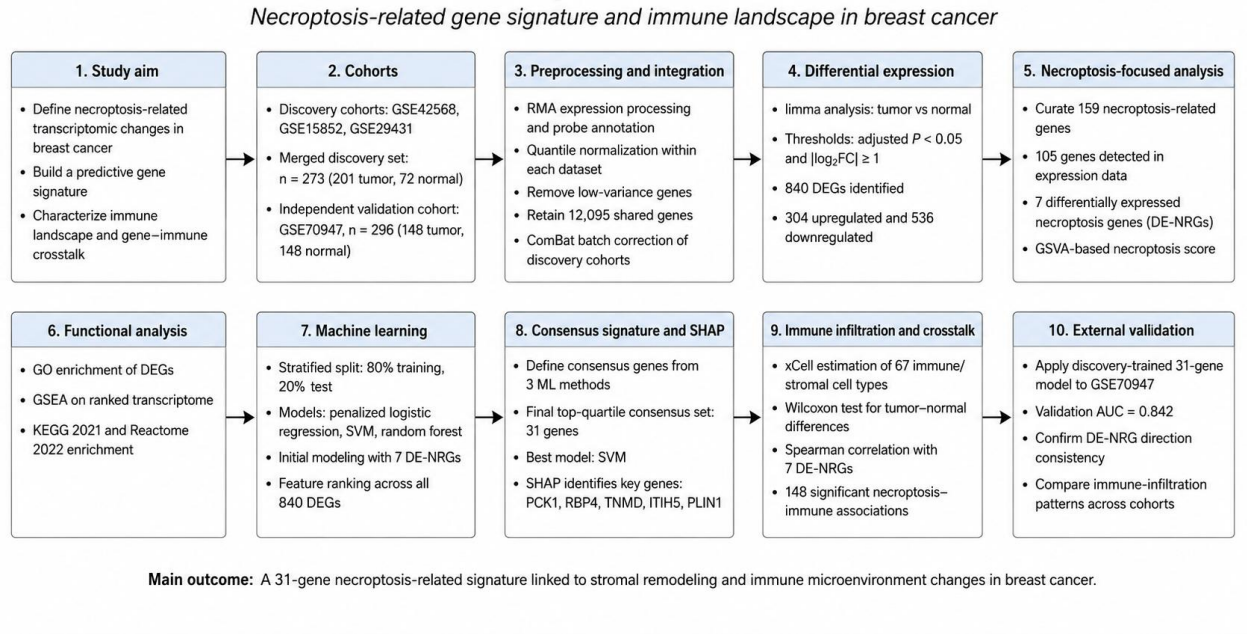


Figure 1: **Overall study workflow.** Three GEO datasets were integrated into a discovery cohort and analyzed using preprocessing, ComBat batch correction, differential expression, necroptosis-focused analysis, functional enrichment, machine-learning feature selection, SHAP interpretation, and xCell-based immune profiling. The resulting 31-gene consensus signature and necroptosis–immune associations were subsequently evaluated in the independent GSE70947 validation cohort.

2. Methods

2.1 Dataset acquisition and preprocessing

Four publicly available gene expression datasets were retrieved from the Gene Expression Omnibus (GEO): GSE29431 (GPL570; 54 tumor and 12 normal samples), GSE70947 (GPL570; 148 tumor and 148 normal sample and GSE15852 (GPL96, Affymetrix Human Genome U133A Array; 43 tumor and 43 normal samples), and GSE42568 (GPL570, Affymetrix Human Genome U133 Plus 2.0 Array; 104 tumor and 17 normal samples) [15]. Dataset 1, 2 and 3 were merged to create a discovery cohort of 273 (201 tumor, 72 normal), and GSE70947 was used as an independent validation cohort (n=296).

The robust multi-array average (RMA) algorithm provided by the package affy of R [16] was used to process raw expression values. Platform-specific packages were used to annotate probes to gene symbols (hgu133plus2.db for GPL570, hgu133a.db for GPL96), and if more than one probe mapped to a gene, the probe with the highest mean expression was selected as representative. Values of the expressions were quantile normalized per data set and the genes with the lowest variance when considering the entire gene set were discarded from further analysis.

2.2 Batch correction and cohort integration

To integrate the genes, only those genes shared across all the three discovery-cohort platforms were kept yielding 12,095 common genes. We used the ComBat algorithm embedded in empirical Bayes [17] in the sva package to correct for batch effects between the three discovery datasets. The model matrix reflected tumor/normal status during correction to ensure that only biologically meaningful variation was retained, while eliminating platform-based technical variation. The efficiency of batch correction was measured using the coefficient of determination (R^2) for batch assignment before and after correction, and the a priori value was set to be below 0.01.

2.3 Differential expression analysis

Genes differentially expressed between tumor and normal samples in the integrated discovery cohort were identified with the limma package [18] using empirical Bayes moderated statistics. Genes were called significant at a Benjamini-Hochberg adjusted p-value below 0.05 and an absolute log2 fold change of at least 1.0.

2.4 Necroptosis gene list curation

A reference list of 159 necroptosis-related genes was assembled from the KEGG necroptosis pathway (hsa04217), the Reactome necroptosis pathway (R-HSA-5213460), and relevant published literature [19, 20]. Gene symbols were harmonized to HGNC-approved nomenclature; all 159 genes were successfully mapped, with no unmapped entries.

2.5 Necroptosis score calculation

A per-sample necroptosis activity score was derived using the GSVA algorithm [21], applied to the 105 necroptosis-related genes detected within the expression data. This score reflects the enrichment of the necroptosis gene set relative to the background transcriptome and was calculated with the Gaussian cumulative-distribution-function-based GSVA method using default parameters.

2.6 Enrichment analysis

Gene Ontology (GO) enrichment across all 840 DEGs was performed with clusterProfiler [22] using the org.Hs.eg.db annotation database (p-value cutoff 0.05, q-value cutoff 0.2). Gene Set Enrichment Analysis (GSEA) was conducted with the fgsea algorithm [23] on the complete gene list, ranked by a signed statistic (sign(logFC) × −log10(p-value)). KEGG 2021 and Reactome 2022 pathway enrichment were additionally examined using EnrichR [24].

2.7 Machine learning models

Three complementary machine learning approaches were applied, first to the seven DE-NRGs and subsequently to a 31-gene consensus set derived from feature-importance ranking. The discovery cohort was partitioned into an 80% training set (n=219; 58 normal, 161 tumor) and a 20% test set (n=54; 14 normal, 40 tumor) using stratified sampling to preserve class balance.

Penalized logistic regression (PLR)

Lasso-regularized logistic regression was implemented with the glmnet package [25], with the regularization parameter (lambda.1se) selected via 10-fold cross-validation; only features retaining non-zero coefficients at this lambda were used in the final model.

Support vector machine (SVM)

A radial basis function kernel SVM was trained with the e1071 package [26]. The cost and gamma hyperparameters were optimized through 10-fold cross-validation over a grid search (cost: 0.1, 1, 10, 100; gamma: 0.001, 0.01, 0.1, 1, 10).

Random forest (RF)

A 500-tree ensemble was constructed using the randomForest package [27] with default settings; feature importance was assessed by mean decrease in Gini impurity.

2.8 Model evaluation

Classifier performance was assessed on the held-out test set using area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPRC), accuracy, balanced accuracy, sensitivity, specificity, positive and negative predictive values (PPV, NPV), F1

score, Matthews correlation coefficient (MCC), and Brier score. AUC confidence intervals were estimated from 2000 bootstrap replicates.

2.9 SHAP interpretation

Feature contributions for the best-performing model (SVM) were quantified with SHAP (SHapley Additive exPlanations) values, computed using the fastshap package [28] with 50 Monte Carlo simulations per feature. Overall SHAP importance was defined as the mean absolute SHAP value across all test samples, and dependence plots were generated for the five highest-ranked features to visualize how feature values relate to their SHAP contributions.

2.10 Consensus gene signature

Consensus genes were defined at three levels of stringency: (1) strict consensus, genes selected by all three ML methods (LASSO at lambda.1se, SVM, and RF variable importance); (2) moderate consensus, genes selected by at least two of the three methods; and (3) top-quartile consensus, genes falling within the top 25% of average feature importance across all three methods. The top-quartile consensus set, comprising 31 genes, was carried forward as the final predictive signature.

2.11 Immune infiltration analysis

Immune and stromal cell-type enrichment was estimated using xCell [29], a gene-signature-based deconvolution method that scores 67 distinct cell types from bulk expression profiles; cell types with an underlying cross-validation AUC above 0.50 in xCell were retained for analysis. Differences in immune infiltration between tumor and normal samples were tested with the Wilcoxon rank-sum test, with Benjamini-Hochberg correction for multiple comparisons. Necroptosis-immune crosstalk was assessed via Spearman correlation between the seven DE-NRGs and the significantly differential immune cell scores across all discovery-cohort samples, with correlations considered significant at FDR<0.05.

2.12 External validation

The complete analytical pipeline was reproduced independently in the GSE70947 validation cohort. The 31-gene consensus signature was tested by applying the random forest model trained in the discovery cohort directly to the validation expression data; the single gene not measured on the validation platform (n=1) was imputed as zero. Immune infiltration patterns and necroptosis-immune correlations from the two cohorts were compared using the Spearman correlation of their respective effect sizes.

3. Results

3.1 Dataset characteristics and preprocessing

The discovery cohort combined 273 samples from three independent datasets: GSE42568 (104 tumor, 17 normal), GSE15852 (43 tumor, 43 normal), and GSE29431 (54 tumor, 12 normal). The independent validation cohort, GSE70947, contributed 296 samples (148 tumor, 148 normal). Following quantile normalization and removal of low-variance genes, 12,095 genes common to all three discovery platforms were retained for integration. The median R^2 for batch assignment (before correction) was 0.736, indicating significant technical differences in the underlying datasets. ComBat correction resulted in a median R^2 of 0.003, while keeping the tumor-normal biological signal by adding the class covariate in the correction model. Post correction, the results of PCA indicated that samples were grouped mostly according to the biological class (tumor vs. normal) and not by the dataset of origin, thus demonstrating the effectiveness of the batch removal (Figure 2a-b).

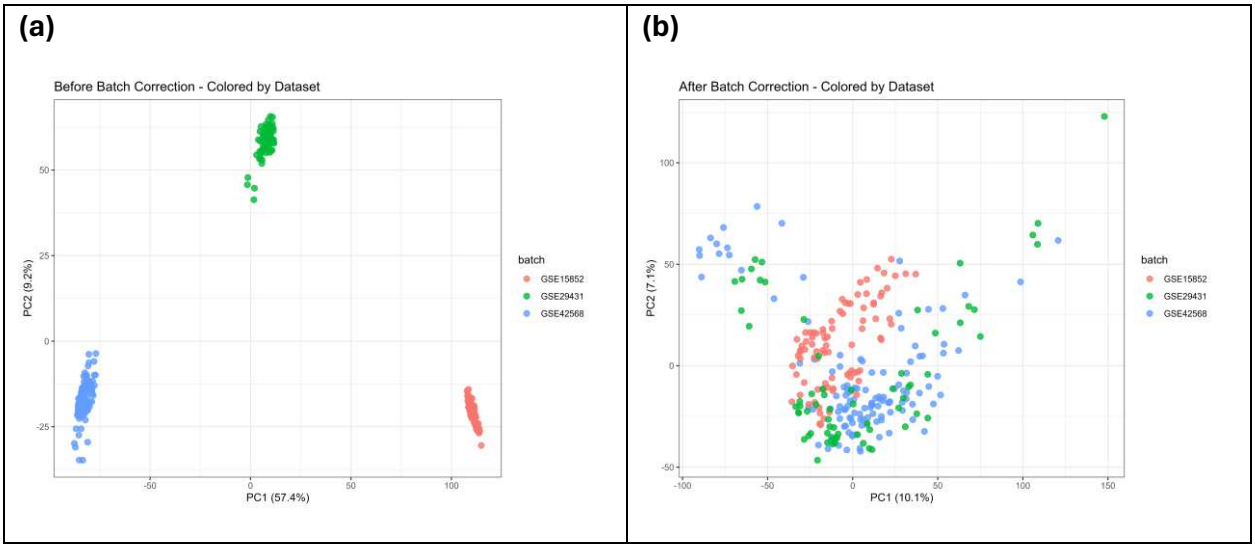


Figure 2: Principal component analysis before and after batch-effect correction. (A) PCA of the three discovery datasets before ComBat correction, showing strong clustering according to dataset of origin. (B) PCA after correction, demonstrating substantially improved cross-dataset integration while retaining biological separation between tumor and normal breast samples.

3.2 Differential expression landscape

In the integrated discovery cohort, samples from 201 tumors were compared to samples from 72 normal tissues. 90% of the identified genes were differentially expressed, with 304 genes (90%) showing log2FC≥1.0 and/or adjusted p<0.05. 536 genes were downregulated and 5 genes were upregulated in the tumor tissue. The most strongly upregulated genes included KRT19 (logFC=+4.56, p=6.62e-51), a marker of luminal epithelial differentiation [30], together with CD24 (+3.62, p=2.49e-40), EPCAM (+3.16, p=1.80e-46), SDC1 (+2.82, p=6.76e-44), and SPINT2 (+2.52, p=2.07e-48). The most strongly downregulated genes were RBP4 (logFC=-5.21, p=1.07e-58),

PCK1 (-4.17, p=2.17e-58), PDE3B (-4.05, p=1.43e-57), SDPR (-4.00, p=5.46e-50), and CIDEC (-3.87, p=9.90e-42) (**Figure 3A-B**). This is a simultaneous loss of metabolic (PCK1, RBP4, PDE3B) and adipocyte-associated transcripts (CIDEC, PLIN1, LPL) are consistant with the stromal remodeling and metabolic reprogramming previously reported in breast tumors [31, 32].

GO enrichment of all 840 DEGs showed enrichment in extracellular matrix organization (GO:0030198), The collagen-containing extracellular matrix (GO:0062023) and receptor ligand activity. The most prominent terms were (GO:0048018) and others. GSEA also identified co-ordinated. Immune-related pathways, such as interferon-gamma response and inflammatory pathways, were enriched. The changes in the immune microenvironment, as seen in the Figure, are typical of signaling. breast tumors [33].

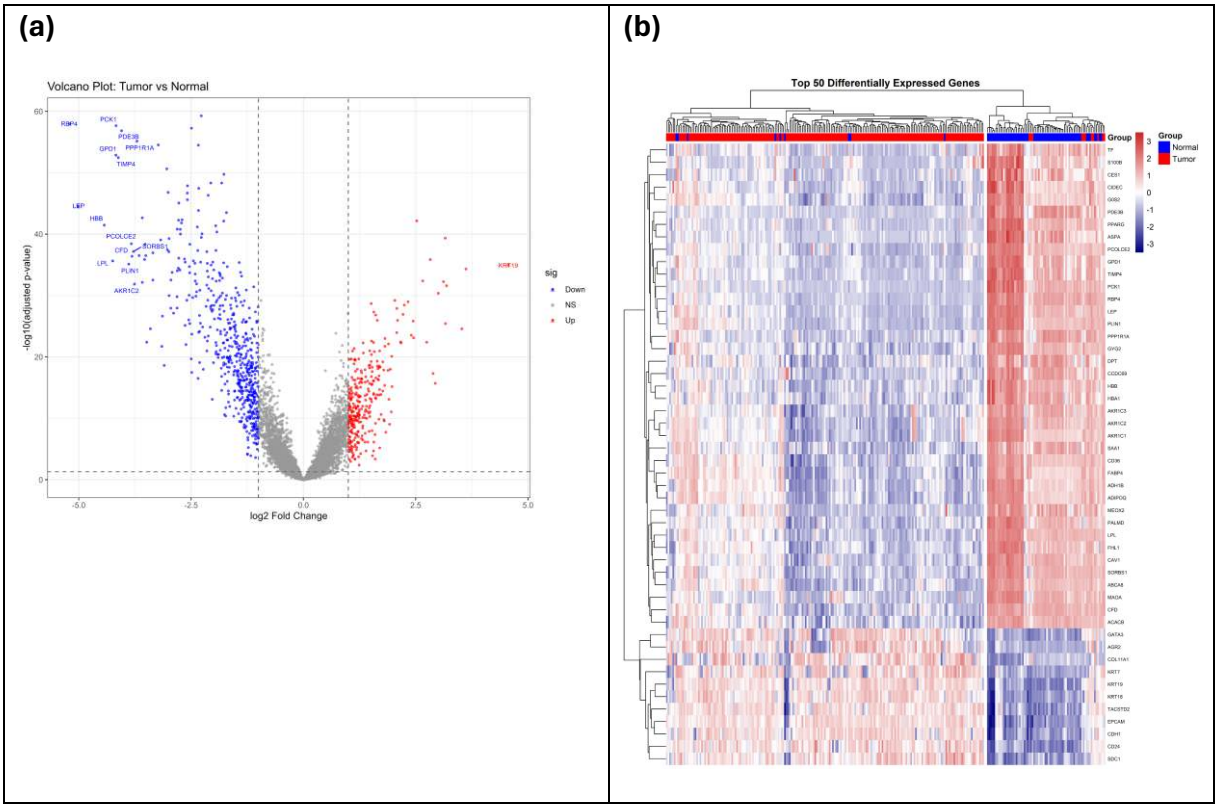


Figure 3: Differential expression landscape of the integrated breast cancer discovery cohort. (A) Volcano plot showing differentially expressed genes between tumor and normal breast tissue using an adjusted $P < 0.05$ and $|\log_2 FC| \geq 1$ threshold. (B) Heatmap of the most significantly differentially expressed genes across discovery samples, illustrating distinct tumor and normal expression patterns. A total of 840 DEGs were identified, including 304 upregulated and 536 downregulated genes.

3.3 Necroptosis-related gene expression in breast cancer

The processed expression data contained data for 105 (66.0%) of the 159 curated necroptosis genes, with the remaining 54 genes missing due to platform coverage limitations, including a large proportion of histone variants, type I interferon family members and PLA2G4 subfamily paralogs. Among the 840 DEGs, seven were necroptosis-related (DE-NRGs): STAT1 (logFC=+1.23,

adj.p=1.16e-09), PLA2G4A (-2.16, 5.38e-17), STAT5A (-1.25, 1.02e-06), PYGL (-1.75, 1.70e-12), TLR4 (-1.05, 2.89e-07), MAPK10 (-1.26, 3.69e-06), and IL33 (-1.42, 2.18e-13) (**Figure** 4a).

Six of the seven DE-NRGs were shown to be downregulated (unless STAT1 was the only gene). showing increased expression pattern indicating broad suppression of necroptotic machinery In line with previous reports of RIPK3 silencing in cancer [11], we found that breast tumors exhibited decreased RIPK3 expression. Consistent with this, The GSVA-based necroptosis score was significantly higher (p=6.04e-07) in tumor samples (**Figure 4b**), confirming pathway-level suppression. Among the DE-NRGs, STAT5A showed the strongest positive correlation with the necroptosis score (Spearman r=0.516, p=5.31e-20), followed by TLR4 (r=0.491, p<1e-300), PLA2G4A (r=0.446, p=9.30e-15), and IL33 (r=0.438, p=3.40e-14). STAT1 correlated only weakly and non-significantly with the pathway score (r=0.106, p=0.081) (**Figure 4c-d**), suggesting it may be regulated at least partly independently of the core necroptosis machinery.

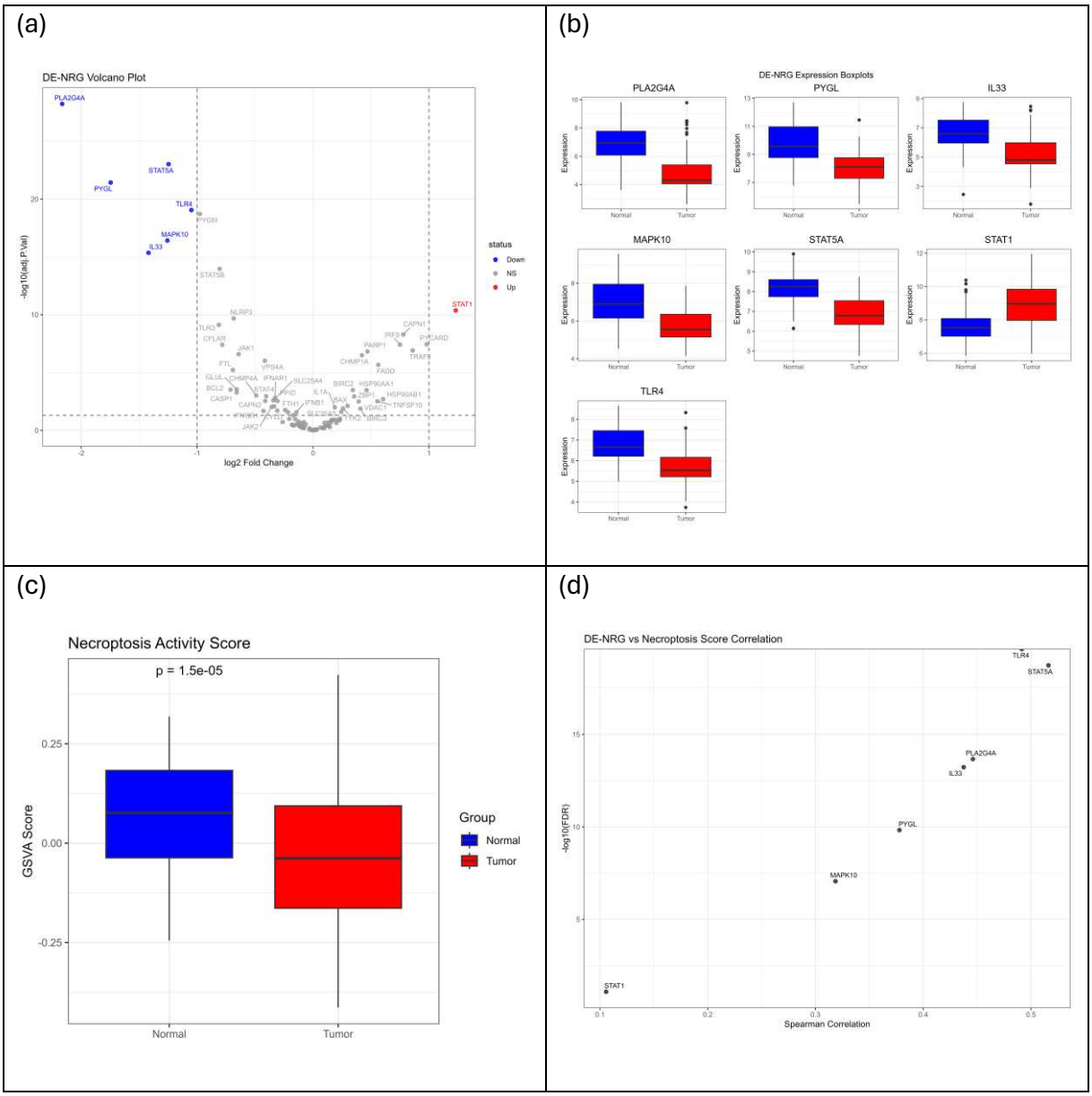


Figure 4: Expression and pathway-level activity of necroptosis-related genes in breast cancer. (A) Volcano plot highlighting the seven differentially expressed necroptosis-related genes (DE-NRGs). (B) Boxplots comparing expression of PLA2G4A, STAT5A, PYGL, TLR4, MAPK10, IL33, and STAT1 between tumor and normal samples. (C) Comparison of GSVA-derived necroptosis activity scores between the two groups. (D) Correlations between individual DE-NRG expression and the necroptosis activity score.

3.4 Machine learning-based gene signature development

Using the seven DE-NRGs as the initial feature set, three classifiers were trained and compared. Lasso-regularized logistic regression, tuned by 10-fold cross-validation, retained five features (PLA2G4A, STAT5A, PYGL, TLR4, STAT1), excluding MAPK10 and IL33 (**Figure 5a**). On the held-out test set (n=54), all three classifiers achieved perfect specificity (1.000), with sensitivity ranging from 0.429 (PLR) to 0.714 (RF); corresponding AUC values were 0.954 (PLR), 0.954 (SVM), and 0.971 (RF) (**Figure 5b,c,d**).

To strengthen the signature, all 840 DEGs were ranked by their average importance across the three ML methods, and the top quartile yielded a 31-gene consensus set. This set incorporated all seven DE-NRGs alongside metabolic and stromal genes such as ITIH5, AKAP12, GPD1, CAV1, LPL, PLIN1, PCK1, PPARG, RBP4, and TNMD. Retraining on these 31 genes improved performance across all three models: SVM achieved the highest AUC (0.968), AUPRC (0.987), and the lowest Brier score (0.049), while PLR and RF achieved AUCs of 0.952 and 0.942, respectively. All three models achieved identical accuracy (0.944), balanced accuracy (0.916), and MCC (0.853) on the test set (**Figure 5e,f**).

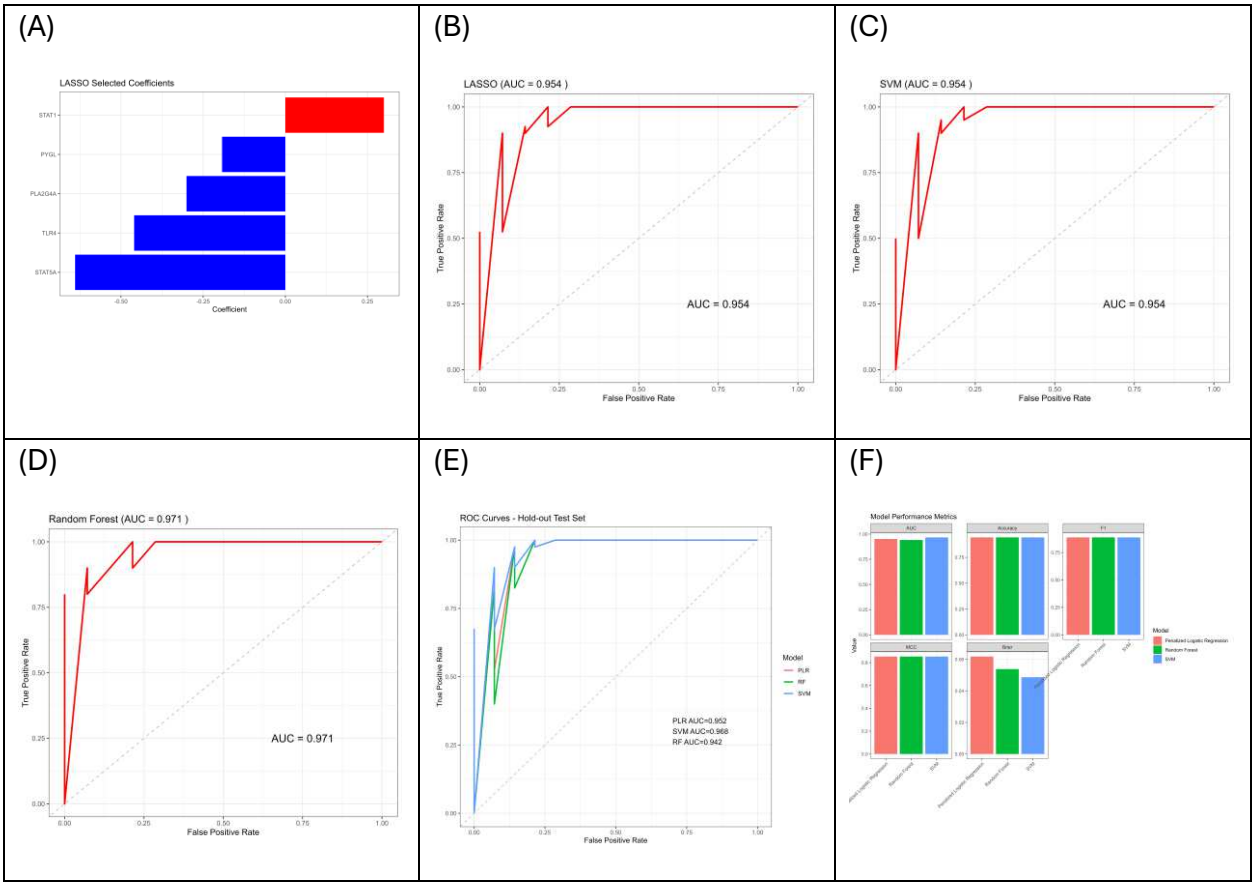


Figure 5: Machine-learning development and evaluation of the necroptosis-related gene signature. (A) LASSO-selected coefficients for the initial DE-NRG feature set. (B–D) ROC curves for penalized logistic regression, support vector machine, and random forest models on the held-out test set. (E) Comparative ROC curves for the models developed using the 31-gene consensus signature. (F) Comparison of classification performance metrics across the final models. The SVM showed the strongest overall discrimination with an AUC of 0.968.

3.5 SHAP interpretation of the best-performing model

SHAP analysis of the SVM model ranked PCK1 as the most influential feature (mean |SHAP|=0.062), followed by RBP4 (0.057), TNMD (0.056), ITIH5 (0.049), and PLIN1 (0.047) (**Figure 6a**) a ranking dominated by metabolic genes. PCK1, a rate-limiting gluconeogenic enzyme frequently silenced across cancer types [34], and RBP4, an adipocyte-expressed retinol transporter, both pushed model predictions toward the tumor class when their expression was

low, consistent with their pronounced downregulation in tumor tissue (PCK1: -4.17; RBP4: -5.21). TNMD and ITIH5, both extracellular matrix-related genes, followed the same directional pattern. Dependence plots for the top features were monotonic **Figure 6b-f**, with lower expression corresponding to higher SHAP values consistent with these genes functioning as suppressed markers of stromal integrity in tumor tissue.

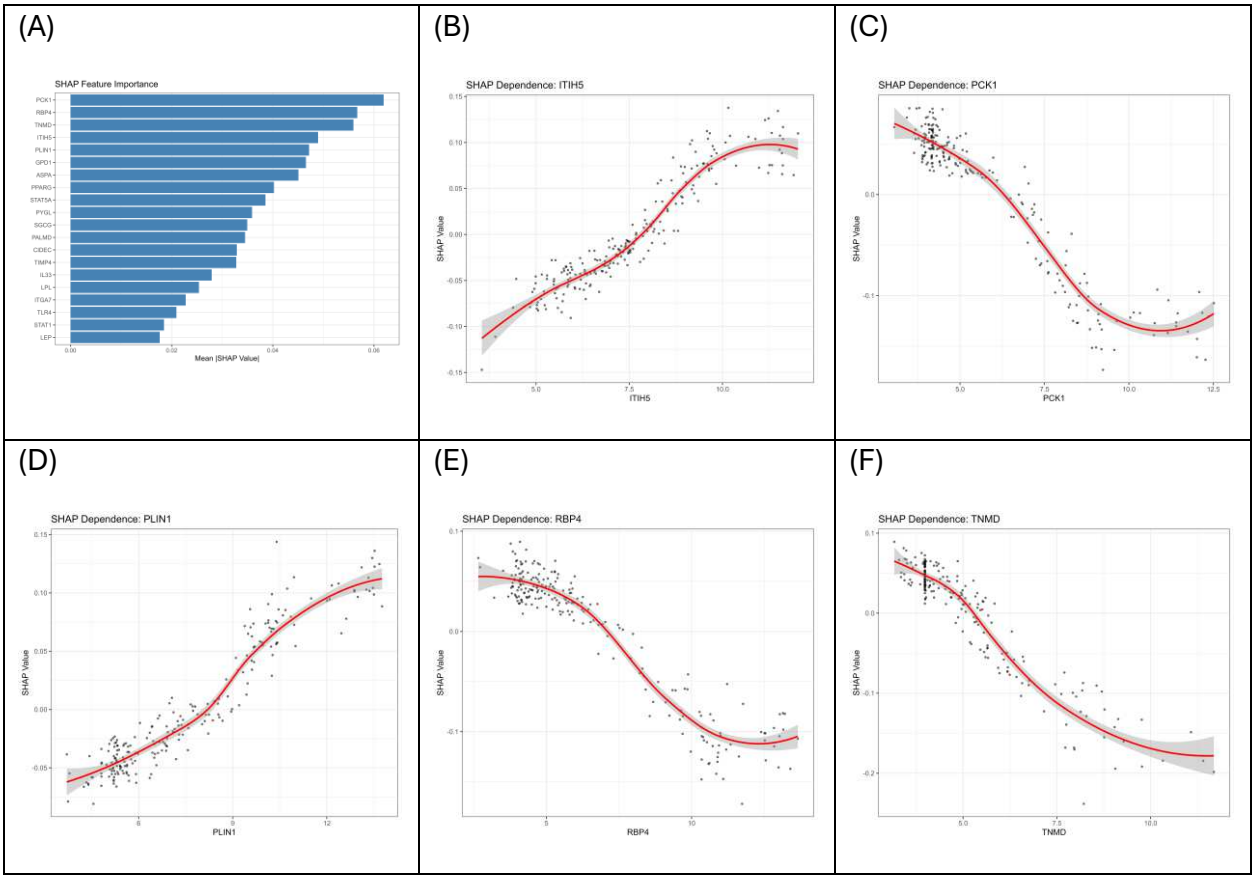


Figure 6: SHAP-based interpretation of the best-performing SVM model. (A) Global feature-importance ranking based on mean absolute SHAP values for the 31-gene signature. (B–F) SHAP dependence plots for the leading predictive genes, including ITIH5, PCK1, PLIN1, RBP4, and TNMD. Lower expression of several metabolic and stromal genes was associated with predictions toward the tumor class

3.6 Immune infiltration landscape

xCell profiling showed that there was widespread remodeling of the immune and stromal compartments in breast tumors. In a total of 67 cell types, epithelial cell scores are significantly elevated in tumors. than normal tissue (tumor mean 0.381 vs normal mean 0.231, $p<0.001$, FDR<0.001), reflecting the predicted enrichment of cancerous epithelial cells. Stromal populations were markedly depleted, including adipocytes (0.216 vs 0.442, p=8.02e-27), overall stroma score (0.201 vs 0.440, p=8.20e-25), microvascular endothelial cells (0.175 vs 0.296, p=1.13e-20), lymphatic endothelial cells (0.113 vs 0.223, p=3.09e-20), fibroblasts (0.070 vs 0.217, p=9.22e-17), and hematopoietic stem cells (0.104 vs 0.273, p=2.04e-15) (**Figure7**). These

patterns combine to represent the Malignant transformation of normal adipose and stromal tissue, which is a well-established phenomenon. hallmark of the progression of breast cancer [35].

Some stromal cell subsets that are limited in tumor tissue - adipocytes, fibroblasts and Endothelial cells have been shown to control the initiation of necroptosis via paracrine pathways [36]. All 67 xCell cell types were correlated with the seven DE-NRGs revealing 148 significant which were associated with each other (FDR<0.05) (Figure 7b). Of all the genes, STAT5A exhibited the widest range of correlations. positive associations to multiple stromal populations and negative associations with epithelial The luminal layer was organized, as it is normally in other gastrointestinal organs, in a manner consistent with its known function in maintaining luminal differentiation [13]. TLR4 correlated positively with populations of myeloid-derived cells, consistent with its known function in innate immune sensin.

.

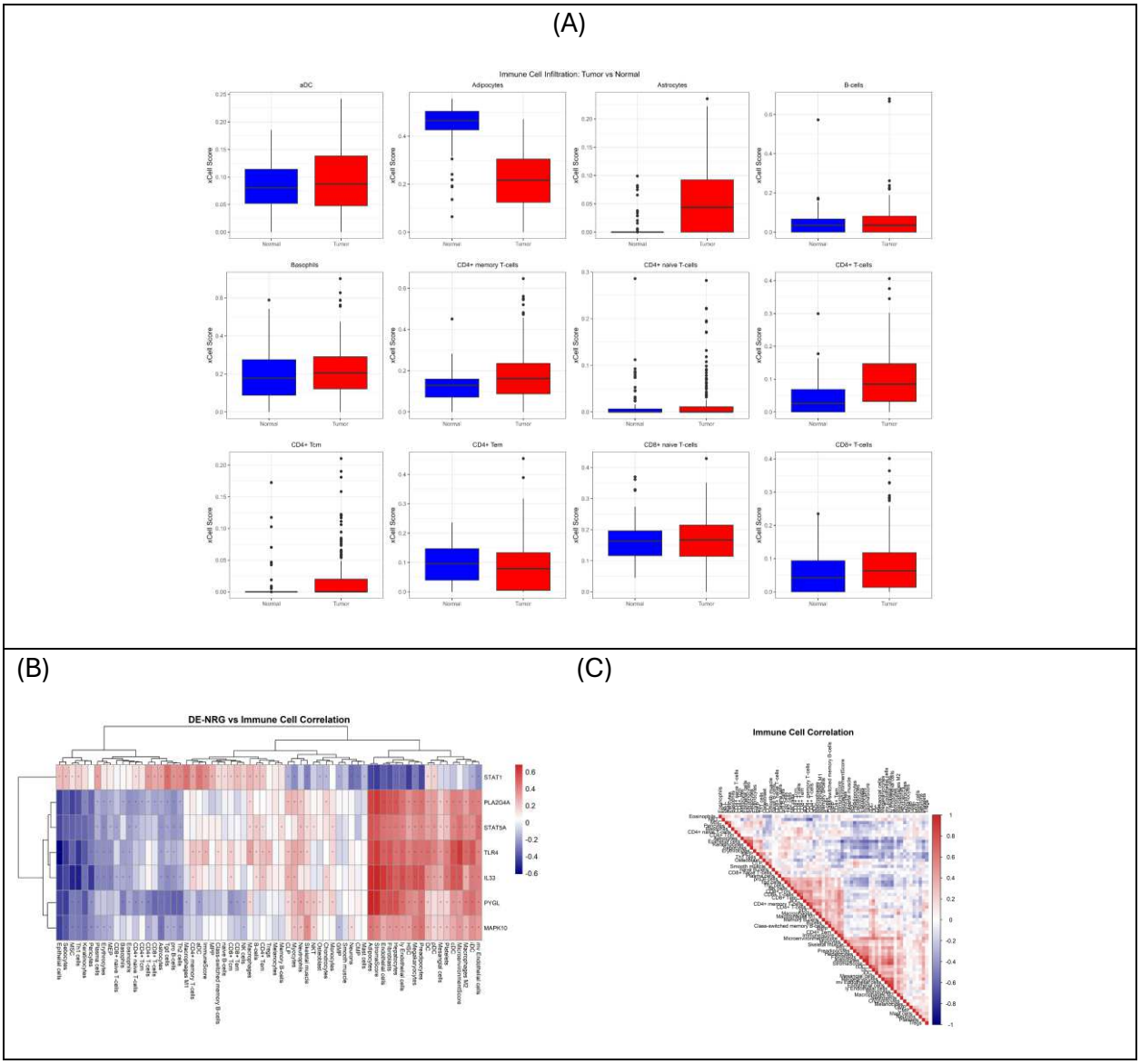


Figure 7: Immune and stromal landscape of breast cancer and its relationship with necroptosis-related genes. (A) Representative xCell comparisons of immune and stromal cell-type enrichment between tumor and normal samples. (B) Heatmap of Spearman correlations between the seven DE-NRGs and xCell-derived immune and stromal scores. (C) Correlation structure among immune and stromal cell populations. Tumors showed prominent epithelial enrichment together with depletion of adipocytes, fibroblasts, endothelial cells, and other stromal populations.

3.7 Necroptosis-immune crosstalk

The necroptosis score at the pathway level was positively correlated with stromal adipocyte, fibroblast and endothelial cell signatures across the 273 discovery samples (r=0.68, r=0.54, r=0.51, respectively), and negatively with epithelial cell content (r=-0.48) (Figure 8). It suggests that increased activity of the necroptosis pathway correlates to well-preserved stromal microenvironment, and the decrease in the expression of necroptosis genes in tumours corresponds to stromal depletion. This relationship is not discernable from observational data

alone, and two non-exclusive interpretations are possible: 1) stromal cells may be the main source of expression of necroptosis genes in breast tissue; or 2) necroptosis signaling itself may play a role in maintaining stromal homeostasis.

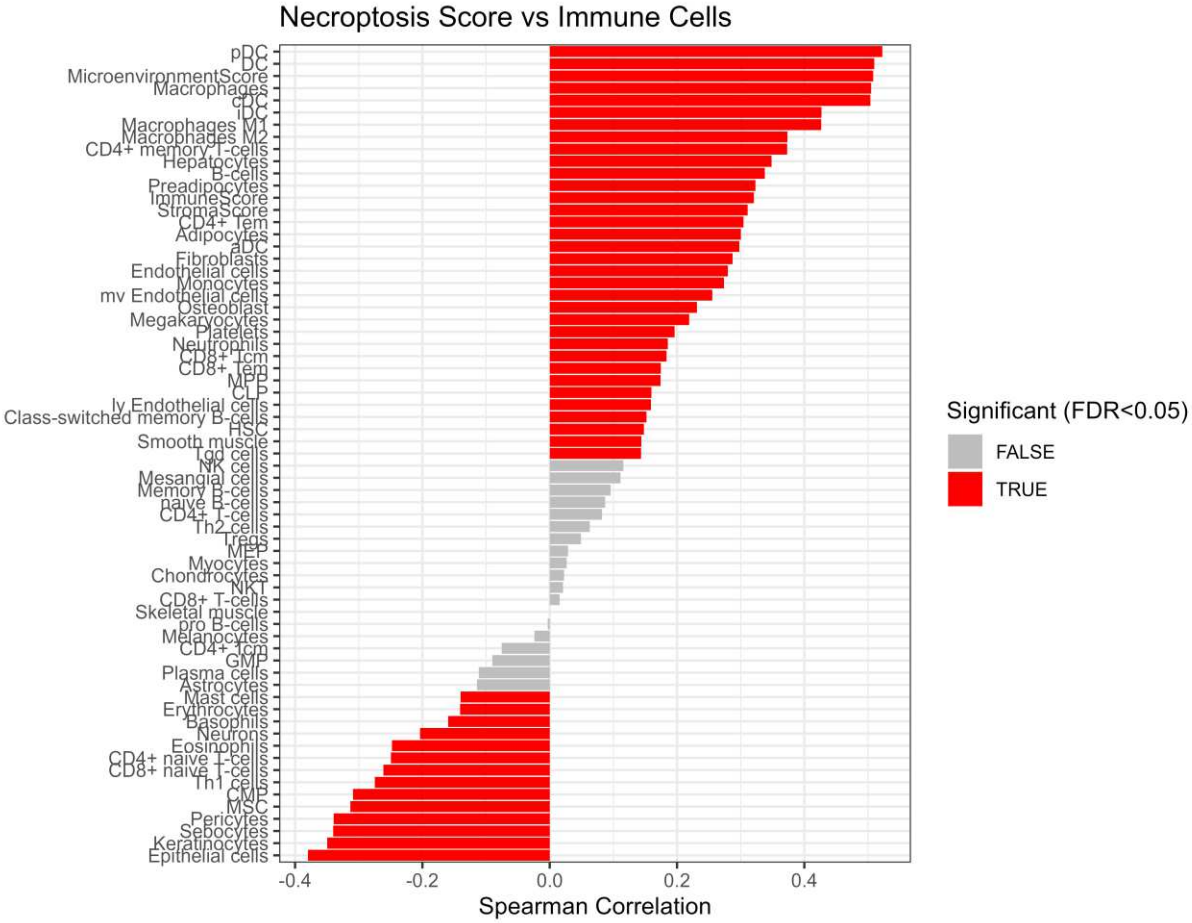


Figure 8: Association between necroptosis activity and immune/stromal cell infiltration in breast cancer. Correlation analysis between the GSVA-derived necroptosis score and xCell-estimated immune and stromal cell populations across the discovery cohort. Higher necroptosis scores were positively associated with stromal components, including adipocytes, fibroblasts, and endothelial cells, and negatively associated with epithelial-cell abundance, indicating a close relationship between necroptosis-associated transcriptional activity and tumor-microenvironment composition

3.8 External validation in an independent cohort

The 31-gene signature was tested in an independent cohort of 296 peripheral blood samples from patients with prostate cancer (GSE70947). A total of 148 breast tumor and 148 normal tissues were obtained. Thirty out of the 31 consensus genes Not represented and thus not detected were PCOLCE2. imputed as zero. The AUC obtained using the discovery-trained random forest model was We evaluated the signature in an independent dataset and found 0.842, which shows that the signature has a discriminatory performance. cohort (Fig. 9A). The model was strongly sensitive (0.993), and moderately specific (0.216) (Fig. 9b). There may be differences in the stromal

composition between cohorts, especially as the signature is based on stromal depletion and was trained with cohorts. batch-corrected discovery data.

All seven DE-NRGs were measured in the validation cohort, and their directions of dysregulation were consistent with those observed in the discovery analysis **(Fig. 9C)**. PLA2G4A, STAT5A, PYGL, TLR4, MAPK10, and IL33 showed lower expression in tumor tissues, whereas STAT1 showed higher expression. Notably, IL33 remained downregulated, with logFC values of −1.42 in the discovery cohort and −1.33 in the validation cohort. TLR4 also remained downregulated, with corresponding logFC values of −1.05 and −0.93.

Immune-infiltration patterns were also broadly conserved between the two cohorts, with a Spearman correlation of 0.757 between the tumor-versus-normal effect sizes across all 67 evaluated cell types. Consistent with the discovery analysis, epithelial-cell enrichment ($p < 0.001$) and adipocyte depletion ($p = 4.49 \times 10^{-24}$) were reproduced in the validation cohort **(Fig. 9D)**.

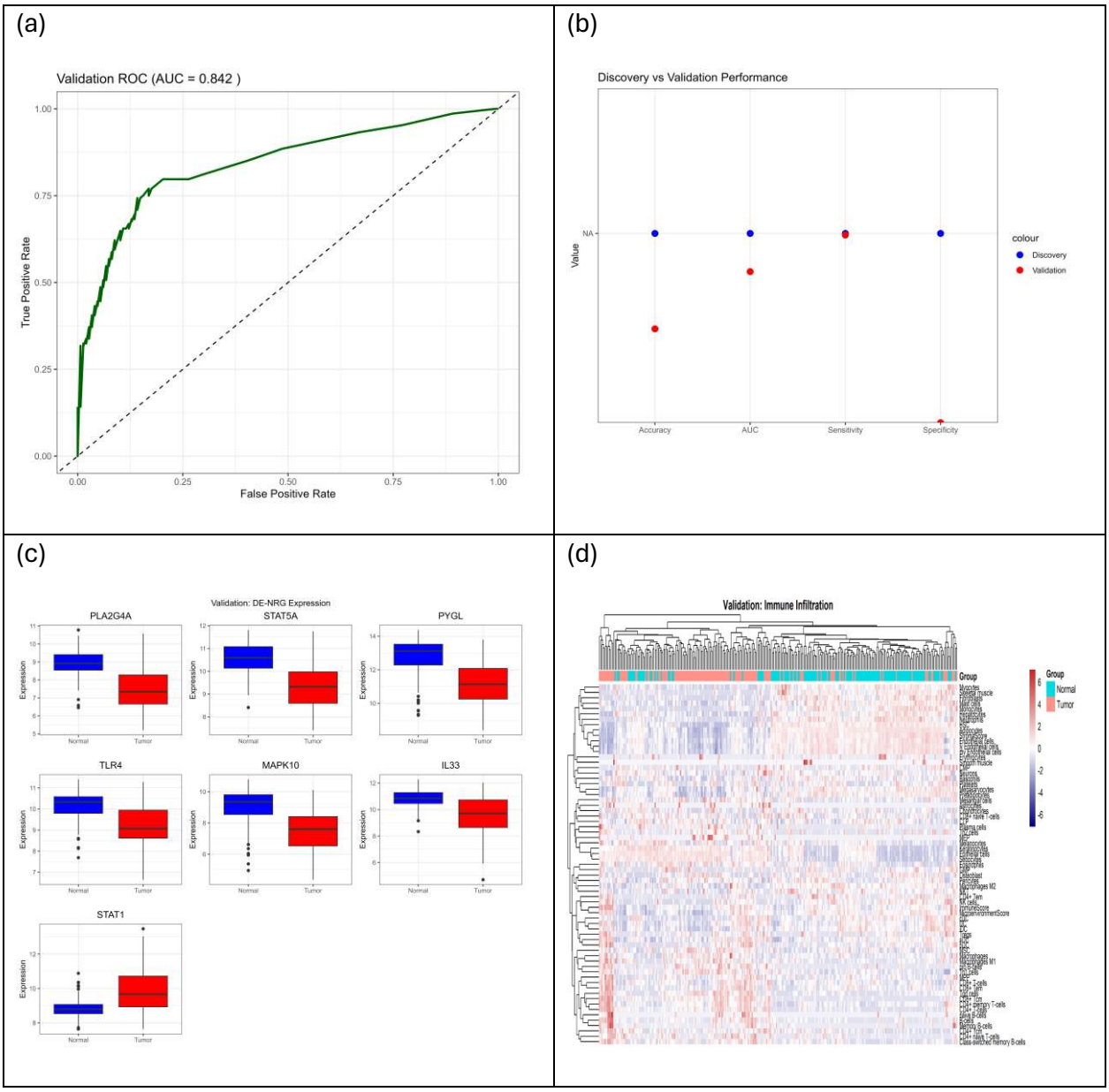


Figure 9: Independent validation of the 31-gene consensus signature in GSE70947. (A) ROC curve of the discovery-trained random forest classifier in the independent validation cohort, yielding an AUC of 0.842. (B) Comparison of predictive performance between the discovery and validation cohorts. (C) Expression patterns of the seven DE-NRGs in validation tumor and normal samples, demonstrating preservation of their direction of dysregulation. (D) Heatmap of immune and stromal infiltration patterns in the validation cohort.

3.9 Enrichment and pathway analysis

GO enrichment of the 840 DEGs identified 47 significantly enriched Biological Process terms, including extracellular matrix disassembly, collagen metabolic process, and response to TGF-beta. Cellular Component enrichment centered on collagen-containing extracellular matrix, basement membrane, and focal adhesion, while Molecular Function terms included extracellular matrix structural constituent, glycosaminoglycan binding, and growth factor binding. KEGG pathway analysis via EnrichR highlighted complement and coagulation cascades, PPAR signaling,

and ECM-receptor interaction, while Reactome analysis emphasized extracellular matrix organization and degradation together with interleukin signaling. GSEA showed coordinated upregulation of immune activation programs (interferon-gamma signaling, antigen processing) alongside downregulation of metabolic and developmental programs **(Fig 10**).

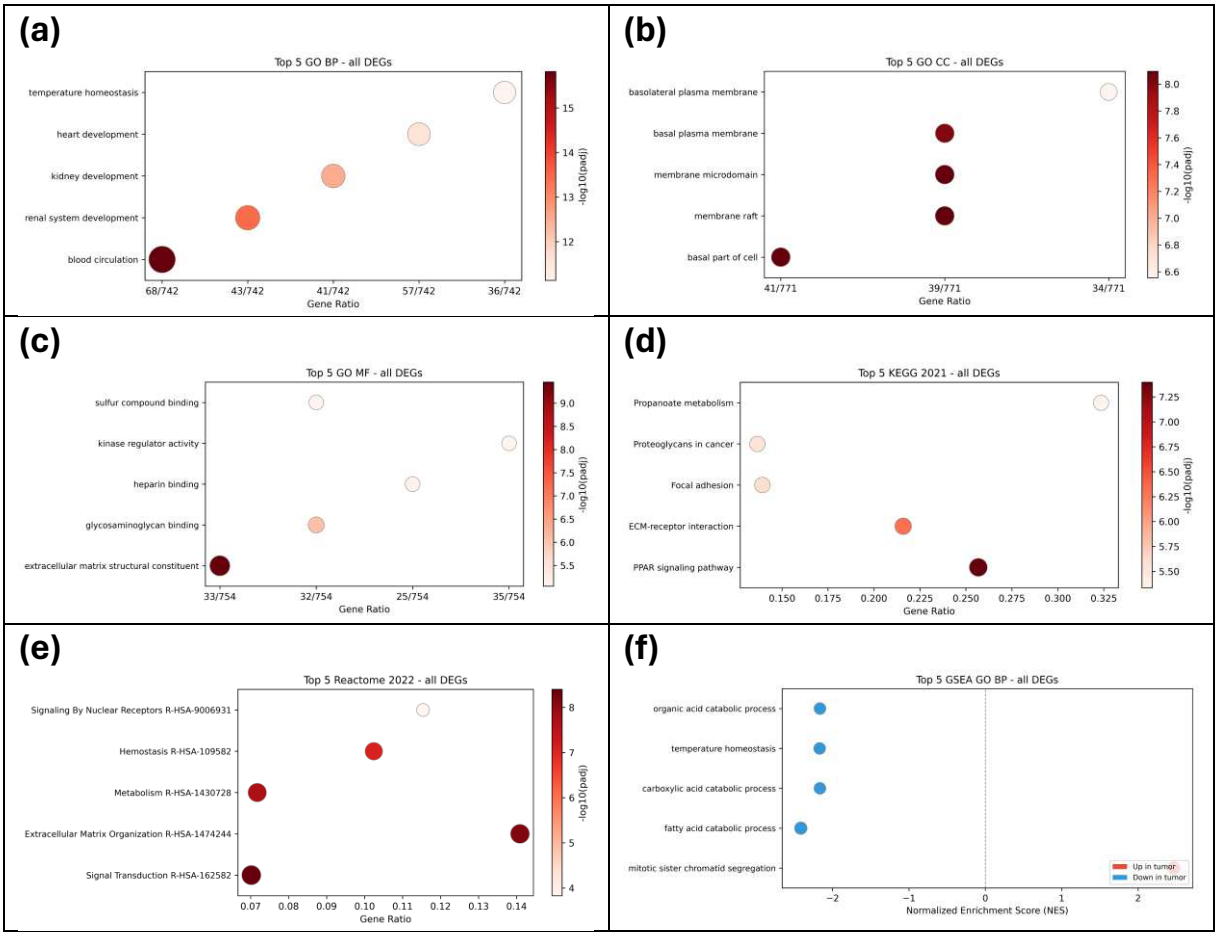


Figure 10: Functional and pathway enrichment analysis of differentially expressed genes. (A–C) Gene Ontology enrichment for Biological Process, Cellular Component, and Molecular Function categories. (D) KEGG 2021 pathway enrichment. (E) Reactome 2022 pathway enrichment. (F) Gene set enrichment analysis of the ranked transcriptome. Enriched processes prominently involved extracellular-matrix organization, PPAR signaling, focal adhesion, metabolic regulation, and immune-related pathways

4. Discussion

The main results of this study are the widespread suppression of activity of the necroptosis pathway in breast tumors: six out of seven DE-NRGs were downregulated and the GSVA pathway score was This suppression was confirmed at system level. This pattern is thought to contribute to a view of necroptosis as a One tumor-suppressive pathway that may allow tumors to survive and progress is one that has been silenced [8, 11]. This one upregulated DE-NRG was STAT1, which is interesting and requires further investigation.

As STAT1 is a dual-function transcription factor that regulates interferon signaling, and can both induce and inhibit. Under certain circumstances, activation of the pathway leads to apoptosis and under other circumstances, it leads to necroptosis [12]. Its selective upregulation, in the background Elsewhere in pathway, the suppression raises two possibilities: that there is selective pressure to maintain anti-tumor immune signaling by interferon, and disable the execution arm of Involving necroptotic death, or that the immune signaling activities of STAT1 are mostly independent from its role in cell death. in the death of cells, the observed signal being not from the cancer itself, but rather from the infiltrating immune cells. epithelium [37]. Single-cell transcriptomic data would allow a distinction to be made between these possibilities.

The necroptosis signature consists of 31 genes including core necroptosis regulators and an expanded group of stromal and metabolic genes. The top genes controlling metabolic pathways that dominated the SHAP were PCK1, RBP4, PLIN1 and PPARG. metabolic reprogramming, which links the dysregulation of necroptosis with importance ranking, characterizes breast cancer. Whereas, loss of PCK1 has been previously linked with higher glycolytic Meanwhile, there is flux between multiple cancer types [34] and RBP4, which is mainly expressed in adipocytes, demonstrates. An orderly decrease in the level of sharpness, which corresponds with the transformation of the adipose tissue into malignant epithelial tissue. As biomarkers, these genes are likely to work well as sentinels of stromal Breast tumorigenesis disrupts the integrity of which is disrupted early [35].

Immune infiltration patterns were highly reproducible across cohorts. Depletion of adipocytes, fibroblasts, and endothelial cells, together with epithelial cell enrichment, replicated with a cross-cohort correlation of 0.757 — a notably high figure given the differences in array platform, patient population, and sample handling between datasets. The positive correlation between the necroptosis score and stromal signatures, and its negative correlation with epithelial content, raises the possibility that necroptosis pathway activity in bulk breast tissue is concentrated primarily within the stromal compartment. If correct, a low necroptosis score in a bulk tumor sample may partly reflect stromal loss rather than pathway silencing within malignant cells themselves — a distinction with implications for how such signatures should be interpreted going forward.

Several limitations warrant acknowledgment. First, reliance on bulk tissue expression limits the ability to assign necroptosis gene expression changes to specific cell populations; single-cell RNA sequencing or spatial transcriptomic approaches would help resolve this. Second, there was less specificity in the validation cohort (0.216) than in the discovery test set, which could be due to the difference in tissue type (core needle biopsy vs surgical resection), or simply because of differing degrees of platform sensitivity or true biological differences between cohorts. Third, 54 of the 159 curated necroptosis genes were not present on the used microarray platforms and RNA-sequencing studies would provide fuller coverage of the pathways. Fourth, the necrosis score is a gene-set based proxy and not a direct functional readout and experimental validation in matched tumour and normal tissue would bolster these observations. Fifth, clinical outcomes were not available, and this could be a potential direction for future research, as the prognostic value of the necroptosis signature could be enhanced through incorporating clinical outcomes. Despite these constraints, some factors contribute to the trustworthiness of the results.

Four separate cohorts were combined with formal batch correction to yield more statistical power than would have been obtained from any individual cohort. Combining three complementary machine learning methods with the SHAP-based interpretation linked predictive performance to underlying biology. Reproducibility of patterns of immune infiltration across cohorts is some of the strongest support that the associations observed are a true reflection of tumor biology and not a dataset-specific artifact. All analysis scripts, gene lists and intermediate results have been shared to facilitate follow-up investigation.

Clinical implications

A low necroptosis score corresponds to higher stromal loss and might indicate more aggressive disease. The association of necroptosis with immunogenic cell death suggests that the proposed signature may be used in principle for patient stratification in immunotherapy, as tumours with higher levels of necroptosis pathway activity might be more sensitive to checkpoint blockade [38] and this will need to be tested in prospective studies. In a therapeutic context, the general downregulation of the necroptosis pathway genes suggests that reactivation of the pathway by means of inhibiting SMAC interaction with the inhibitor of apoptosis proteins (IAP) or by blocking interaction between RIPK3 and IAPs using specific drugs, such as SMAC mimetics, IAP antagonists or RIPK3 overexpression, may be beneficial [39]. There is also the fact that other regulators of necroptosis serve other functions unrelated to death like STAT1 in interferon signaling, and TLR4 in innate immune sensing; if necroptosis was to be reactivated, this would have to be done with great caution to prevent unwanted inflammatory effects.

5. Conclusion

This study provides an integrative transcriptomic characterization of necroptosis-related dysregulation in breast cancer, combining multi-cohort differential expression analysis, machine-learning-based biomarker discovery, and immune-infiltration profiling with external validation. Analysis of the integrated discovery cohort identified 840 differentially expressed genes, including seven differentially expressed necroptosis-related genes (DE-NRGs), six downregulated in tumor tissue alongside upregulation of STAT1. This pattern, supported by GSVA-based pathway analysis, indicates that breast cancer involves coordinated, multi-component disruption of necroptosis-associated transcriptional programs rather than dysregulation of a single regulatory node.

Machine-learning analysis identified a 31-gene consensus signature that robustly discriminated breast tumor from normal tissue, with a support vector machine classifier achieving strong discrimination in the discovery set (AUC = 0.968) and retaining meaningful performance in the independent GSE70947 validation cohort (AUC = 0.842). SHAP-based interpretability identified PCK1, RBP4, TNMD, ITIH5, and PLIN1 as principal contributors to model predictions. The predominance of metabolic, adipocyte-associated, and extracellular-matrix-related genes among these top features suggests that the diagnostic signal extends beyond canonical necroptosis regulators, reflecting broader metabolic and stromal remodeling in the tumor microenvironment.

Immune-infiltration analyses revealed substantial reorganization of the tumor microenvironment, with increased epithelial-cell representation and reduced adipocyte, fibroblast, endothelial, and overall stromal populations. Correlation analysis identified 148 significant associations between DE-NRGs and immune or stromal cell fractions, linking necroptosis-associated transcriptional changes to microenvironmental composition. These patterns, along with the direction of DE-NRG dysregulation, were largely reproduced in the independent validation cohort, supporting the robustness of these findings across cohorts.

Ethics Approval and Consent to Participate

This study used publicly available, de-identified transcriptomic datasets obtained from established public repositories. No new human participants were recruited, and no biological samples were collected directly for this study. Therefore, institutional ethical approval and informed consent were not required for the present computational analysis. All datasets were used in accordance with the terms and policies of their respective repositories and the original studies.

Consent for Publication

Not applicable. This study did not involve the collection or publication of identifiable personal information from individual participants.

Competing Interests

The authors declare that they have no competing financial or non-financial interests that could have influenced the work reported in this manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement

All datasets analyzed in this study were obtained from publicly accessible databases and repositories. The breast cancer gene-expression datasets were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under the following Series accession numbers:

GSE29431 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29431),

GSE70947 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70947),

GSE15852 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15852),

and **GSE42568** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42568).

These records provide access to the corresponding expression data, sample information, platform annotations, and associated supplementary files. The GSE15852 record, for example, contains the original breast tumor and paired normal-tissue expression data and associated raw/processed files.

Pathway enrichment and biological interpretation were performed using publicly available pathway resources, including the **Kyoto Encyclopedia of Genes and Genomes (KEGG)** and **Reactome Pathway Database**. KEGG pathway information and pathway maps are available through the specific KEGG pathway resources at:

https://www.kegg.jp/kegg/pathway.html and

https://www.kegg.jp/pathway/.

KEGG provides pathway maps and associated biological annotations that can be accessed using individual KEGG pathway identifiers. Reactome pathway data and analysis resources are available at

https://reactome.org/ and

https://reactome.org/download-data/, including pathway information, pathway hierarchies, identifier mappings, and downloadable pathway datasets.

The analysis code is available on

GitHub:

https://github.com/shehzadkhalil911-cloud/WGCNA/tree/main/breast_cancer_5_necroptosis

Author Contributions

SK: Conceptualization, methodology, formal analysis, bioinformatics analysis, visualization, data interpretation, and writing of the original manuscript.

Acknowledgements

The authors have no specific acknowledgements to declare.

Clinical trial number:

not applicable

References

1. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
2. Waks, A.G. and E.P. Winer, *Breast cancer treatment: a review.* Jama, 2019. **321**(3): p. 288-300.
3. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence.* Clinical cancer research, 2007. **13**(15): p. 4429-4434.
4. DeNardo, D.G. and L.M. Coussens, *Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression.* Breast cancer research, 2007. **9**(4): p. 212.
5. Binnewies, M., et al., *Understanding the tumor immune microenvironment (TIME) for effective therapy.* Nature medicine, 2018. **24**(5): p. 541-550.
6. Galluzzi, L., et al., *Necroptosis: mechanisms and relevance to disease.* Annual review of pathology: mechanisms of disease, 2017. **12**: p. 103-130.
7. Kaczmarek, A., P. Vandenabeele, and D.V. Krysko, *Necroptosis: the release of damage-associated molecular patterns and its physiological relevance.* Immunity, 2013. **38**(2): p. 209-223.
8. Aaes, T.L., et al., *Vaccination with necroptotic cancer cells induces efficient anti-tumor immunity.* Cell reports, 2016. **15**(2): p. 274-287.
9. Seifert, L., et al., *The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression.* Nature, 2016. **532**(7598): p. 245-249.
10. Najafov, A., H. Chen, and J. Yuan, *Necroptosis and cancer.* Trends in cancer, 2017. **3**(4): p. 294-301.
11. Koo, G.-B., et al., *Methylation-dependent loss of RIP3 expression in cancer represses programmed necrosis in response to chemotherapeutics.* Cell research, 2015. **25**(6): p. 707-725.
12. Thapa, R.J., et al., *Interferon-induced RIP1/RIP3-mediated necrosis requires PKR and is licensed by FADD and caspases.* Proceedings of the National Academy of Sciences, 2013. **110**(33): p. E3109-E3118.
13. Yamashita, H., et al., *Stat5 expression predicts response to endocrine therapy and improves survival in estrogen receptor-positive breast cancer.* Endocrine-related cancer, 2006. **13**(3): p. 885-893.
14. Mercogliano, M.F., et al., *Harnessing tumor necrosis factor alpha to achieve effective cancer immunotherapy.* Cancers, 2021. **13**(3): p. 564.
15. Clarke, C., et al., *Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis.* Carcinogenesis, 2013. **34**(10): p. 2300-2308.
16. Gautier, L., et al., *affy—analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-315.
17. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-127.
18. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47-e47.

19. Weinlich, R., et al., *Necroptosis in development, inflammation and disease.* Nature reviews Molecular cell biology, 2017. **18**(2): p. 127-136.
20. Pasparakis, M. and P. Vandenabeele, *Necroptosis and its role in inflammation.* Nature, 2015. **517**(7534): p. 311-320.
21. Hänzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data.* BMC bioinformatics, 2013. **14**(1): p. 7.
22. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The innovation, 2021. **2**(3).
23. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the national academy of sciences, 2005. **102**(43): p. 15545-15550.
24. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC bioinformatics, 2013. **14**(1): p. 128.
25. Friedman, J.H., T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent.* Journal of statistical software, 2010. **33**: p. 1-22.
26. Chang, C.-C. and C.-J. Lin, *LIBSVM: A library for support vector machines.* ACM transactions on intelligent systems and technology (TIST), 2011. **2**(3): p. 1-27.
27. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
28. Maksymiuk, S., A. Gosiewska, and P. Biecek, *Landscape of R packages for eXplainable Artificial Intelligence.* arXiv preprint arXiv:2009.13248, 2020.
29. Aran, D., Z. Hu, and A.J. Butte, *xCell: digitally portraying the tissue cellular heterogeneity landscape.* Genome biology, 2017. **18**(1): p. 220.
30. Bartek, J., et al., *Differential expression of keratin 19 in normal human epithelial tissues revealed by monospecific monoclonal antibodies.* The Histochemical Journal, 1986. **18**(10): p. 565-575.
31. DeBerardinis, R.J. and N.S. Chandel, *Fundamentals of cancer metabolism.* Science advances, 2016. **2**(5): p. e1600200.
32. Bussard, K.M., et al., *Tumor-associated stromal cells as key contributors to the tumor microenvironment.* Breast Cancer Research, 2016. **18**(1): p. 84.
33. Thorsson, V., et al., *The immune landscape of cancer.* Immunity, 2018. **48**(4): p. 812-830. e14.
34. Montal, E.D., et al., *Inhibition of phosphoenolpyruvate carboxykinase blocks lactate utilization and impairs tumor growth in colorectal cancer.* Cancer & metabolism, 2019. **7**(1): p. 8.
35. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* cell, 2011. **144**(5): p. 646-674.
36. Lisanti, M.P., et al., *JNK1 stress signaling is hyper-activated in high breast density and the tumor stroma: connecting fibrosis, inflammation, and stemness for cancer prevention.* Cell cycle, 2014. **13**(4): p. 580-599.
37. Meissl, K., et al., *The good and the bad faces of STAT1 in solid tumours.* Cytokine, 2017. **89**: p. 12-20.
38. Kroemer, G., et al., *Immunogenic cell death in cancer therapy.* Annual review of immunology, 2013. **31**(1): p. 51-72.
39. Fulda, S., *Promises and challenges of Smac mimetics as cancer therapeutics.* Clinical Cancer Research, 2015. **21**(22): p. 5030-5036.