Volume 3, Number 3, 2026
Pages 1010-1025
Bioscience Insights
ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)
BiosciInsights.com

GENE EXPRESSION PROFILING OF INFLAMMATION-RELATED GENES IN BREAST CANCER TISSUES: INSIGHTS INTO THE TUMOR MICROENVIRONMENT

Sadia Shahbaz[*1], Amina Sehar[2], Asad Zulfiqar[3], Rumeesa[4], Areeb Anwar Usmani[5]

[*1]*M.Phil. Scholar, Molecular Biology and Biotechnology Department, Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore*
[2]*University of Central Punjab*
[3]*MPhil Zoology at Mirpur University of Science and Technology (MUST), Mirpur, Azad Kashmir*
[4]*Government College Women University Sialkot*
[5]*Biomedical and Biological Sciences Sohail University Karachi*

[*1]sadiaraishahbaz@gmail.com, [2]aminaseharleo26@gmail.com, [3]asadzulfiqar25@gmail.com, [4]rumeesashafique25@gmail.com, [5]areeb.usmani@sohailuniversity.edu.pk

https://doi.org/10.5281/zenodo.23727174

Article Information

Article History

Received:	June 19, 2026
Revised:	August 19, 2026
Accepted:	September 07, 2026
Available Online:	October 28, 2026

Keywords:
Breast cancer; inflammation; gene expression; tumor microenvironment; RNA sequencing; quantitative analysis.

Copyright @Author
Corresponding Author: *
Sadia Shahbaz

Abstract

Inflammation-related gene expression may provide insights into breast cancer biology; however, bulk tissue measurements do not directly determine immune infiltration or pathway activation. This study compared selected inflammatory transcripts in breast cancer and matched adjacent-normal tissues using an exploratory, quantitative, retrospective secondary analysis of six tissue pairs from GSE133998. After filtering, 23,813 features were analyzed with paired tests of $\log_2$(CPM + 1) and Benjamini–Hochberg adjustment. Eight inflammatory genes and six expression signatures were evaluated. Negative binomial sensitivity analyses and external microarray comparisons supplemented the primary analysis. Seven genes were estimable, while TNF was not. TGFB1 demonstrated the largest positive mean paired difference (0.992; 95% confidence interval, 0.344–1.640; adjusted p = 0.139). TGFB1, IL1B, and NFKB1 had unadjusted p-values below 0.05, but none achieved genome-wide significance. No expression signature reached its adjusted significance threshold. TGFB1 met the panel-adjusted threshold in the pooled-dispersion sensitivity model ($\log_2$ fold change, 1.007; adjusted p = 0.0096). External analysis identified 18,028 significant probes among 54,675, but did not establish gene-level replication. Inflammatory transcripts showed heterogeneous directional differences; however, primary analyses did not identify statistically significant differences in the selected genes or signatures after adjustment. TGFB1 merits further investigation, although its sensitivity finding depends on model specification and testing scope.

The small sample size and incomplete analytical documentation limit interpretation. Larger, independently validated studies incorporating cellular and protein measurements are required.

INTRODUCTION

Breast cancer comprises a group of biologically diverse diseases that differ in molecular characteristics, clinical behavior, and treatment response. This diversity arises from variation both between tumors and among cellular populations within individual tumors (Polyak, 2011). While molecular classification has enhanced the selection of systemic treatments, a comprehensive understanding of breast cancer also necessitates consideration of the tissue environment surrounding malignant cells (Waks & Winer, 2019). Analysis of inflammation-related gene expression offers a means to characterize this environment and to identify molecular differences between tumor tissue and adjacent nonmalignant tissue.

The tumor micro-environment consists of immune cells, fibro blasts, endothelial cells, adipocytes, and extracellular matrix components that interact with malignant epithelial cells. These interactions affect tissue organization, vascular development, and responses to local signals (Balkwill et al., 2012). Micro environmental influences are present throughout various stages of tumor progression, with stromal compartments contributing to conditions that facilitate invasion and metastasis (Quail & Joyce, 2013). As a result, a tumor's expression profile reflects input from multiple interacting cell populations. This complexity underscores the biological relevance of inflammatory profiling and necessitates careful interpretation of the cellular processes underlying observed differences.

Inflammation serves as a critical link between tissue injury, immune responses, and malignant progression. Coussens and Werb (2002) demonstrated that inflammatory cells and mediators can influence proliferation, tissue remodeling, and the local conditions that support tumor development. Later studies differentiated between inflammatory responses arising from pre-existing tissue conditions and those triggered by oncogenic changes within developing tumors (Mantovani et al., 2008). More recently, Greten and Grivennikov (2019) highlighted that inflammatory processes evolve during tumor development and involve significant plasticity among participating cells. These perspectives suggest that inflammation-related genes should be analyzed as components of interacting networks rather than as uniformly elevated cancer markers.

Cytokines, chemokines, and transcriptional regulators constitute complementary elements of these networks. IL6, TNF, and IL1B are involved in inflammatory communication; CCL2 and CXCL8 participate in chemokine signaling; and STAT3 and NFKB1 are linked to regulatory pathways that connect extracellular stimuli to cellular responses. The biological significance of these molecules justifies their joint examination, although messenger RNA abundance does not directly indicate the activity of the corresponding proteins or pathways. Inflammatory signaling relies on interactions among ligands, receptors, regulatory proteins, and responding cell populations, which may vary considerably between tissue environments (Greten & Grivennikov, 2019).

TGFB1 adds complexity because TGF-β signaling regulates cellular behavior in a highly context-dependent manner. Its effects depend on cellular differentiation, transcriptional state, and interactions with other signaling systems (Massagué, 2012). Stromal cells are especially relevant in this context. Fibroblasts can modify extracellular matrix organization and communicate with malignant and immune cells via secreted factors, thereby influencing the biological properties of the tumor environment (Kalluri, 2016). Consequently, variation in TGFB1 and other inflammation-related transcripts may reflect multiple overlapping processes, such as changes in stromal composition and transcriptional regulation. Determining which process predominates requires evidence beyond bulk expression data.

Interpreting immune-associated expression requires distinguishing between immune presence and immune function. The concept of immune contexture highlights the type, density, functional orientation, and location of immune cells within tumors (Fridman et al., 2012). Similarly, the cancer–immune set-point framework characterizes antitumor immunity as the result of interactions among tumor, host, and environmental factors (Chen & Mellman, 2017). These frameworks clarify why increased expression of certain immune markers does not necessarily equate to more effective tumor control. Composite

transcript scores can summarize expression patterns but cannot independently determine immune-cell proportions, spatial organization, or functional activity.

A paired comparison of tumor and adjacent-normal tissues provides a valuable approach for investigating local expression differences while minimizing inter-individual variation. However, adjacent tissue should not be assumed to be biologically equivalent to healthy tissue. Research on field cancerization indicates that molecular alterations can precede morphological abnormalities and may extend throughout a tissue field (Curtius et al., 2018). Although field effects were not directly evaluated in this study, the literature supports a cautious interpretation of the comparator. The paired contrast addresses differences between tumor and adjacent-normal tissue within patients, rather than the full distinction between cancer and health.

Transcriptomic analysis necessitates alignment between biological questions and statistical inference. Established expression-analysis frameworks highlight the importance of considering experimental design, variability, and the large number of simultaneous comparisons (Ritchie et al., 2015). In small paired datasets, positive or negative estimates may be biologically interesting but may not provide sufficiently precise evidence of differential expression. Therefore, confidence intervals and multiple-testing-adjusted results are essential for distinguishing directional observations from statistically supported findings. External comparisons also require careful consideration of sample independence and differences between gene-level sequencing measurements and microarray probes.

This study investigated IL6, TNF, IL1B, CXCL8, CCL2, STAT3, NFKB1, and TGFB1 in six matched pairs of breast cancer and adjacent-normal tissues. The primary objective was to estimate the direction, magnitude, and uncertainty of within-patient expression differences. Secondary objectives included exploring inflammatory and immune-marker expression signatures, evaluating sensitivity to alternative statistical models, and assessing external microarray comparisons. The study aimed to provide exploratory transcript-level insights into the breast tumor microenvironment, while distinguishing expression patterns from evidence of immune infiltration, pathway activation, or clinical biomarker performance.

Literature Review

Research on inflammation-related gene expression in breast cancer utilizes molecular profiling, experimental studies, and investigations of immune and stromal cell function. While these approaches provide complementary evidence, each addresses distinct research questions. Tissue profiling identifies expression patterns, experimental interventions elucidate mechanisms, and clinical studies assess associations with patient outcomes. Distinguishing among these forms of evidence is essential when interpreting results from a small paired transcriptomic analysis.

Allinen et al. (2004) established a foundational understanding by characterizing gene expression across the various cellular populations comprising normal breast tissue and breast carcinomas. Their findings demonstrated that molecular alterations in breast cancer extend beyond malignant epithelial cells to include surrounding tissue compartments. This supports the investigation of inflammation-related transcripts within the broader tumor microenvironment. However, it also highlights a limitation of bulk-tissue profiling: observed expression differences cannot be automatically attributed to cancer cells, as stromal and immune populations contribute to the measured signal. Consequently, cell-specific profiling and bulk analysis yield related but distinct information.

Finak et al. (2008) further demonstrated the clinical relevance of stromal expression by identifying stromal gene-expression patterns associated with breast cancer outcomes. Their study established that the nonmalignant compartment can contain clinically informative molecular signals. However, a prognostic signature derived from stromal tissue is distinct from an inflammatory panel measured in mixed tissue. Prognostic interpretation requires clinical outcome data and validation, whereas a paired tumor–adjacent-normal comparison primarily estimates tissue-associated differences. Therefore, while the relevance of stromal expression to prognosis justifies further investigation, it does not establish prognostic value for individual genes examined in a separate dataset.

These observations align with the broader model proposed by Hanahan and Coussens (2012), which posits that recruited stromal populations support cancer development through reciprocal communication with neoplastic cells. Their synthesis emphasizes that the tumor environment is actively organized via

recruitment, activation, and persistence of diverse cell populations. Similarly, Hinshaw and Shevde (2019) described the influence of innate immune and stromal components on tumor progression. Collectively, these reviews indicate that inflammatory expression should be interpreted as a product of interactions among cellular populations. They do not support the expectation that all inflammation-related transcripts will increase simultaneously or that uniform expression patterns will occur across all tumors.

Experimental studies provide direct evidence of how stromal cells generate inflammatory signals. Erez et al. (2010) identified a proinflammatory program in cancer-associated fibroblasts during early neoplastic development. Their principal mechanistic experiments utilized a mouse model of skin carcinogenesis, with related inflammatory signatures also observed in mammary and pancreatic tumors. Inhibition of NF-κB signaling reduced several tumor-supporting activities of these fibroblasts. These findings demonstrate that inflammatory transcription can originate in the stromal compartment and have functional consequences. However, measuring NFKB1 or IL1B expression in bulk breast tissue does not replicate the experimental evidence necessary to establish fibroblast-specific NF-κB activity.

The IL-6/JAK/STAT3 axis constitutes another significant link between inflammatory communication and tumor biology. Johnson et al. (2018) reviewed evidence indicating that this pathway can influence malignant-cell survival and the functional behavior of immune cells within tumors. This includes feedback mechanisms and regulatory events that measuring IL6 or STAT3 transcripts alone does not capture. This distinction is critical when expression estimates differ in direction or do not reach statistical significance. A transcript-level result reflects abundance under the sampled conditions but does not independently establish whether the signaling pathway is activated, inhibited, or functionally unchanged.

Macrophage research underscores the necessity of distinguishing cellular identity from cellular function. Noy and Pollard (2014) described the involvement of tumor-associated macrophages in angiogenesis, invasion, and metastatic processes. Their review emphasized the interactions between macrophages and malignant cells rather than suggesting a single uniform macrophage phenotype. DeNardo and Ruffell (2019) subsequently examined macrophages as regulators of tumor immunity and immunotherapy, highlighting the influence of local environmental conditions on macrophage behavior. These findings support the examination of macrophage-associated transcripts but indicate that a marker score alone cannot fully characterize the functional state of the contributing cells. An elevated score may reflect differences in macrophage-associated expression without clarifying the underlying cause or biological consequence.

Macrophage activity can influence treatment response. Ruffell and Coussens (2015) reviewed mechanisms by which macrophages contribute to therapeutic resistance, including support for malignant-cell survival and modification of the local immune environment. Their work demonstrates that treatment response depends in part on properties external to the cancer-cell genome. However, treatment-related mechanisms cannot be inferred from expression differences in untreated tissue without appropriate clinical information and functional measurements. In exploratory studies lacking treatment-response outcomes, macrophage markers should be interpreted as components of the tissue expression profile rather than as indicators of resistance or therapeutic sensitivity.

The broader immune literature cautions against equating inflammatory activity with tumor promotion. Gonzalez et al. (2018) described immune functions throughout tumor initiation and metastatic progression, encompassing both antitumor responses and processes that facilitate disease advancement. Joyce and Pollard (2009) demonstrated that microenvironmental influences operate at multiple stages of metastasis, from local invasion to growth at distant sites. These perspectives clarify why similar inflammatory mediators may have different implications depending on location, timing, and cellular context. They also establish a critical boundary for tissue profiling: expression differences in primary breast samples alone cannot demonstrate metastatic behavior or immune escape.

Expression signatures provide a practical approach to summarizing immune-associated information across multiple genes. Nirmal et al. (2018) developed immune-cell gene signatures for profiling solid-tumor microenvironments and demonstrated variation in immune-associated expression across tumors. Their work underscores the importance of systematic marker selection and evaluation of how well a signature represents its intended cellular population. A small investigator-defined panel may yield a useful descriptive score, but its interpretation depends on the constituent genes, standardization, aggregation

procedure, and validation. Without these details, comparisons with published signatures are limited. Furthermore, a score based on transcript abundance should not be equated with cell proportion or interpreted as direct histological infiltration.

Analytical methodology represents an additional source of variation between studies. Law et al. (2014) developed the voom approach to model the relationship between mean expression and variability in RNA-sequencing data, assigning precision weights prior to linear-model analysis. Their findings demonstrate that applying a logarithmic transformation alone does not address all statistical properties of sequencing measurements. This consideration is particularly relevant to small paired analyses, where individual observations may strongly influence estimates. Published methods provide a benchmark for evaluating uncertainty and planning reproducible reanalysis. They should not be cited as procedures performed unless they were actually implemented in the study.

Collectively, the literature supports a biologically informed examination of inflammatory transcripts while delineating the limits of conclusions that can be drawn from bulk expression data. Cell-specific and experimental studies establish plausible mechanisms, whereas tissue comparisons identify patterns that may warrant further investigation. This study addresses whether selected inflammation-related genes differ between matched breast cancer and adjacent-normal tissues. Its contribution is to estimate within-patient differences and their associated uncertainty, supplemented by exploratory expression signatures and sensitivity analyses. Determining the cellular origin, functional consequences, or clinical relevance of these patterns requires additional evidence beyond the paired transcriptomic comparison.

Methodology

Study Design

An exploratory quantitative retrospective secondary analysis of publicly available transcriptomic data was conducted to examine inflammation-related gene expression in breast cancer tissues. The primary comparison employed a within-patient paired design, comparing breast cancer tissue to adjacent-normal tissue from the same patient. Complementary analyses assessed inflammatory and immune-marker expression signatures, sensitivity to statistical model specification, and external microarray expression patterns.

The patient was considered the biological unit of replication. Tissue samples from the same patient were analyzed as matched observations rather than as independent samples. The study involved computational analysis of previously collected expression data; no new tissue collection or laboratory experiments were performed.

Data Sources and Analytical Samples

The primary dataset was obtained from the National Center for Biotechnology Information Gene Expression Omnibus under accession GSE133998. The tissue count matrix, GSE133998_cancer_and_adjacent_normal.txt.gz, was used for the paired analysis. Blood-expression measurements were not combined with tissue measurements.

The analytical sample included six breast cancer tissues and six matched adjacent-normal tissues from patients identified as H36, H37, H38, H39, H40, and H42. Tissue classification and patient identifiers were used to maintain the matched structure. Adjacent-normal tissues served as within-patient comparators rather than as independent healthy controls.

The external microarray dataset GSE42568 comprised 104 breast cancer and 17 normal breast tissue arrays measured using the Affymetrix Human Genome U133 Plus 2.0 platform. This dataset was analyzed separately because its groups were unpaired and its measurements were probe-based. Two additional datasets, GSE45827 and GSE65216, were examined for potential expression-profile overlap before considering their suitability for independent replication.

Data Cleaning and Expression Filtering

The primary count matrix was examined for incomplete measurements, nonnumeric values, duplicated identifiers, and negative counts. Features with incomplete measurements or duplicated identifiers were excluded. Missing measurements were not replaced with zero, and no expression values were imputed.

Sample library totals were calculated from the cleaned count matrix prior to expression filtering. Features were retained for hypothesis testing if they contained at least 10 counts in six or more of the 12 tissue samples. The same filtering criterion was applied uniformly across the dataset without reference to statistical significance.

The initial matrix contained 61,155 features. After cleaning, 61,149 complete, uniquely identified features remained, of which 23,813 satisfied the expression-filtering criterion.

Normalization and Transformation

Expression counts were normalized to counts per million to account for differences in sample library totals:

$$\mathrm{CPM}_{gj} = \frac{C_{gj}}{L_j} \times 10^6,$$

where C_{gj} denotes the count for feature g, in sample j, and L_j, denotes the total count in the corresponding cleaned sample library before expression filtering.

Normalized expression values were transformed as:

$$Y_{gj} = \log_2(\mathrm{CPM}_{gj} + 1).$$

The constant of one permitted transformation of zero counts. The resulting values were used in the primary paired analysis. Differences on this scale were reported as differences in transformed expression rather than conventional $\log_2$ fold changes.

Inflammation-Related Gene Panel

The selected panel comprised eight inflammation-related genes: IL6, TNF, IL1B, CXCL8, CCL2, STAT3, NFKB1, and TGFB1. Expression estimates for these genes were extracted from the primary genome-wide analysis.

For each evaluable gene, the analysis reported the mean paired expression difference, its 95% confidence interval, the unadjusted *p-value*, and the genome-wide adjusted *p-value*. Genes without sufficient informative measurements for estimation were reported as non-estimable. Non-estimability was not interpreted as evidence of absent biological expression or absence of a tissue-associated difference.

Primary Paired Statistical Analysis

Within-patient expression differences were calculated by subtracting adjacent-normal expression from tumor expression:

$$d_{gi} = Y_{g,T_i} - Y_{g,N_i},$$

where Y_{g,T_i} and Y_{g,N_i} represent transformed expression for feature g in the tumor and adjacent-normal tissues of patient i, respectively.

The mean paired difference was calculated as:

$$\bar{d}_g = \frac{1}{n}\sum_{i=1}^{n} d_{gi}, \qquad n = 6.$$

Positive estimates indicated higher mean transformed expression in tumor tissue, whereas negative estimates indicated lower mean transformed expression.

A two-sided paired Student's *t* test assessed whether the mean within-patient difference was zero:

$$t_g = \frac{\bar{d}_g}{s_{d,g}/\sqrt{n}},$$

where $s_{d,g}$ represents the standard deviation of the paired differences. The test used five degrees of freedom. Gene-specific 95% confidence intervals were calculated as:

$$\bar{d}_g \pm t_{0.975,5}\frac{s_{d,g}}{\sqrt{6}}.$$

Benjamini–Hochberg adjustment was applied across all 23,813 tested features. A genome-wide adjusted *p-value* below 0.05 defined statistical significance in the primary analysis. Unadjusted *p-values* were reported separately and were not used to establish genome-wide significance.

The paired *t* procedure assumes approximately normally distributed within-patient differences. Given the limited number of pairs and the use of transformed-expression tests rather than a dispersion-moderated count model, the primary analysis was interpreted as exploratory. Confidence intervals were gene-specific and did not provide simultaneous genome-wide coverage.

Inflammatory and Immune-Marker Expression Signatures

Six investigator-defined expression signatures were examined to summarize transcript patterns associated with inflammatory signaling and immune-cell markers.

Expression signature	Constituent genes
TNF/NF-κB-associated	NFKB1, NFKBIA, RELA, TNFAIP3, TRAF2
IL-6/JAK/STAT3-associated	IL6, IL6R, JAK1, JAK2, STAT3, SOCS3
TGF-β-associated	TGFB1, TGFBR1, TGFBR2, SMAD2, SMAD3, SMAD7
Chemokine	CCL2, CXCL8, CXCL9, CXCL10, CXCL12
Macrophage-marker	CD68, CSF1R, CD163, MRC1, LST1
T-cell-marker	CD3D, CD3E, CD2, CD8A, CD4, TRAC

The available signature outputs summarized standardized transcript measurements and reported mean tumor-minus-adjacent-normal score differences. Multiple-testing adjustment was applied separately across the six signature comparisons.

The underlying standardization procedure, aggregation formula, and signature-level statistical test were not documented in the available analysis record. Therefore, these outputs were retained as supplementary exploratory findings rather than as fully reproducible confirmatory analyses. The signatures were interpreted as bulk-tissue expression summaries, not as direct measurements of immune-cell proportions, histological infiltration, or pathway activation.

Negative-Binomial Sensitivity Analyses

The available sensitivity analysis used a paired negative-binomial model incorporating tissue status, patient effects, and library-size offsets. Patient effects accounted for the matched sampling structure, while the offsets accounted for differences in library totals. A common pooled dispersion estimate of 0.142 was reported.

Tissue-associated estimates were expressed as tumor-versus-adjacent-normal $\log_2$ fold changes with 95% confidence intervals. Benjamini–Hochberg adjustment was applied across the seven evaluable genes. These panel-adjusted results were distinguished from the genome-wide adjusted results of the primary analysis.

An additional gene-specific NB2 sensitivity fit was reported separately. Non-estimable genes and convergence failures were retained in the reporting rather than being assigned numerical results. Due to the unavailability of complete fitting specifications and software versions, these analyses were treated as supporting sensitivity outputs. They were not interpreted as independent replication or as replacements for the primary paired analysis.

External Microarray Comparison

The external GSE42568 analysis compared available probe-level expression measurements between breast cancer and normal breast tissue groups. Two-sided Welch's *t* tests were used to accommodate unequal group sizes and potentially unequal variances. Benjamini–Hochberg adjustment was applied across 54,675 complete probe records.

The analysis remained at probe level because probe-to-gene mapping and gene-level aggregation were not completed. Therefore, significant probe counts were not interpreted as counts of distinct genes or as independent confirmation of the selected inflammatory panel.

The preprocessing and transformation scale of the analyzed microarray matrix were not fully documented in the available record. As a result, the external findings were treated as supplementary descriptive comparisons rather than as definitive cross-platform validation.

Assessment of Dataset Overlap

Potential overlap between GSE45827 and GSE65216 was examined using available sample annotations and expression-profile comparisons across 754 inspected common probes. Correlation coefficients and absolute expression differences were used to assess similarity between closely matching profiles.

Near-identical profiles were considered evidence of potential shared or closely related measurements. Different repository identifiers alone were not regarded as sufficient evidence of independent biological donors. The datasets were therefore not pooled or counted as separate independent replications.

Ethical Considerations

The study used publicly accessible transcriptomic datasets and involved no direct participant contact, new biological sampling, or experimental intervention. Analyses were restricted to repository-provided expression measurements and associated sample annotations, and no attempt was made to identify participants.

Results

The analysis examined six matched pairs of breast cancer and adjacent-normal tissues from the GSE133998 dataset. Of the 61,155 initial expression features, 23,813 satisfied the filtering criteria and were included in the primary paired differential-expression analysis. Five features met the genome-wide false-discovery-rate threshold (*FDR* < 0.05); however, none of the seven evaluable inflammation-related genes achieved genome-wide statistical significance. The selected genes demonstrated heterogeneous expression differences, with *TGFB1* exhibiting the largest positive mean paired difference. None of the six inflammatory and immune-marker expression signatures remained significant after adjustment for multiple testing. In the targeted negative-binomial sensitivity analysis, *TGFB1* achieved panel-adjusted significance ($p = 0.0096$). Supplementary microarray analyses revealed additional probe-level differences but did not provide independent gene-level replication. The results are organized by dataset characteristics, individual gene expression, expression signatures, sensitivity analyses, and external comparisons.

Table 1. Characteristics of the tissue dataset and expression-filtering results

Characteristic	Number
Patients with matched tissue samples	6
Breast cancer tissue samples	6
Adjacent-normal tissue samples	6
Initial expression features	61,155
Features excluded because of incomplete measurements or duplicated identifiers	6
Complete, uniquely identified features	61,149
Features retained after expression filtering	23,813
Features meeting genome-wide FDR < 0.05	5

Expression filtering required at least 10 counts in six or more samples. The biological sample size was six patients, each contributing one tumor and one adjacent-normal tissue sample. FDR, false discovery rate.

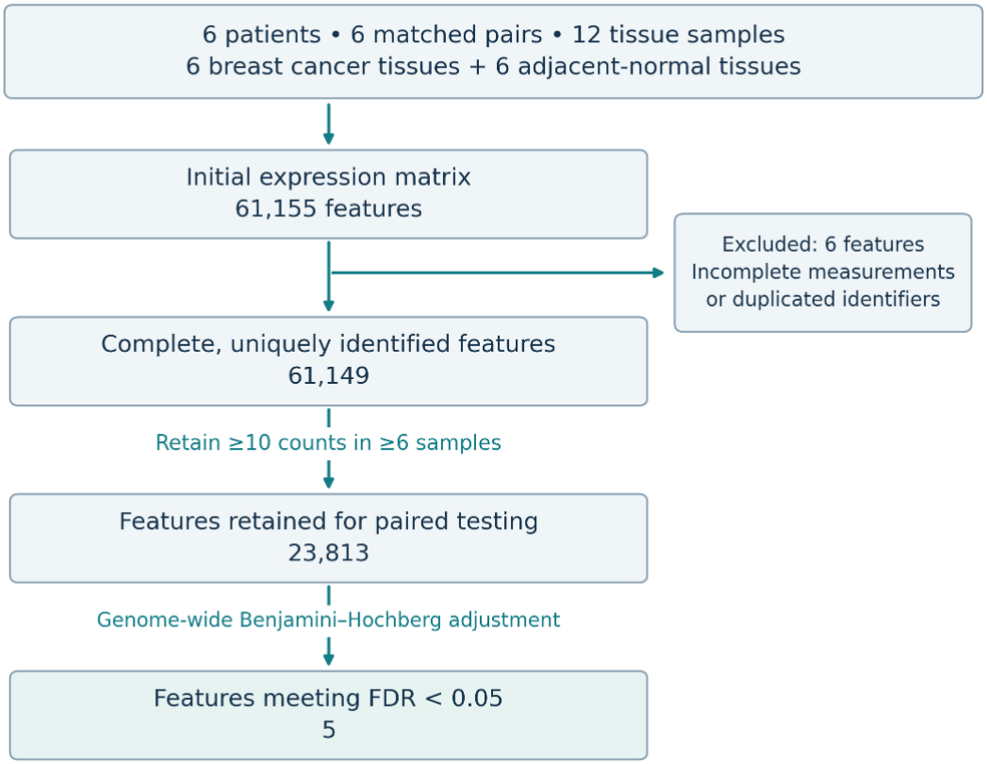


Figure 1 Tissue sampling, expression filtering, and genome-wide differential-expression testing.

Expression of inflammation-related genes

Seven of the eight selected inflammation-related genes were evaluable. TNF had insufficient informative counts for estimation. Five genes—TGFB1, CCL2, IL1B, CXCL8, and NFKB1—showed positive mean tumor-minus-adjacent-normal differences, whereas IL6 and STAT3 showed negative differences.

TGFB1 had the largest positive mean difference, at 0.992 $\log_2$(CPM + 1) units (95% CI, 0.344–1.640), followed by CCL2 at 0.673 (95% CI, −0.383–1.729) and IL1B at 0.574 (95% CI, 0.055–1.093). The estimated differences for CXCL8 and NFKB1 were 0.209 and 0.169, respectively. IL6 and STAT3 had estimated differences of −0.126 and −0.170, respectively.

TGFB1, IL1B, and NFKB1 had unadjusted *p-values* below 0.05. However, none remained statistically significant after genome-wide multiple-testing adjustment; adjusted *p-values* ranged from 0.139 to 0.863 (Table 2).

Table 2. Paired tumor-minus-adjacent-normal expression differences for selected inflammation-related genes

Gene	Mean paired difference	95% confidence interval	Unadjusted *p*	Genome-wide adjusted *p*
IL6	−0.126	−1.233 to 0.980	0.781	0.863
TNF	Not estimable	—	—	—
IL1B	0.574	0.055 to 1.093	0.0361	0.179
CXCL8	0.209	−1.611 to 2.029	0.780	0.862

Gene	Mean paired difference	95% confidence interval	Unadjusted *p*	Genome-wide adjusted *p*
CCL2	0.673	−0.383 to 1.729	0.162	0.336
STAT3	−0.170	−0.566 to 0.225	0.319	0.498
NFKB1	0.169	0.000585 to 0.337	0.0495	0.198
TGFB1	0.992	0.344 to 1.640	0.0110	0.139

Estimates represent mean within-patient differences in $\log_2$(counts per million + 1) across six matched pairs. Positive values indicate higher mean transformed expression in tumor tissue. Confidence intervals are unadjusted, gene-specific intervals. Benjamini–Hochberg adjustment covered 23,813 tested features. These estimates are differences in transformed expression, rather than conventional $\log_2$ fold changes.

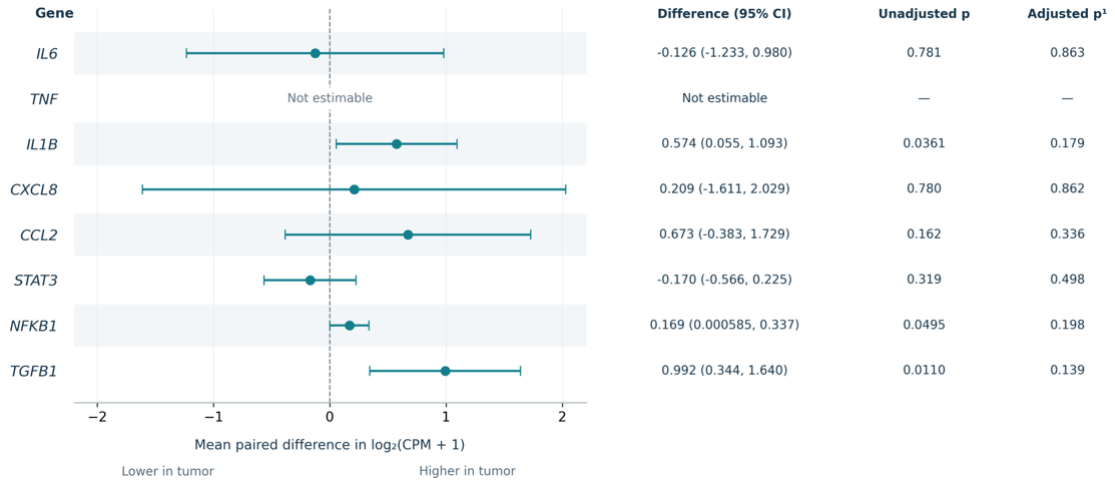


Figure 2 Paired tumor-minus-adjacent-normal expression differences with unadjusted 95% confidence intervals across six patients

Inflammatory and immune-marker expression signatures

All six expression signatures had positive mean tumor-minus-adjacent-normal score differences. The T-cell-marker signature showed the largest numerical difference (1.106), followed by the TNF/NF-κB-associated signature (0.528) and the chemokine signature (0.450). The corresponding differences for the TGF-β-associated, macrophage-marker, and IL-6/JAK/STAT3-associated signatures were 0.225, 0.193, and 0.058, respectively.

None of the signature comparisons met the FDR threshold of 0.05 after adjustment across the six tests. The lowest adjusted *p-value* was observed for the T-cell-marker signature (0.109), followed by the TNF/NF-κB-associated and chemokine signatures (both 0.133; Table 3).

Table 3. Paired differences in inflammatory and immune-marker expression signatures

Expression signature	Number of genes	Mean paired score difference	Adjusted *p*
TNF/NF-κB-associated	5	0.528	0.133

Expression signature	Number of genes	Mean paired score difference	Adjusted *p*
IL-6/JAK/STAT3-associated	6	0.058	0.883
TGF-β-associated	6	0.225	0.473
Chemokine	5	0.450	0.133
Macrophage-marker	5	0.193	0.715
T-cell-marker	6	1.106	0.109

Values represent mean within-patient differences in standardized expression scores, calculated as tumor minus adjacent-normal tissue. Benjamini–Hochberg adjustment covered six signature comparisons. These investigator-defined signatures summarize bulk-tissue transcript patterns; they do not directly measure immune-cell proportions or pathway activation.

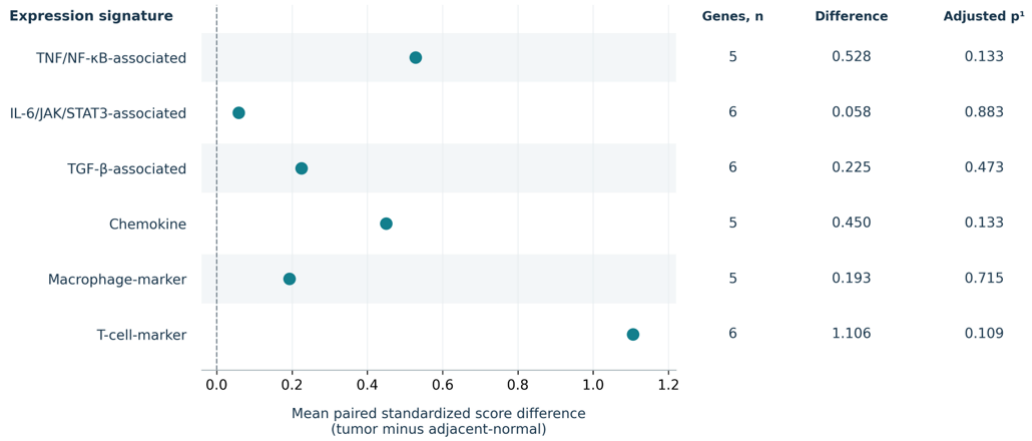


Figure 3 Paired differences in inflammatory and immune-marker expression signatures

Sensitivity analysis of individual genes

A targeted paired negative-binomial analysis using pooled dispersion produced effect estimates with the same direction as the primary analysis for all seven evaluable genes. TGFB1 showed a positive $\log_2$ fold change of 1.007 (95% CI, 0.391–1.624) and met the panel-level FDR threshold (adjusted $p = 0.0096$). IL1B and CCL2 each had a panel-adjusted *p-value* of 0.0655. The remaining genes did not meet the panel-level significance threshold (Table 4).

Table 4. Targeted paired negative binomial sensitivity analysis

Gene	$\log_2$ fold change	95% confidence interval	Panel-adjusted *p*
IL6	−0.280	−0.964 to 0.404	0.5907
IL1B	0.742	0.103 to 1.381	0.0655
CXCL8	0.592	−0.060 to 1.244	0.1312
CCL2	0.695	0.075 to 1.315	0.0655

Gene	Log_2 fold change	95% confidence interval	Panel-adjusted *p*
STAT3	−0.169	−0.784 to 0.446	0.5907
NFKB1	0.171	−0.446 to 0.787	0.5907
TGFB1	1.007	0.391 to 1.624	0.0096

The model included patient effects and library-size offsets, with a common pooled dispersion estimate of 0.142. Multiple-testing adjustment covered seven evaluable genes.

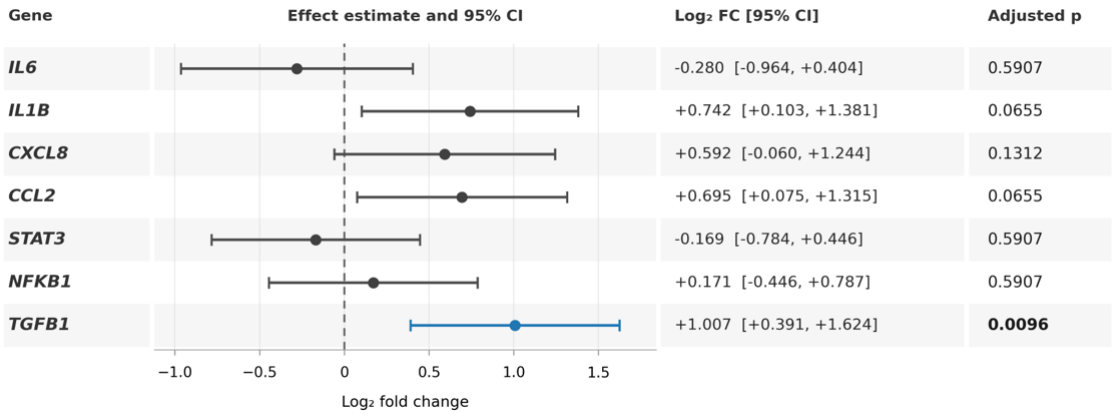


Figure 4 Forest plot of log_2 fold changes with 95% confidence intervals and panel-adjusted p-values from the targeted paired negative-binomial sensitivity analysis.

A separate gene-specific NB2 sensitivity fit did not converge for STAT3, while TNF remained non-estimable. Accordingly, the sensitivity analyses did not provide a complete, consistently estimable comparison across all eight selected genes.

External expression comparison

The external GSE42568 microarray dataset comprised 104 breast cancer and 17 normal breast tissue arrays. Of 54,675 complete probe records evaluated using unpaired comparisons, 18,028 met FDR < 0.05. These results represented probe-level tumor–normal differences. Gene-level replication of the selected inflammatory panel was not established because probe-to-gene mapping and panel-specific comparisons had not been completed.

An overlap assessment of GSE45827 and GSE65216 identified near-identical expression profiles within a subset of 754 common probes. The closest matches had a median correlation of approximately 1.000 and a median absolute difference of approximately 0.000003. These datasets were therefore not treated as independent replication cohorts.

The primary paired analysis identified differences in the direction and magnitude of inflammatory transcript expression, but no selected gene or expression signature met its corresponding multiple-testing threshold. TGFB1 met the panel-level threshold only in the pooled-dispersion sensitivity analysis.

Discussion

This exploratory analysis evaluated inflammation-related gene expression in six matched breast cancer and adjacent-normal tissue pairs. The principal finding was that none of the seven evaluable inflammatory genes met the genome-wide false-discovery-rate threshold, although TGFB1, IL1B, and NFKB1 had

unadjusted *p* values below 0.05. Five selected genes exhibited positive tumor-minus-adjacent-normal estimates, whereas IL6 and STAT3 showed negative estimates. None of the six inflammatory or immune-marker signatures met its respective multiple-testing threshold. These results indicate uncertainty regarding tissue-associated differences in the selected panel, rather than providing evidence for the absence of inflammatory biology. This distinction aligns with evidence that cancer-associated inflammation involves interacting, context-dependent processes instead of a uniformly elevated transcriptional response (Grivennikov et al., 2010).

TGFB1 exhibited the largest positive primary estimate among the selected genes, with a mean paired difference of 0.992 $\log_2$(CPM + 1) units. Its genome-wide adjusted *p-value* was 0.139, while the targeted pooled-dispersion negative-binomial analysis yielded a $\log_2$ fold change of 1.007 and a panel-adjusted *p-value* of 0.0096. The consistent positive direction across these analyses suggests that TGFB1 is a reasonable candidate for further investigation; however, the sensitivity result does not establish a genome-wide significant finding. Biologically, TGF-β is involved in immune regulation and tumor–stroma interactions, with effects dependent on cellular and disease context (Batlle & Massagué, 2019). Therefore, the present transcript-level estimate cannot confirm increased active TGF-β signaling, immune suppression, or a specific tumor-promoting mechanism.

IL1B and NFKB1 also demonstrated positive estimates with unadjusted *p-values* below 0.05, but their genome-wide adjusted values were 0.179 and 0.198, respectively. In contrast, IL6 and STAT3 exhibited negative estimates with confidence intervals that included zero. These differences do not provide evidence for coordinated activation or suppression of an inflammatory pathway. Signaling networks involving IL-6, STAT3, and NF-κB are regulated through mechanisms beyond messenger RNA abundance, including protein activation and cellular context (Yu et al., 2009). Therefore, lower STAT3 transcript expression cannot be equated with reduced STAT3 activity, and higher NFKB1 expression alone does not demonstrate NF-κB activation.

The chemokine results also require cautious interpretation. CCL2 and CXCL8 exhibited positive mean paired differences, but neither reached the genome-wide significance threshold. The particularly broad confidence interval for CXCL8, ranging from −1.611 to 2.029, indicates substantial uncertainty regarding both the magnitude and direction of the underlying difference. Chemokine function depends on ligand availability, receptor expression, and regulatory mechanisms beyond transcription (Cambier et al., 2023). Therefore, the observed estimates provide exploratory insights into chemokine transcripts but do not establish increased leukocyte recruitment. TNF was non-estimable and should be distinguished from a gene with an estimated null association; its analytical status does not indicate biological absence.

At the signature level, all six mean differences were positive, with the largest numerical difference observed for the T-cell-marker signature. However, its adjusted *p-value* of 0.109, along with nonsignificant results for the remaining signatures, did not provide statistically supported evidence of higher signature scores in tumor tissue. Additionally, these investigator-defined scores are not validated estimates of cellular abundance. Newman et al. (2015) demonstrated that expression-based estimation of cell subsets requires a defined deconvolution framework and validation against reference measurements. Therefore, the present marker summaries cannot establish greater T-cell or macrophage infiltration. Incomplete documentation of the scoring procedure further limits reproducibility and supports retaining these findings as supplementary and exploratory.

Cellular heterogeneity may explain variation in the observed transcript patterns, although this factor was not directly assessed in the present study. Single-cell and spatial analyses have identified diverse malignant, immune, and stromal populations within breast cancers (Wu et al., 2021). Changes in bulk expression may therefore reflect differences in cellular proportions, transcription within individual cell populations, or both. The current analysis could not distinguish among these possibilities. More broadly, the immune microenvironment is defined by the location, functional state, and abundance of its cellular components (Binnewies et al., 2018). These aspects remain beyond the scope of the present bulk-tissue comparison.

The choice of adjacent-normal tissue also shapes the interpretation. Pairing reduced between-patient variation, but the comparator was not equivalent to tissue from healthy individuals. Aran et al. (2017) showed that adjacent-normal tissues can exhibit transcriptional characteristics distinct from both healthy and tumor tissues. Local inflammatory or stromal responses in adjacent tissue could therefore influence

the estimated contrasts, although their contribution cannot be quantified from the current data. The results should be understood as within-patient differences between two tissue environments, rather than comprehensive differences between breast cancer and health.

The sensitivity analyses highlight the importance of distinguishing model assumptions from the scope of multiple testing. The primary analysis adjusted across 23,813 features, whereas the targeted negative-binomial model adjusted across seven genes. Consequently, the TGFB1 result addressed a different testing family and employed a different statistical model. The individual contributions of these differences cannot be separated based on the reported summaries. RNA-sequencing methods such as DESeq2 explicitly address dispersion estimation and stabilization of expression effects, emphasizing the importance of modeling uncertainty in small datasets (Love et al., 2014). In this study, pooled dispersion, incomplete fitting documentation, and the separate NB2 convergence failure for STAT3 limit the strength of the sensitivity evidence.

The external microarray comparison did not resolve these uncertainties. Although 18,028 of 54,675 probes met the adjusted significance threshold in GSE42568, this aggregate result did not establish replication of the selected inflammatory genes. Probe-to-gene mapping and panel-specific comparisons were not completed, and the microarray groups were unpaired. Potential overlap between GSE45827 and GSE65216 further precluded their use as independent confirmations. Therefore, this study provides an exploratory paired analysis with supplementary external observations, rather than independently validated inflammatory biomarkers.

The principal strengths of this study include preservation of patient matching, explicit reporting of uncertainty, and separation of genome-wide from panel-level adjustment. Key limitations include the small six-patient sample, limited clinical characterization, bulk-tissue measurement, incomplete analytical documentation, and absence of completed gene-level replication. No diagnostic, prognostic, or treatment-selection claims are supported. Given the context-dependent role of the breast cancer immune microenvironment described by Harris et al. (2024), future research should examine larger, clinically characterized cohorts and integrate reproducible transcriptomic analysis with cell-resolved or protein-level measurements. Such studies could determine whether the observed TGFB1 and other directional estimates represent reproducible tissue differences and clarify their cellular and functional significance.

Conclusion

This exploratory analysis of six matched breast cancer and adjacent-normal tissue pairs identified variable expression patterns among selected inflammation-related genes. None of the seven evaluable genes or six expression signatures reached the multiple-testing threshold in the primary analyses. TGFB1 demonstrated the largest positive gene-level estimate and met the panel-adjusted threshold in a sensitivity analysis, indicating the need for further investigation. These results do not provide evidence for increased immune-cell infiltration or activation of inflammatory pathways. Larger, independently validated studies that incorporate cell-resolved and protein-level measurements are required to clarify the biological significance of these expression patterns within the breast tumor microenvironment.

REFERENCES

Balkwill, F. R., Capasso, M., & Hagemann, T. (2012). The tumor microenvironment at a glance. *Journal of Cell Science, 125*(23), 5591–5596. https://doi.org/10.1242/jcs.116392

Chen, D. S., & Mellman, I. (2017). Elements of cancer immunity and the cancer–immune set point. *Nature, 541*(7637), 321–330. https://doi.org/10.1038/nature21349

Coussens, L. M., & Werb, Z. (2002). Inflammation and cancer. *Nature, 420*(6917), 860–867. https://doi.org/10.1038/nature01322

Curtius, K., Wright, N. A., & Graham, T. A. (2018). An evolutionary perspective on field cancerization. *Nature Reviews Cancer, 18*(1), 19–32. https://doi.org/10.1038/nrc.2017.102

Fridman, W. H., Pagès, F., Sautès-Fridman, C., & Galon, J. (2012). The immune contexture in human tumours: Impact on clinical outcome. *Nature Reviews Cancer, 12*(4), 298–306. https://doi.org/10.1038/nrc3245

Greten, F. R., & Grivennikov, S. I. (2019). Inflammation and cancer: Triggers, mechanisms, and consequences. *Immunity, 51*(1), 27–41. https://doi.org/10.1016/j.immuni.2019.06.025

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nature Reviews Cancer, 16*(9), 582–598. https://doi.org/10.1038/nrc.2016.73

Mantovani, A., Allavena, P., Sica, A., & Balkwill, F. (2008). Cancer-related inflammation. *Nature, 454*(7203), 436–444. https://doi.org/10.1038/nature07205

Massagué, J. (2012). TGFβ signalling in context. *Nature Reviews Molecular Cell Biology, 13*(10), 616–630. https://doi.org/10.1038/nrm3434

Polyak, K. (2011). Heterogeneity in breast cancer. *The Journal of Clinical Investigation, 121*(10), 3786–3788. https://doi.org/10.1172/JCI60534

Quail, D. F., & Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine, 19*(11), 1423–1437. https://doi.org/10.1038/nm.3394

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), Article e47. https://doi.org/10.1093/nar/gkv007

Waks, A. G., & Winer, E. P. (2019). Breast cancer treatment: A review. *JAMA, 321*(3), 288–300. https://doi.org/10.1001/jama.2018.19323

Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., Schnitt, S., Sellers, W. R., & Polyak, K. (2004). Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell, 6*(1), 17–32. https://doi.org/10.1016/j.ccr.2004.06.010

DeNardo, D. G., & Ruffell, B. (2019). Macrophages as regulators of tumour immunity and immunotherapy. *Nature Reviews Immunology, 19*(6), 369–382. https://doi.org/10.1038/s41577-019-0127-6

Erez, N., Truitt, M., Olson, P., Arron, S. T., & Hanahan, D. (2010). Cancer-associated fibroblasts are activated in incipient neoplasia to orchestrate tumor-promoting inflammation in an NF-κB-dependent manner. *Cancer Cell, 17*(2), 135–147. https://doi.org/10.1016/j.ccr.2009.12.041

Finak, G., Bertos, N., Pepin, F., Sadekova, S., Souleimanova, M., Zhao, H., Chen, H., Omeroglu, G., Meterissian, S., Omeroglu, A., Hallett, M., & Park, M. (2008). Stromal gene expression predicts clinical outcome in breast cancer. *Nature Medicine, 14*(5), 518–527. https://doi.org/10.1038/nm1764

Gonzalez, H., Hagerling, C., & Werb, Z. (2018). Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes & Development, 32*(19–20), 1267–1284. https://doi.org/10.1101/gad.314617.118

Hanahan, D., & Coussens, L. M. (2012). Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell, 21*(3), 309–322. https://doi.org/10.1016/j.ccr.2012.02.022

Hinshaw, D. C., & Shevde, L. A. (2019). The tumor microenvironment innately modulates cancer progression. *Cancer Research, 79*(18), 4557–4566. https://doi.org/10.1158/0008-5472.CAN-18-3962

Johnson, D. E., O'Keefe, R. A., & Grandis, J. R. (2018). Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nature Reviews Clinical Oncology, 15*(4), 234–248. https://doi.org/10.1038/nrclinonc.2018.8

Joyce, J. A., & Pollard, J. W. (2009). Microenvironmental regulation of metastasis. *Nature Reviews Cancer, 9*(4), 239–252. https://doi.org/10.1038/nrc2618

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology, 15*, Article R29. https://doi.org/10.1186/gb-2014-15-2-r29

Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., & Freeman, T. C. (2018). Immune cell gene signatures for profiling the microenvironment of solid tumors. *Cancer Immunology Research, 6*(11), 1388–1400. https://doi.org/10.1158/2326-6066.CIR-18-0342

Noy, R., & Pollard, J. W. (2014). Tumor-associated macrophages: From mechanisms to therapy. *Immunity, 41*(1), 49–61. https://doi.org/10.1016/j.immuni.2014.06.010

Ruffell, B., & Coussens, L. M. (2015). Macrophages and therapeutic resistance in cancer. *Cancer Cell, 27*(4), 462–472. https://doi.org/10.1016/j.ccell.2015.02.015

Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirota, M., & Butte, A. J. (2017). Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nature Communications, 8*, Article 1077. https://doi.org/10.1038/s41467-017-01027-z

Batlle, E., & Massagué, J. (2019). Transforming growth factor-β signaling in immunity and cancer. *Immunity, 50*(4), 924–940. https://doi.org/10.1016/j.immuni.2019.03.024

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., Vonderheide, R. H., Pittet, M. J., Jain, R. K., Zou, W., Howcroft, T. K., Woodhouse, E. C., Weinberg, R. A., & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine, 24*(5), 541–550. https://doi.org/10.1038/s41591-018-0014-x

Cambier, S., Gouwy, M., & Proost, P. (2023). The chemokines CXCL8 and CXCL12: Molecular and functional properties, role in disease and efforts towards pharmacological intervention. *Cellular & Molecular Immunology, 20*, 217–251. https://doi.org/10.1038/s41423-023-00974-6

Grivennikov, S. I., Greten, F. R., & Karin, M. (2010). Immunity, inflammation, and cancer. *Cell, 140*(6), 883–899. https://doi.org/10.1016/j.cell.2010.01.025

Harris, M. A., Savas, P., Virassamy, B., O'Malley, M. M. R., Kay, J., Mueller, S. N., Mackay, L. K., Salgado, R., & Loi, S. (2024). Towards targeting the breast cancer immune microenvironment. *Nature Reviews Cancer, 24*(8), 554–577. https://doi.org/10.1038/s41568-024-00714-6

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*, Article 550. https://doi.org/10.1186/s13059-014-0550-8

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., Hoang, C. D., Diehn, M., & Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods, 12*(5), 453–457. https://doi.org/10.1038/nmeth.3337

Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J. R., Bartonicek, N., Wang, T., Larsson, L., Kaczorowski, D., Weisenfeld, N. I., Uytingco, C. R., Chew, J. G., Bent, Z. W., Chan, C.-L., Gnanasambandapillai, V., . . . Swarbrick, A. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics, 53*(9), 1334–1347. https://doi.org/10.1038/s41588-021-00911-1

Yu, H., Pardoll, D., & Jove, R. (2009). STATs in cancer inflammation and immunity: A leading role for STAT3. *Nature Reviews Cancer, 9*(11), 798–809. https://doi.org/10.1038/nrc2734