

RESEARCH ARTICLE

Immunophenotype-associated gene signature in ductal breast tumors varies by receptor subtype, but the expression of individual signature genes remains consistent

Michael Behring^{1,2}  | Yuanfan Ye¹ | Amr Elkholy² | Prachi Bajpai² | Sumit Agarwal² | Hyung-Gyoon Kim² | Akinyemi I. Ojesina^{1,3}  | Howard W Wiener¹ | Upender Manne^{2,3}  | Sadeep Shrestha¹ | Ana I. Vazquez^{4,5}

¹Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA

²Department of Pathology and Surgery, University of Alabama at Birmingham, Birmingham, AL, USA

³Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL, USA

⁴Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA

⁵Institute for Quantitative Health Science & Engineering, East Lansing, MI, USA

Correspondence

Michael Behring and Ana I. Vazquez, Department of Epidemiology, University of Alabama at Birmingham; 1665 University Boulevard Birmingham, AL 35294, USA.

Emails: avazquez@msu.edu (M. B.) and michaelbehring@uabmc.edu (A. I. V.)

Funding information

Behring NCI T32 Cancer Prevention and Control Training Program (NCI T32 CA047888).

Abstract

Background: In silico deconvolution of invasive immune cell infiltration in bulk breast tumors helps characterize immunophenotype, expands treatment options, and influences survival endpoints. In this study, we identify the differential expression (DE) of the LM22 signature to classify immune-rich and -poor breast tumors and evaluate immune infiltration by receptor subtype and lymph node metastasis.

Methods: Using publicly available data, we applied the CIBERSORT algorithm to estimate immune cells infiltrating the tumor into immune-rich and immune-poor groups. We then tested the association of receptor subtype and nodal status with immune-rich/poor phenotype. We used DE to test individual signature genes and over-representation analysis for related pathways.

Results: *CCL19* and *CXCL9* expression differed between rich/poor signature groups regardless of subtype. Overexpression of *CHI3L2* and *FES* was observed in triple negative breast cancers (TNBCs) relative to other subtypes in immune-rich tumors. Non-signature genes, *LYZ*, *CIQB*, *CORO1A*, *EVI2B*, *GBP1*, *PSMB9*, and *CD52* were consistently overexpressed in immune-rich tumors, and *SCUBE2* and *GRIA2* were associated with immune-poor tumors. Immune-rich tumors had significant upregulation of genes/pathways while none were identified in immune-poor tumors.

Conclusions: Overall, the proportion of immune-rich/poor tumors differed by subtype; however, a subset of 10 LM22 genes that marked immune-rich status remained the same across subtype. Non-LM22 genes differentially expressed between the phenotypes suggest that the biologic processes responsible for immune-poor phenotype are not yet well characterized.

KEYWORDS

breast cancer, microenvironment, transcriptomics, tumor-infiltrating immune cells

Sadeep Shrestha and Ana I. Vazquez shared senior authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The immune system has a strong influence on both tumorigenesis and clinical outcomes in all cancers.¹ Recent discoveries have led to the development of new immune-based prognostic markers and innovative therapy. However, host antitumor immune response is a complex process, involving multiple cell types and signaling mechanisms that inhibit or even propagate tumor growth. The interaction between tumor cells and the immune system relies on a conceptual framework of tumor immunoediting, a sequence of tumor suppression, homeostasis, and subversion/evasion.² A vital, but not fully understood feature in this process is the presence of tumor infiltrating immune cells.³ In breast cancer, immune cell infiltrates can constitute as much as half of the tumor mass in patients, and be comprised of many different types of immune cells.⁴ The focus of studies in breast patients has been limited to the prognostic effect of tumor invasive lymphocytes (TIL) cell types,⁵ although there is also increasing evidence to support the relevance of other immune cell types in breast cancer.^{6,7}

Breast cancer is a heterogeneous disease, made up of 4 distinct receptor subtypes Luminal A, Luminal B, human epidermal growth factor receptor 2 positive (Her2+), and triple negative (TNBC). Clinical trials evaluating adjuvant therapy and immune invasiveness found that the presence of TILs in TNBC and Her2+ receptor subtypes is associated with improved treatment response and survival.⁵ In previous research, both Her2+ and TNBC tumors had higher TIL and were more immune-rich than hormone receptor-positive (Luminal A and Luminal B) tumors.⁸ However, the picture of how tumor-invasive immune cells and receptor subtypes are related in breast cancer remains incomplete. Reviews summarizing the association of TIL infiltration with survival in breast tumors show that outcomes depend on how immune cells are measured, which cell types are included, and how they modify the pathogenesis in different subtypes.^{9,10}

There is no gold standard for the measure of lymphocytes and other tumor invasive immune cell types. In clinical settings, the recommended methodology for classifying tumors into immune-poor/rich remains hematoxylin and eosin (H&E) staining of tumor sections.¹¹ Noninfiltrative, immune-poor tumors have been identified as a stronger predictor of survival than any invasive cell type in estrogen receptor negative (ER-) tumors.¹² Thus, the ability to reliably identify and better understand the functional states of immune-poor tumors is of high translational relevance for a large subset of patients, yet it remains a missing gap in the current knowledge of breast tumor immune microenvironment.

The goal of this research study is to use three independent sets of data to create a validated, standard signature of tumor invasive immune cells genes. We aim to better understand how the presence/lack of immune cell infiltration

in a primary tumor relates to distinct molecular/pathologic features of disease, specifically receptor subtype, and lymph node metastasis. Rather than focus on the relative abundance of each invasive immune cell type in tumors, we use a standard gene expression array-based signature of 547 genes to create a binary (immune-rich or immune-poor) variable of immune invasion status across three large breast cancer cohorts. Among patients with immune-rich tumors, we further identify the specific signature genes on each subtype. Lastly, we apply pathway analysis to identify important molecular mechanisms behind both immune-rich and immune-poor tumor groups.

2 | MATERIALS AND METHODS

2.1 | Study patients

We evaluated patient data from three independent publicly available data sources, final data for analysis includes: 650 patients from The Cancer Genome Atlas (TCGA),¹³ 772 patients from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)¹⁴ and 947 patients from Gene Expression Omnibus (GEO) compiled datasets¹⁵ (Table 1). Lymph node metastasis at diagnosis was used as a binary endpoint (yes/no), and receptor subtype (based on hormone receptor status) was included as the only variable common to all sets of data. In order to maximize inclusion of patients with missing progesterone (PR) or estrogen receptor (ER) data, breast subtypes of Luminal A and Luminal B were combined into a single, Luminal variable, thus, the receptor subtype used across datasets are: Luminal, triple negative receptors (TNBC) and Her2+ (HER). Tumor histologic type was an inclusion criterion and breast tissue not classified as invasive ductal carcinoma (IDC) was excluded; (in TCGA 236 samples were excluded and 429 in METABRIC). In addition, any patient with missing stage information was excluded to avoid misclassification of ductal carcinoma in-situ (DCIS), noninvasive tumors (METABRIC only $n = 774$). GEO data had no information regarding histologic tissue type, yet all patients were confirmed non-DCIS using size and grading information.

2.2 | RNA expression data

METABRIC used the Illumina HT-12v3 platform in gene expression analysis.¹⁴ In TCGA, mRNA expression counts were derived from the TCGA Level 3 RNAseqV2 expression data collected using Illumina HiSeq 2000 platform.¹³ The GEO study data are a curated aggregate of six Affymetrix Human Genome HG-U133A array datasets.¹⁵ Gene

TABLE 1 Association between tumor immune cell signature status NM and receptor subtype for across datasets

Immune-signature	METABRIC			GEO			TCGA		
	Poor (N=639)	Rich (N=133)	P	Poor (N=482)	Rich (N=465)	P	Poor (N=327)	Rich (N=323)	P
Nodal Metastasis			<i>ns</i>			<i>ns</i>			<i>ns</i>
negative	328 (51.3%)	61 (45.9%)		323 (68.7%)	297 (65.0%)		136 (41.6%)	157 (48.6%)	
positive	311 (48.7%)	72 (54.1%)		147 (31.3%)	160 (35.0%)		191 (58.4%)	166 (51.4%)	
Receptor subtype			*			*			*
missing	2 (0.3%)	0 (0.0%)		–	–		17 (5.2%)	35 (10.8%)	
HER2	83 (13.0%)	19 (14.3%)		47 (12.6%)	74 (19.6%)		10 (3.1%)	20 (6.2%)	
Luminal	445 (69.6%)	60 (45.1%)		272 (72.9%)	189 (50.0%)		278 (85.0%)	210 (65.0%)	
TNBC	109 (17.1%)	54 (40.6%)		54 (14.5%)	115 (30.4%)		22 (6.7%)	58(18.0%)	

Note: P: *p*-value - *Indicates $p < 0.001$, *ns* indicates not significant at the 0.05 significance threshold

Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) and Array Express (<http://www.ebi.ac.uk/arrayexpress>) accession numbers are: E-TABM-158,¹⁶ GSE6532,¹⁷ GSE3494,¹⁸ GSE1456,¹⁹ GSE7390,²⁰ and GSE2603.²¹ All RNA intensities and read counts were either available at, or converted to gene level measurements. Based on the leukocyte specific gene signature matrix (LM22) immune cell type data, 515 of 547 immune cell specific genes were present in all three sets of patient data. The expression values for each dataset were independently standardized and combined results were evaluated for consistency using standard cutpoints and filtering conditions.

2.3 | Primary tumor immune-rich/immune-poor cell activity

We estimated immune cells abundance in the tumor sample, derived from the LM22 signature employing the immune cell expression profiling using CIBERSORT algorithm.²² This LM22 signature is comprised of 547 genes, which represent unique expression values for 22 specific human leukocyte cell types. LM22 has been validated in a process involving variably pure leukocyte subsets as well as bulk tumors from multiple cancer types, including breast cancer.²² This approach has been shown to be effective in identifying non-tumor cells through comparison to flow sorted validation measures in follicular lymphoma.²³ In breast cancer, leukocyte cell type specific expression has been confirmed as an accurate measure of tumor immune infiltrates.²⁴

CIBERSORT, based on the LM22 gene signature, was used to estimate immune cells infiltrating the tumor.²² CIBERSORT's core algorithm employs a machine learning approach to derive infiltrating cell types from a patient's whole-transcriptome RNA tumor expression. Rather than quantifying the proportion of tumor infiltrating cell types,

we applied the first part of CIBERSORT's algorithm to categorize the overall immune-rich/immune-poor status of a tumor. Briefly, for each tumor sample, genes (the same number as LM22 signature genes) are randomly chosen from the full transcriptome. The correlation between random tumor RNA and signature RNA values was tested using a Pearson product-moment correlation. This process was repeated 1000 times per sample to establish a distribution of null correlation values. The Pearson's correlation coefficient for the real test of tumor RNA to signature RNA was then ranked among the 1000 random test statistics for a *p*-value. Samples under the $p = 0.05$ threshold were termed immune-rich.

2.4 | Association testing and gene set tests

Both METABRIC and TCGA had 0.3% and 8% missing subtype information, respectively. Differential expression analysis, for immune-rich/poor phenotype, was conducted with linear models using the R package limma.²⁵ Empirical Bayes²⁶ shrinkage was used to conduct t-statistics for each gene and multiple comparisons were adjusted with the Benjamini-Hochberg correction.²⁷ In TCGA, raw counts of RNA per-gene were compiled across the transcriptome of each patient and then assembled into a gene-by-patient matrix. Normalization factors for the raw data matrix were calculated, and a negative binomial model of differential counts was used in the edgeR package.²⁸ Over-representation analysis (OVA) for up or downregulated genes from DE models was performed using KEGG (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>) pathways and the limma "kegga" function.²⁵ The scope of OVA analysis is intended to validate the variable of immune-rich through any association with immune-related pathways, and as a discovery approach for immune-poor pathways.

TABLE 2 Differentially expressed LM22 genes between immune-rich and immune-poor tumors, for all patients, and receptor-stratified

Gene	Log Fold Change (Log FC)								
	All receptor subtypes			Luminal type only			TNBC type only		
	(M)	(G)	(T)	(M)	(G)	(T)	(M)	(G)	(T)
<i>CCL19</i>	2.09	1.77	1.24	2.08	1.54	1.01	2.02	1.36	1.60
<i>CCL5</i>	1.77	1.35	1.18	1.73	1.32	1.01	1.31	1.15	1.05
<i>CD2</i>	1.76	1.01	1.05	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
<i>CD3D</i>	1.88	1.00	1.22	1.80	1.00	1.10	<i>ns</i>	<i>ns</i>	<i>ns</i>
<i>CXCL10</i>	1.68	1.39	1.55	1.67	1.12	1.01	<i>ns</i>	<i>ns</i>	<i>ns</i>
<i>CXCL9</i>	2.29	3.10	1.83	2.31	2.70	1.76	1.50	2.69	1.24
<i>GZMA</i>	1.55	1.14	1.12	1.46	1.14	1.01	<i>ns</i>	<i>ns</i>	<i>ns</i>
<i>GZMK</i>	1.93	1.17	1.12	1.92	1.25	1.07	<i>ns</i>	<i>ns</i>	<i>ns</i>
<i>MMP9</i>	1.26	1.28	1.50	1.16	1.45	1.58	<i>ns</i>	<i>ns</i>	<i>ns</i>
<i>SELL</i>	1.34	1.20	1.48	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>

2.4.1 | Differential expression analysis

We first tested if there were genes with differential expression in the following categories: immune-rich and immune-poor and nodal metastasis status and breast cancer receptor subtype. We then evaluated if expression in LM22 signature genes was changed in immune-rich samples across the conditions of subtype or nodal status. Differential expression (DE) among immune-rich and -poor tumors was assessed within each subtype and compared against all other subtypes (grouped together, not pairwise). Results lists from each source of patient data (METABRIC, GEO, and TCGA) were compared against one another for validation, and only intersecting genes with association FDR-adjusted $p < 0.05$ and log fold change in the same direction were considered. Since LM22 is used to identify primarily immune-rich tumors, we conducted a secondary, genome-wide, DE analysis excluding LM22 (non-LM22) genes in order to identify any shared genes expressed in immune-poor groups, and to check for additionally over expressed non-LM22 genes in immune-rich tumors. All LM22 genes were excluded in this testing, and results were filtered/validated according to FDR adjusted $p < 0.05$ and log fold change in the same direction.

2.5 | RESULTS

2.6 | Tumor immune signature groups

Receptor subtype was associated with both nodal status and immune-rich/poor tumors. Approximately 61% (1454/2369) of tumors in all datasets were Luminal type. Luminal patient tumors were more frequently immune-poor in all data. Conversely, TNBC and Her2+ patients consistently had more immune-rich tumors across all datasets (Table 1). No

association was found between nodal metastasis status and immune-poor/rich tumor groups in any of the three datasets (Table 1). In addition, no definitive connection between NM and immune-cell signature was identified in LM22 signature gene analysis (data not shown).

2.7 | mRNA differential association between immune phenotypes

While overall immune-rich/poor status differed significantly across receptor subtype, individual signature genes seemed to be the same across the three subtypes, suggesting that the same genes responsible for immune infiltration do not differ across subtype. There were 10 total LM22 signature genes with consistently higher expression in immune-rich tumors (FDR $p < 0.05$, see Table 2). However, the Her2 subtypes sample sizes were too small for estimates ($n = 30$ in TCGA), several of the TNBC only genes were not significant, most likely due to lack of power. TNBC receptor subtype showed no unique differences in genes associated with immune-rich status when compared to all receptor subtypes. Presence of lymph node metastasis had no consistent effect upon tumor immune-rich/poor status across datasets (analysis not shown).

2.7.1 | LM22 immune-rich differences across receptor subtypes

Within immune-rich tumors only, differential expression of genes was observed between both Luminal and TNBC subtypes. Nine LM22 genes had significant differential expression (FDR adjusted $p < 0.05$) and consistent direction of log fold change across the three datasets, (7 in TNBC and 2 in Luminal with respect to the other subtypes). Both

KYNU and *EPHA1* transcripts were significantly decreased in the Luminal immune-rich tumor group. In TNBC patients with immune-rich samples, expression of *CHI3L2* and *FES* was significantly upregulated relative to all other subtypes. Decreases of expression in *TRPM*, *ABCB9*, and *EPHA1* were unique to the TNBC immune-rich set of patients (Table 3). No consistent differences were found in the Her2 receptor subtype.

2.7.2 | Non-LM22 DE and pathway analysis

Excluding all LM22 signature genes, we performed a whole transcriptome analysis of differential expression between immune-rich and immune-poor tumors, adjusting for receptor subtype and nodal metastasis. Differential expression analysis identified 1,951 total genes shared across datasets with expression associated to immune-rich status. Of these results, nine genes showed statistically significant DE (FDR adjusted $p < 0.05$) between the 2 immune phenotypes consistently across the three datasets (Table 4). The expression of *GRIA2* and *SCUBE2* genes was associated with immune-poor tumors. Importantly, gene OVA applied to the models above found no KEGG pathways to be consistently upregulated in immune-poor tumors. In tumors with an immune-rich signature, 51 sets of KEGG pathways had significantly over-represented, and upregulated genes in all datasets (Figure 1). Over-represented pathways with the strong, consistent associations across datasets included; antigen processing and presentation, natural killer cell mediated cytotoxicity, NOD-like receptor signaling pathway, and various immune response to infection (Figure 2).

3 | DISCUSSION

Our approach classified 2369 patients into two groups (immune-rich, and immune-poor) using an immune cell expression signature. The overall frequency of immune-rich versus immune-poor tumors differed by receptor subtype; however, the specific signature genes remained the same, suggesting that signature expression is a stronger determinant of immune infiltration than receptor subtype. Common to both TNBC and Luminal patients, the genes for C-C motif chemokine ligand genes of *CCL5* and *CCL19* as well for the C-X-C motif chemokine *CXCL9* exhibited the highest differential expression in immune-rich tumors compared to immune-poor tumors, regardless of subtype. Granzyme genes encoding *GZMA* and *GZMK* were consistently high across cohorts of Luminal patients but not for TNBC. While Luminal tumors had the smallest proportion of immune-rich tumors, they also had the more significantly different LM22 signature genes than TNBC. This non-intuitive finding is partially due to the large overall proportion of breast tumors that are Luminal, adding more sample size and power for DE analysis. In immune-rich only tumor analysis, no significant increases in the expression of LM22 genes was found in Luminal versus all other subtypes. TNBC patients had the highest percentage of immune-rich versus immune-poor tumors per subtype. Both *CHI3L2* and *FES* showed TNBC-differential over expression in immune-rich subset analysis. Recent breast cancer models in mice have also indicated increased *FES* expression as linked to tumor aggressiveness, macrophage infiltration, and involvement in the JAK3-Fes-PLD2 signaling pathway.^{29–32} Chitinase 3-like-2 (*CHI3L2*) has been identified as a marker of M2 tumor associated

Gene	METABRIC		GEO		TCGA	
	Log FC (M)	FDR p -value (M)	Log FC (G)	FDR p -value (G)	Log FC (T)	FDR p -value (T)
Subtype						
Luminal vs other						
<i>KYNU</i>	−1.02	<0.001	−3.12	<0.001	−1.82	<0.001
<i>EPHA1</i>	−0.64	<0.001	−0.11	0.01	−0.76	<0.001
TNBC vs other						
<i>CHI3L2</i>	1.18	<0.001	0.84	<0.001	2.10	<0.001
<i>FES</i>	0.74	<0.001	0.09	0.01	0.77	0.01
<i>FFAR2</i>	−0.35	0.03	−0.39	<0.001	−2.09	<0.001
<i>TRPM4</i>	−0.53	<0.001	−0.23	<0.001	−1.42	<0.001
<i>FAM174B</i>	−0.97	<0.001	−0.22	0.01	−1.27	<0.001
<i>ABCB9</i>	−0.60	0.03	−0.16	<0.001	−1.07	<0.001
<i>EPHA1</i>	−0.60	<0.001	−0.10	0.04	−0.61	0.05

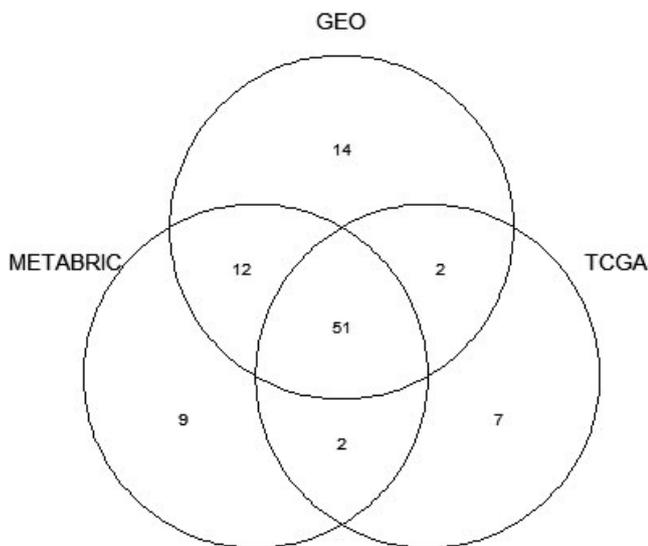
TABLE 3 Top LM22 genes differentially expressed by subtype within immune-rich tumors

Abbreviation: Log FC =Log Fold Change (subtype/all other subtypes).

TABLE 4 Expression of top non-LM22 signature genes: immune-rich versus immune-poor tumors

Gene	Log FC (METABRIC)	Log FC (GEO)	Log FC (TCGA)
<i>CIQB</i>	1.16**	0.81**	1.34**
<i>CD52</i>	1.89**	1.06**	2.63**
<i>CORO1A</i>	1.15**	0.82**	1.37**
<i>EVI2B</i>	1.42**	0.81**	1.36**
<i>GBP1</i>	1.02**	0.97**	1.47**
<i>LYZ</i>	1.83**	1.01**	2.05**
<i>PSMB9</i>	1.17*	0.84**	1.21**
<i>GRIA2</i>	-0.58**	-2.35*	-1.30**
<i>SCUBE2</i>	-0.57**	-0.54*	-1.21**

Abbreviation: Log FC =Log Fold Change, **FDR $p < 0.0001$, *FDR $p < 0.01$.

**FIGURE 1** Upregulated pathways in over-representation analysis (OVA) and overlap among datasets

macrophages, yet its effect on tumor progression in breast cancer is not fully known.^{33,34}

We also observed several new relationships between non-LM22 genes and immune-rich tumors, specifically over-expression of *LYZ*, *CIQB*, *CORO1A*, *EVI2B*, *GBP1*, *PSMB9*, and *CD52* in immune-rich tumors independent of receptor subtype or nodal metastasis. Likewise, overexpression of both *SCUBE2* and *GRIA2* was consistently associated with immune-poor status across datasets. Interestingly, *SCUBE2* has been associated with ER/PR positive tumors³⁵ and proposed as a tumor suppressor in breast cancer,³⁶ yet it has also been shown as the only common gene in outcome-predictive signature assays in breast cancer.³⁷ The gene *GRIA2* is a member of a family of glutamate receptors, and its expression has been linked to the modulation of both tumor and immune cell function.³⁸ We identified several novel markers which were

overexpressed in immune-rich tumors. CAMPATH-1 antigen (*CD52*) is expressed in mature lymphocytes. It is the target of the monoclonal antibody therapy alemtuzumab in several diseases including chronic lymphocytic leukemia, cutaneous T-cell lymphoma (CTCL), and multiple sclerosis (MS).³⁹ In breast cancer stromal tissue, *CD52* has been linked to an expression signature of poor outcome,^{40,41} our results give further evidence of the potential of its overexpression as marker of immune cell invasion in breast tumors.

One of the genes most differentially expressed between immune-rich and immune-poor tumors was lysozyme (*LYZ*). Whether increased lysozyme production originates from normal or cancerous breast cells or from invasive immune cells, or both, is less clear. Lysozyme is secreted by neutrophils and macrophages in inflammatory response,⁴² and is also present in the breast secretions of ~98% of healthy non-lactating women.⁴³ Previous research of the breast milk protein in breast tumors showed an inverse relationship between cancer severity and increased expression of *LYZ*.⁴⁴ In the immune response, lysozyme degrades the bacterial cell wall with components that activate molecules downstream of pattern recognition receptors such as inflammasomes, Toll-like receptors, and NOD-like receptors.⁴⁵ In OVA, NOD-like receptor signaling was upregulated in invasive immune cells, particularly the guanylate binding proteins of *GBP1*, *GBP4*, and *GBP5*. GBP-1 has been associated with various antimicrobial activities as well as paclitaxel resistance in cell lines of some tumors with the notable exception of breast.⁴⁶ While its contributions to prognostic factors across cancer types is varied, *GBP1* has been associated with improved survival in breast cancer.⁴⁷

There was no clear association between our signature of tumor immune cell infiltration and the extrinsic phenotype of nodal metastasis status. No single gene result from differential expression testing was both significant and consistent in direction of association across datasets. In addition, no interaction by subtype was observed which might have obscured the association between NM and RNA expression by immune signature group. These results could suggest that while draining lymph nodes are a vital part of the antitumor immune response, the process of lymph node metastasis in breast cancer is not demonstrably linked to the infiltration of tumor invasive immune cells. Importantly, our results of LM22 genes being associated with immune-poor phenotype are expected, given the use of LM22 in classifying immune cells. Furthermore, we found no consistently over/under represented pathways for immune-poor tumors, and only 2 non-LM22 genes were significantly upregulated in immune-poor analysis.

In summary, our analysis identified several signature tumor transcripts associated with immune-rich and immune-poor status. Future research on the topic may be informed by the finding that, while there are individual immune-rich

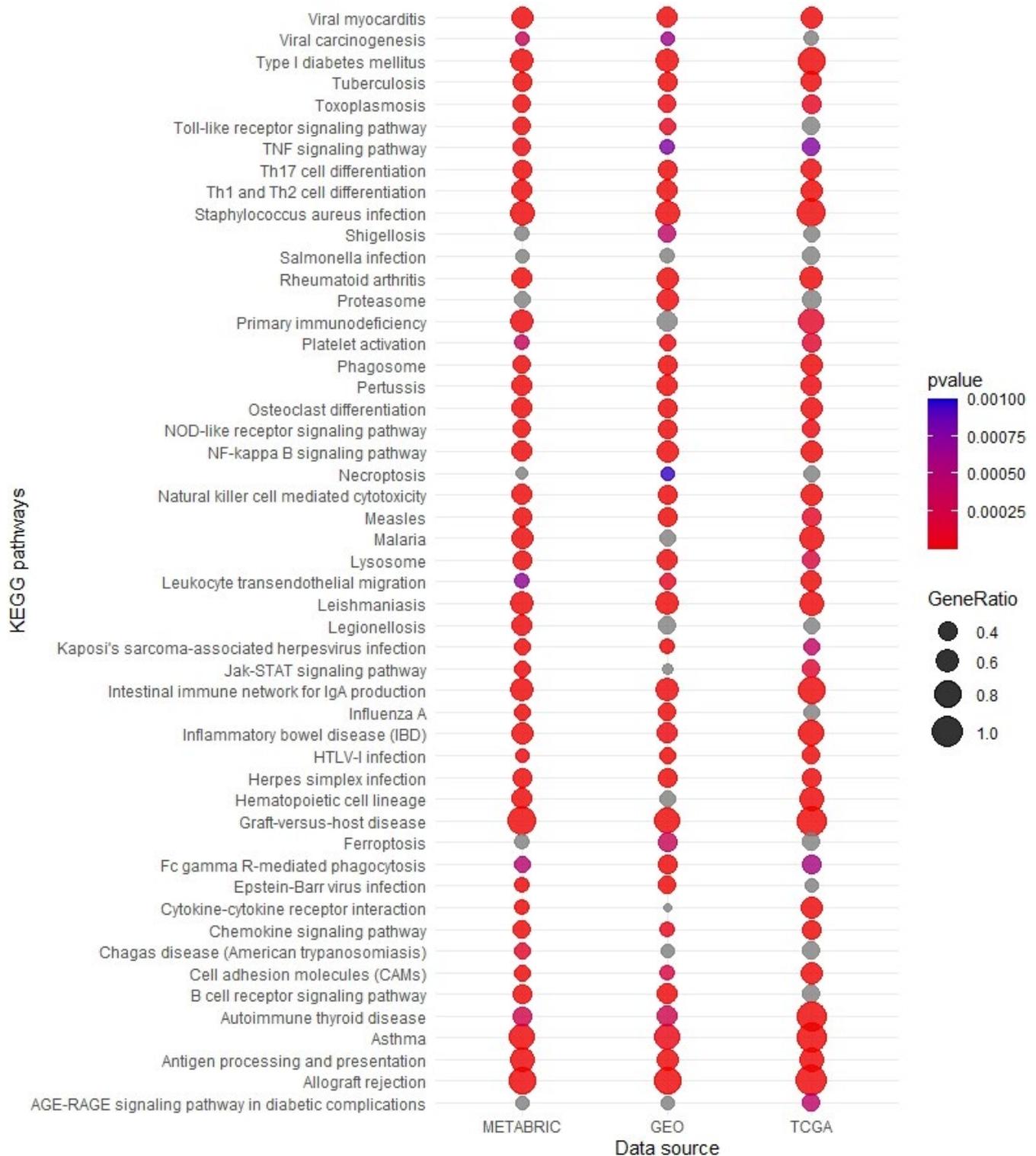


FIGURE 2 Over-representation analysis results of 51 upregulated KEGG pathways in immune-rich tumor expression. Gene ratio is number of upregulated genes/total genes in pathway

genes which are uniquely expressed in TNBC and Luminal tumors, the majority of validated, immune-rich gene expression is not affected by receptor subtype or nodal metastasis. Additionally, these data provide two potential novel targets for the study of the immune microenvironment of immune-poor tumors; *SCUBE2* and *GRIA2*.

ACKNOWLEDGEMENT

We would like to thank: The UAB Comprehensive Cancer Center Young Investigator Grant (Dr. Vazquez) & Quetelet Endowment Professorship Fund (Dr. Shrestha). Dr. Behring would like to thank the Cancer Prevention and Control Training Program (NCI T32 CA047888)

CONFLICT OF INTEREST

None.

AUTHORS' CONTRIBUTIONS

Conception and design: M. Behring, S. Shrestha, A.I. Vazquez, **Development of methodology:** M Behring, A.I. Vazquez, **Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** M. Behring, S. Shrestha, A.I. Vazquez, **Writing, review, and/or revision of the manuscript:** M. Behring, Y. Ye, A. Elkholy, P. Bajpai, S. Agarwal, H. Kim, H. W. Wiener, A. I. Ojesina, S. Shrestha, U. Manne, A.I. Vazquez, **Study supervision:** S. Shrestha, U. Manne, A.I. Vazquez.

ORCID

Michael Behring  <https://orcid.org/0000-0002-5162-322X>

Akinyemi I. Ojesina  <https://orcid.org/0000-0003-0755-3639>

Upendar Manne  <https://orcid.org/0000-0002-1545-3032>

REFERENCES

- Liu XS, Mardis ER. Applications of immunogenomics to cancer. *Cell*. 2017;168(4):600-612. <https://doi.org/10.1016/j.cell.2017.01.014>
- Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity*. 2004;21:137-148. <https://doi.org/10.1016/j.immuni.2004.07.017>
- Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39(4):782-795. <https://doi.org/10.1016/j.immuni.2013.10.003>
- Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol*. 2002;3(11):991-998. <https://doi.org/10.1038/ni1102-991>
- Salgado R, Denkert C, Demaria S, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILS Working Group 2014. *Ann Oncol*. 2015;26(2):259-271. <https://doi.org/10.1093/annonc/mdu450>
- Williams CB, Yeh ES, Soloff AC. Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy. *NPJ Breast Cancer*. 2015;2016(2):15025. <https://doi.org/10.1038/npjbcancer.2015.25>
- Baxevasis CN, Sofopoulos M, Fortis SP, Perez SA. The role of immune infiltrates as prognostic biomarkers in patients with breast cancer. *Cancer Immunol Immunother*. 2019;68(10):1671-1680. <https://doi.org/10.1007/s00262-019-02327-7>
- Stanton SE, Adams S, Disis ML. Variation in the incidence and magnitude of tumor-infiltrating lymphocytes in breast cancer subtypes a systematic review. *JAMA Oncol*. 2016;2(10):1354-1360. <https://doi.org/10.1001/jamaoncol.2016.1061>
- Ibrahim EM, Al-Foheidi M, Al-Mansour MM, Kazkaz GA, Yunus TE. The prognostic and predicting roles of tumor-infiltrating lymphocytes in breast cancer: A meta-analysis. *Open Breast Cancer J*. 2014;6(1):9-19.
- Solinas C, Carbognin L, De Silva P, Criscitiello C, Lambertini M. Tumor-infiltrating lymphocytes in breast cancer according to tumor subtype: current state of the art. *Breast*. 2017;35:142-150. <https://doi.org/10.1016/j.breast.2017.07.005>
- Desmedt C, Zoppoli G, Gundem G, et al. Genomic characterization of primary invasive lobular breast cancer. *J Clin Oncol*. 2016;34(16):1872-1880. <https://doi.org/10.1200/JCO.2015.64.0334>
- Ali HR, Chlon L, Pharoah PDP, Markowitz F, Caldas C. Patterns of immune infiltration in breast cancer and their clinical implications: a gene-expression-based retrospective study. *PLoS Med*. 2016;13(12):1-24. <https://doi.org/10.1371/journal.pmed.1002194>
- Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70. <https://doi.org/10.1038/nature11412>
- Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-352.
- Zhao XI, Rødland EA, Sørbye T, et al. Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer*. 2014;14(1):211. <https://doi.org/10.1186/1471-2407-14-211>
- Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*. 2006;10(6):529-541. <https://doi.org/10.1016/j.ccr.2006.10.009>
- Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol*. 2007;25(10):1239-1246. <https://doi.org/10.1200/JCO.2006.07.1522>
- Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA*. 2005;102(38):13550-13555. <https://doi.org/10.1073/pnas.0506230102>
- Pawitan Y, Bjöhle J, Amler L, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res*. 2005;7(6):R953. <https://doi.org/10.1186/bcr1325>
- Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res*. 2007;13(11):3207-3214. <https://doi.org/10.1158/1078-0432.CCR-06-2765>
- Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. *Nature*. 2005;436(July):518-524. <https://doi.org/10.1038/nature03799>
- Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Meth*. 2016;21(3):193-201. <https://doi.org/10.1016/j.molmed.2014.11.008>
- Dave SS, Wright G, Tan B, et al. Prediction of survival in Follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med*. 2004;351:2159-2169.
- Nagalla S, Chou JW, Willingham MC, et al. Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol*. 2013;14(4):R34. <https://doi.org/10.1186/gb-2013-14-4-r34>
- Ritchie ME, Phipson B, Wu DI, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. <https://doi.org/10.1093/nar/gkv007>

26. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3(1):1-25. <https://doi.org/10.2202/1544-6115.1027>
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289-300. <https://doi.org/10.2307/2346101>
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140. <https://doi.org/10.1093/bioinformatics/btp616>
29. Zhang S, Chitu V, Stanley ER, Elliott BE, Greer PA. Fes tyrosine kinase expression in the tumor niche correlates with enhanced tumor growth, angiogenesis, circulating tumor cells, metastasis, and infiltrating macrophages. *Cancer Res*. 2011;71(4):1465-1473. <https://doi.org/10.1158/0008-5472.CAN-10-3757>
30. Kwok E, Everingham S, Zhang S, Greer PA, Allingham JS, Craig AWB. FES kinase promotes mast cell recruitment to mammary tumors via the stem cell factor / KIT receptor signaling axis. *Mol Cancer Res*. 2012;16:881-891. <https://doi.org/10.1158/1541-7786.MCR-12-0115>
31. Ye Q, Kantonen S, Henkels KM, Gomez-Cambronero J. A new signaling pathway (JAK-Fes-phospholipase D) that is enhanced in highly proliferative breast cancer cells. *J Biol Chem*. 2013;288(14):9881-9891. <https://doi.org/10.1074/jbc.M113.450593>
32. Gomez-Cambronero J. Phospholipase D in cell signaling: From a myriad of cell functions to cancer growth and metastasis. *J Biol Chem*. 2014;289(33):22557-22566. <https://doi.org/10.1074/jbc.R114.574152>
33. Kzhyshkowska J, Gudima A, Moganti K, Gratchev A, Orekhov A. Perspectives for monocyte/macrophage-based diagnostics of chronic inflammation. *Transfus Med Hemotherapy*. 2016;43(2):66-77. <https://doi.org/10.1159/000444943>
34. Zhao T, Su Z, Li Y, Zhang X, You Q. Chitinase-3 like-protein-1 function and its role in diseases. *Signal Transduct Target Ther*. 2020;5(1):1-20. <https://doi.org/10.1038/s41392-020-00303-7>
35. Cheng C-J, Lin Y-C, Tsai M-T, et al. SCUBE2 suppresses breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer. *Cancer Res*. 2009;69(8):3634-3641. <https://doi.org/10.1158/0008-5472.CAN-08-3615>
36. Lin Y-C, Chen C-C, Cheng C-J, Yang R-B. Domain and functional analysis of a novel breast tumor suppressor protein, SCUBE2. *J Biol Chem*. 2011;286(30):27039-27047. <https://doi.org/10.1074/jbc.M111.244418>
37. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med*. 2009;360:790-800. <https://doi.org/10.1056/NEJMr a0801289>
38. Stepulak A, Rola R, Polberg K, Ikonomidou C. Glutamate and its receptors in cancer. *J Neural Transm*. 2014;121(8):933-944. <https://doi.org/10.1007/s00702-014-1182-6>
39. Hale G, Bright S, Chumbley G, et al. Removal of T cells from bone marrow for transplantation: a monoclonal antilymphocyte antibody that fixes human complement. *Blood*. 1983;62(4):873-882.
40. Finak G, Bertos N, Pepin F, et al. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med*. 2008;14(5):518-527. <https://doi.org/10.1038/nm1764>
41. Wang J, Zhang G, Sui Y, et al. CD52 Is a prognostic biomarker and associated with tumor microenvironment in breast cancer. *Front Genet*. 2020;11(November): 1350-1358. <https://doi.org/10.3389/fgene.2020.578002>.
42. Callewaert L, Michiels CW. Lysozymes in the animal kingdom. *J Biosci*. 2010;35(1):127-160. <https://doi.org/10.1007/s12038-010-0015-5>
43. Petrakis NL. Physiologic, biochemical, and cytologic aspects of nipple aspirate fluid. *Breast Cancer Res Treat*. 1986;8(1):7-19. <https://doi.org/10.1007/BF01805919>
44. Vizoso F, Plaza E, Vazquez J, et al. Lysozyme expression by breast carcinomas, correlation with clinicopathologic parameters, and prognostic significance. *Ann Surg Oncol*. 2001;8(8):667-674. <https://doi.org/10.1245/aso.2001.8.8.667>
45. Ragland SA, Criss AK. From bacterial killing to immune modulation : Recent insights into the functions of lysozyme. *PLoS Pathog*. 2017;13:1-22. <https://doi.org/10.1371/journal.ppat.1006512>
46. Tipton AR, Nyabuto GO, Trendel JA, et al. Guanylate-Binding Protein-1 protects ovarian cancer cell lines but not breast cancer cell lines from killing by paclitaxel. *Biochem Biophys Res Commun*. 2016;478(4):1617-1623. <https://doi.org/10.1016/j.bbrc.2016.08.169>
47. Ascierto M, Idowu MO, Zhao Y, et al. Molecular signatures mostly associated with NK cells are predictive of relapse free survival in breast cancer patients. *J Transl Med*. 2013;11(1):145. <https://doi.org/10.1186/1479-5876-11-145>

How to cite this article: Behring M, Ye Y, Elkholy A, et al. Immunophenotype-associated gene signature in ductal breast tumors varies by receptor subtype, but the expression of individual signature genes remains consistent. *Cancer Med*. 2021;10:5712-5720. <https://doi.org/10.1002/cam4.4095>