Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: October 31, 2022

Citation: Chen Y., 2022. A Cancer Proliferation Gene Signature Supervised by Ki-67 Strata Specific to Luminal A, Estrogen Receptor-Positive, and HER2-Negative Ductal Carcinomas, Medical Research Archives, [online] 10(10). https://doi.org/10.18103/mra.v10i10.3160

DOI
https://doi.org/10.18103/mra.v10i10.3160

ISSN: 2375-1924

RESEARCH ARTICLE

A Cancer Proliferation Gene Signature Supervised by Ki-67 Strata Specific to Luminal A, Estrogen Receptor-Positive, and HER2-Negative Ductal Carcinomas

Youdinghuan Chen, Ph.D.[1,2]

1. Faculty of Bioinformatics and Data Science, College of Health Professions and Natural Sciences, Wilmington University. 320 N DuPont Highway, New Castle, DE 19720.
2. National Coalition of Independent Scholars, 125 Putney Road, Battleboro, VT 05301

*ydavidkchen@gmail.com

ABSTRACT
Clinically determined Ki-67 is a well-established marker for assessing proliferation potential in breast and other cancers. However, Ki-67 and the recommended thresholds for clinical decision-making vary systematically across breast cancer subtypes. In this study, an analysis of published gene expression data against Ki-67 in ER+/HER2- and Luminal A ductal carcinomas identified 127 of 14,997 protein-coding genes (elastic-net coefficient ≠ 0). The most upregulated genes associated with high Ki-67 are involved in cancer proliferation and were known in breast cancer studies, while the downregulated genes are involved in diverse signaling transduction processes. Application of the identified gene signature to ER+/HER2- and Luminal A ductal carcinomas consistently stratified two independent, population-based breast cancer cohorts. Although the ER+/HER2- clinical and the Luminal A intrinsic subtypes typically show good prognosis, one subpopulation identified by the signature showed an elevated risk of disease recurrence (hazards ratios 1.59 [95% CI 1.02, 2.47] and 3.80 [95% CI 0.83, 17.27] in two independent application cohorts). The present study identifies a proliferation gene signature specific to ER+/HER2- and Luminal A ductal carcinomas, provides biological insight into the more proliferative cancers, and could be a basis for future therapeutic development.
Keywords: Gene signature, Ki-67, ER+/HER2-, Luminal A, invasive ductal carcinoma, survival

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor 2; LumA, Luminal A intrinsic subtype; IDC, invasive ductal carcinoma; RNA-seq, RNA sequencing; HR, hazards ratios; CI, confidence interval

CONFLICT OF INTEREST STATEMENT: None

FUNDING SOURCES: None

SUPPLEMENTAL MATERIALS. Supplemental Figures (PDF) and Supplemental Tables (Excel) are hosted under FigShare accession https://doi.org/10.6084/m9.figshare.21187843.v1.

INTRODUCTION

Elevated cellular proliferation is a hallmark of cancer. Unchecked proliferation prevents cells from responding to programmed cell death signals and contributes to carcinogenesis as well as cancer metastasis[1]. One of the most well-studied proliferation biomarkers is Ki-67, a multi-functional protein and positive regulator of the cell cycle[2,3]. In the breast tissue context, Ki-67 expression is minimal in non-cancer cells but elevated in tumors. High Ki-67 expression predicts poorer survival but a positive response to neoadjuvant hormonal therapy and cytotoxic chemotherapy[2,3]. The Ki-67 level also varies greatly across breast cancer subtypes, and as a result, the recommended threshold for clinical decision-making based on Ki-67 varies[2]. For example, Ki-67 of 10% to 20% is recommended for hormone receptor-positive cases, while a 30% or higher threshold is recommended for the more aggressive triple-negative cases[2–4].

The ER+/HER2- subtype, defined based on the presence of estrogen receptor (ER) and the absence of human epidermal growth factor 2 (HER2), accounts for the largest fraction of breast cancer cases. The prognostic value of Ki-67 is well-known in this tumor subtype. Most ER+/HER2- cases are also classified as the Luminal (LumA and LumB) intrinsic subtype based on gene expression[5–7]. Although typically treatable by hormonal endocrine therapies and shows favorable prognosis, this tumor subset still accounts for the largest fraction of breast cancer cases and is therefore an urgent health-care need.

While the utility of Ki-67 is well established, the level of this biomarker is not always available. The rise of "omics" data represents new opportunities to develop biomarkers for precision medicine. Transcriptomic datasets have shown their potential in helping to identify biomarkers in many diseases and subtypes[7]. For example, The Cancer Genome Atlas Breast Cancer (TCGA-BRCA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) projects comprehensively profiled and integrated breast cancer "omics" with clinical features at the population scale, providing the research community with new biological knowledge and clinical insight into the disease as well as rich data resources[5,6].

Since Ki-67 level and gene expression profiles differ between the more proliferative tumors and their less proliferative counterparts, a gene signature that indicates cellular proliferation potential likely exists. However, a breast cancer subtype-specific gene signature for tumor proliferation remains under-explored. This could partly be because of the concurrent lack of clinical measurement of biomarkers such as Ki-67 and high-quality transcriptomic profiles. Because of the heterogeneity in Ki-67 levels across breast cancer subtypes, there is a need for subtype-specific analyses. The present work presents a transcriptome signature identified by a supervised re-analysis of clinically measured Ki-67 in invasive ductal carcinoma (IDC) having both the ER+/HER2- clinical and the LumA intrinsic subtypes, together referred to as "ER+/HER2-/LumA IDCs". When applied to two independent, population-scale breast cancer cohorts, the gene signature consistently revealed two major subpopulations. In each application cohort, one subpopulation displayed distinct clinical characteristics and worse disease-free survival. The identified gene signature revealed biological processes underlying a benign breast cancer subset, which could have clinical utility.

METHODS

Study design. The workflow of this study is summarized in **Figure 1A**. Briefly, a set of ER+/HER2-/LumA IDCs with both RNA-seq gene expression and clinically measured Ki-67 staining from Wu et al.[8] was used to identify the proliferation gene signature. Subsequently, hierarchical clustering of the expression of the identified gene set was used to stratify two population-scale breast cancer cohorts[5,6] for clinical knowledge discovery, including the comparison of patient covariates and survival outcomes.

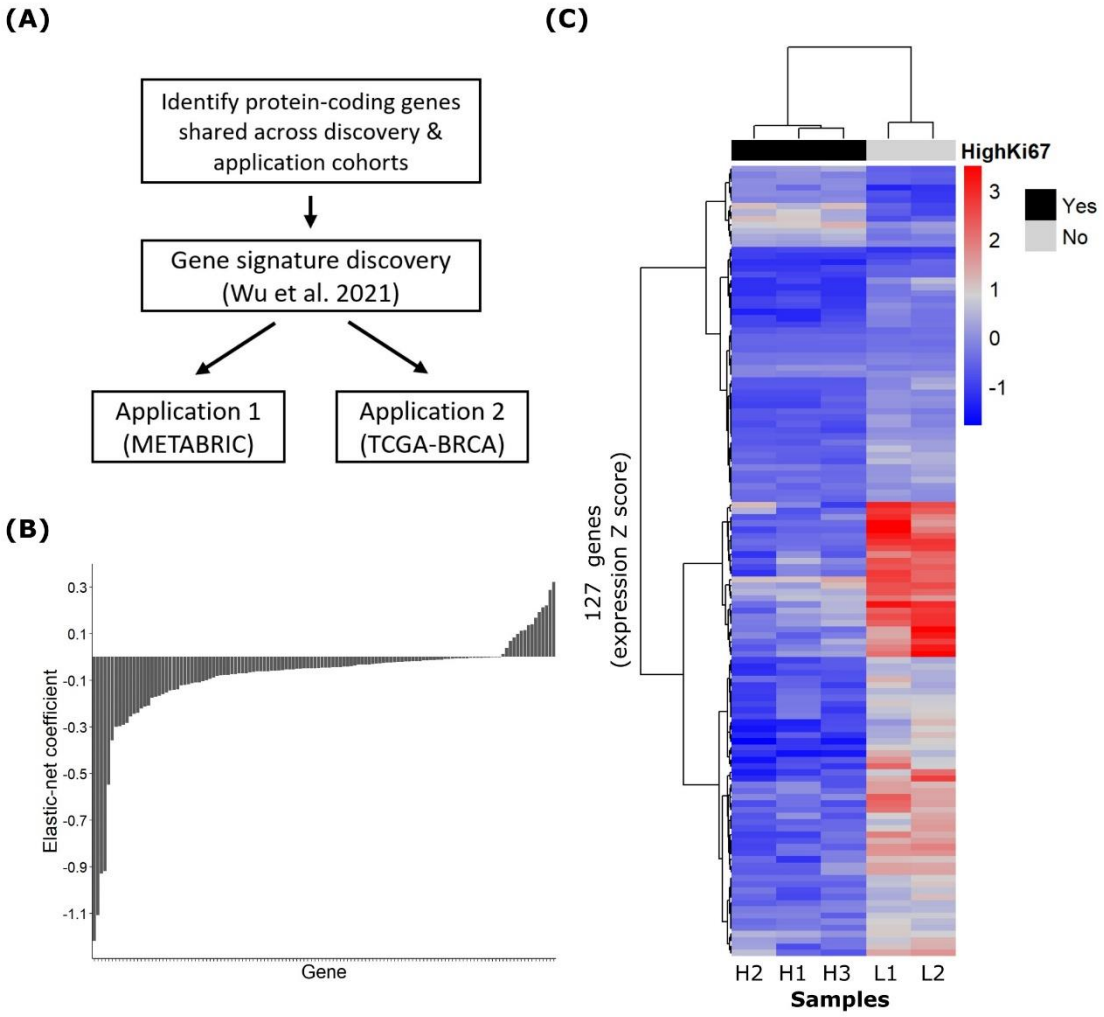

Figure 1. Identification of the proliferation gene signature specific to ER+/HER2-/LumA IDCs. (A) Workflow summary. **(B)** Rank-ordered distribution of the gene signature elastic-net regression coefficients; related to Supplemental Table 2. **(C)** Hierarchical clustering of the 127 genes differentially expressed between Ki-67 strata in the discovery cohort.

Selection and processing of the background gene set. A three-way intersection was performed on the discovery and application cohorts' gene expression data to obtain a background gene set. The selected background genes were further filtered to contain only protein-coding genes from the Hugo Nomenclature annotation (available at www.genenames.org). Each cohort had 14,997 shared protein-coding genes, which were subsequently mean-centered and standard deviation-scaled. The transformed expression data distribution was inspected and then constrained within the ±3.50 range.

Identification of the gene signature. An elastic-net logistic regression model (hyperparameter *alpha* = 0.5, representing a 50:50 mix of L_1 and L_2 regression penalties; *glmnet* R package version 4.1-4)[9] was built using the background genes of the discovery cohort as the input and the binary Ki-67 strata as the target. The optimal model had the minimum *lambda* parameter (= 0.009). The gene signature was defined as the panel of genes whose elastic-net regression coefficients were non-zero in the optimal model. Putative functions of the most differential genes were determined by querying the GeneCards database[10].

Stratification of the application cohorts by the identified gene signature. The processed gene expression of the application cohorts was restricted to ER+/HER2- clinical, LumA intrinsic, and IDC histological subtypes. Hierarchical clustering of the identified gene signature with Euclidean distance and Ward D linkage was used for binary stratification (base R function *dist* and *hclust*; visualization by *pheatmap*). Cluster membership (Cluster 1 and Cluster 2) was automatically assigned using the *cutree* base R function at the first tree split (hyperparameter $k = 2$).

Statistical analysis. All statistical tests were two-sided. Welch *t*-tests and Fisher's exact tests were performed in R 4.2.1 (www.rproject.org). The comparison of clinical characteristics was performed by R package *tableone* 0.13.2. All survival analyses were censored at the ten-year mark and implemented in R packages *survival* 3.3-1 and *survminer* 0.4.9. The covariate-adjusted Cox proportional hazards model of recurrence-free survival was performed on the subset of application-cohort cases with complete age, tumor stage, recurrence-free interval and status.

Data availability. All datasets used in this study are publicly available, as shown in **Table 1**. High-quality transcriptomic profiles and patient covariates of Wu et al. were downloaded from Gene Expression Omnibus accession GSE176078 and the original publication, respectively[8]. TCGA-BRCA gene expression and clinical annotation were downloaded from the Synapse database (https://doi.org/10.7303/syn300013, accession syn2320114) and National Cancer Institute Genomic Data Commons (https://gdc.cancer.gov/about-data/publications/pancanatlas), respectively. METABRIC gene expression and clinical annotations were downloaded from cBioPortal (www.cbioportal.org)[6,11]. Only samples with gene expression data were included.

Code availability. The code for statistical analysis and data visualization can be found in the GitHub repository https://github.com/ydavidchen/ki67-pilot-signature.

Table 1. Datasets with transcriptomic-wide gene expression used in this study. ER+/HER2-, estrogen receptor-positive and human epidermal growth factor 2-negative breast tumors; LumA, Luminal A intrinsic subtype; IDC, invasive ductal carcinoma.

	Publication	Expression assay	n (%), IDC	n (%), ER+/HER2-	n (%), LumA	n (%), ER+/HER2-/LumA IDC with gene expression
Discovery Cohort	Wu et al. (*Nature Genetics* 2021)	RNA-seq	22 (84.6)	12 (46.2)	5 (19.2)	5 (19.2)
Application Cohort 1	METABRIC (Curtis et al. *Nature* 2012)	Microarray	1,810 (72.9)	1,382 (55.1)	700 (27.9)	444 (17.7)
Application Cohort 2	TCGA-BRCA (Cancer Genome Atlas Network, *Nature* 2012)	RNA-seq	779 (71.9)	585 (54.0)	499 (46.1)	232 (21.4)

RESULTS

The rationale for Ki-67 stratification and case selection for supervised gene signature identification. A systematic difference in Ki-67 was observed across breast cancer molecular subtypes among patients and samples published by Wu et al. (the discovery cohort)[8]. Although ER+/HER2- tumors show a wide range of Ki-67 levels (3-75%, median 12.5%), the LumA ER+/HER- subset and low (< 10%) Ki-67 expression were significantly associated suggesting potential confounding by intrinsic subtype (Fisher's exact test $p < 0.05$; **Supplemental Figure 1 table inset**). This finding was corroborated by Ki-67 mRNA expression in two population-scale breast cancer cohorts, METABRIC and TCGA-BRCA (the application

cohorts)[5,6], where ER+/HER2- tumors that were also LumA showed significantly lower expression than their LumB counterpart (both Welch *t* test $p < 2.2e\text{-}16$ and both mean difference in expression < 0; **Supplemental Figure 1**). The observed systematic difference in Ki-67 levels across subtypes motivated downstream analyses restricting to ER+/HER2- cases that are also LumA.

Supervised identification of a proliferation gene signature specific to ER+/HER2-/LumA tumors. Restricting to the ER+/HER2- invasive ductal carcinomas (IDCs) that were also LumA (hence referred to as "ER+/HER2-/LumA IDCs"), five patients from the discovery cohort were identified (**Table 2**). These tumors were assigned to "High Ki-67" and "Low Ki-67" groups using a 10% threshold. The strata showed a similar distribution in age, tumor grade, and tumor stage (**Table 2**). Applying an elastic-net logistic regression model to the background gene set of the discovery cohort, 127 genes showed non-zero coefficients indicating their differential expression between Ki-67 strata (hence referred to as the "gene signature"; Supplemental Table 1). Hierarchical clustering of this gene signature perfectly segregated Ki-67 high vs. low tumors in the discovery cohort (**Figure 1C**). In addition, the identified gene signature did not show an overlap with genes currently used in the established breast cancer diagnostic tests (Prosigna/PAM50, OncotypeDx, Mammaprint, MammoStrat, and Breast Cancer Index)[7], while up to 80% of the diagnostic-test genes were included in the background gene set (**Supplemental Table 2**).

Table 2. Characteristics of the discovery cohort. All subjects had the ER+/HER2- clinical, LumA intrinsic, and IDC histological subtypes as well as bulk-tumor RNA-seq data.

Subject	Original ID	Age > 50	Grade	T Stage	Ki-67 (%)	Ki-67 group
H1	3941	No	II	T1c	10	High
H2	3948	Yes	III	T2	10	High
H3	4290	Yes	II	T4b	10	High
L1	4067	Yes	II	T2	≤4	Low
L2	4530	No	II	T3	5	Low

Potential clinical utility of the identified gene signature. To determine its potential prognostic utility, the gene signature was applied to the two population-based application cohorts, METABRIC and TCGA-BRCA, "the application cohorts". Restricting to the ER+/HER2-/LumA IDCs, the identified gene signature stratified each application cohort into two subpopulations (**Figure 2A**). The cluster structure of each cohort showed striking similarities, hinting at the commonality of the identified gene signature. A comparison of clinical characteristics revealed that patients in Cluster 1 were significantly younger (both cohorts' Welch *t* test $p < 0.01$; **Table 3**). There was also a significant relationship between menopause status and cluster membership (both cohorts' Fisher's exact test $p < 0.01$;

Table 4). In application cohort 2 where race is available, the association between cluster membership and race was significant (Fisher's exact test $p < 0.05$; **Table 3**).

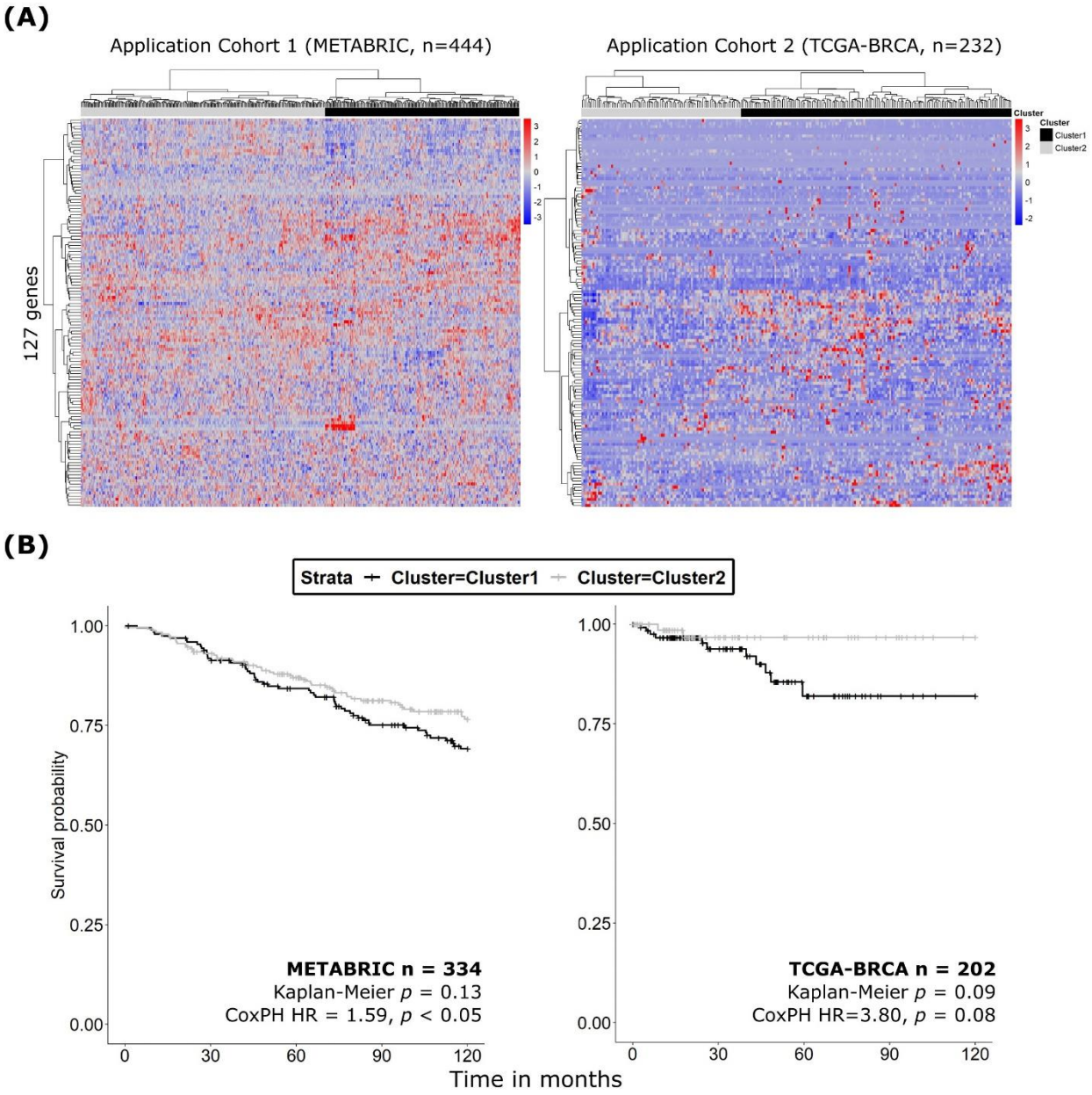

Figure 2. Application of the proliferation gene signature to ER+/HER2-/LumA IDCs to population-based breast cancer cohorts. **(A)** Hierarchical clustering for cluster identification. **(B)** Disease free survival on subset of cases with complete survival, age at diagnosis, and tumor stage. HR, hazards ratios; CI, confidence interval.

Table 3. Comparison of clinical characteristics of ER+/HER2-/LumA IDCs in the application cohorts stratified on the gene-signature clusters. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

	Application Cohort 1 (METABRIC)			Application Cohort 2 (TCGA-BRCA)		
Cluster by gene signature	**Cluster 1**	**Cluster 2**	***p***	**Cluster 1**	**Cluster 2**	***p***
n	197	247		146	86	
Age at diagnosis (mean (SD))	66.32 (11.87)	60.24 (13.11)	***4.72E-7	59.47 (13.58)	54.27 (13.38)	**0.005
Tumor mutational burden (mean (SD))	7.39 (4.21)	6.81 (3.94)	0.14	1.45 (2.18)	4.07 (19.47)	0.11
Tumor stage (%)			0.23			0.56
I-II	137 (69.5)	185 (74.9)		118 (80.8)	68 (79.1)	
III-IV	4 (2.0)	8 (3.2)		27 (18.5)	16 (18.6)	
Not available	56 (28.4)	54 (21.9)		1 (0.7)	2 (2.3)	
Menopause status (%)			**0.002			**0.003
Pre	25 (12.7)	61 (24.7)		29 (19.9)	32 (37.2)	
Peri	0 (0.00)	0 (0.00)		4 (2.7)	7 (8.1)	
Post	172 (87.3)	186 (75.3)		100 (68.5)	42 (48.8)	
Not available	0 (0.00)	0 (0.00)		13 (8.9)	5 (5.8)	
Race category (%)						*0.049
White				110 (75.3)	59 (68.6)	
Black or African American	(Data not available)			15 (10.3)	18 (20.9)	
Asian				5 (3.4)	5 (5.8)	
Not available				16 (11.0)	4 (4.7)	

Table 4. Multivariate Cox proportional hazards regression model for recurrence-free survival with respect to the gene signature-driven clusters in the application cohort. All tumors are ER+, HER2-, LumA, and IDC. Only the cases with complete survival, age, or stage data were included. HR, hazards ratio; CI, confidence interval. $^{*}p < 0.05$, $^{\dagger}p < 0.1$.

Covariates	**Application Cohort 1 (METABRIC, n = 334)**		**Application Cohort 2 (TCGA-BRCA, n = 202)**	
	HR (95% CI)	***p***	**HR (95% CI)**	***p***
Age at diagnosis	1.00 (0.98, 1.02)	1.00	0.98 (0.94, 1.02)	0.26
Tumor stage				
I-II	(reference)		(reference)	
III-IV	1.32 (0.48, 3.64)	0.59	1.95 (0.52, 7.33)	0.32
Cluster by gene signature				
Cluster 2	(reference)		(reference)	
Cluster 1	1.59 (1.02, 2.47)	*0.041	3.80 (0.83, 17.27)	†0.084

In both application cohorts, Cluster 1 showed worse five-year disease-free survival in the univariate Kaplan-Meier analysis (**Figure 2B**). An age- and tumor stage-adjusted Cox proportional hazards model revealed a consistent, elevated risk of disease recurrence in Cluster 1 cases (hazards ratios 1.59 [95% CI 1.02-2.47] and 3.80 [95% CI 0.83-17.27] in the application cohorts 1 and 2; **Table 4**). In addition, Cluster 1 also showed worse five-year overall survival (hazards ratios 1.30 [95% CI 0.86-1.95] and 1.87 [95% CI 0.59-5.90] in the application cohorts 1 and 2; **Supplemental Figure 2** and **Supplemental Table 3**).

DISCUSSION

In this study, an ER+/HER2-/LumA IDC-specific gene signature with an expression profile that was differential between high and low Ki67 strata was identified by a supervised re-analysis. When applied to two large-scale, population-based cohorts, the gene signature stratified ER+/HER2-/LumA IDCs into two major sub-populations, one of which showed worse survival. The finding here suggests that even among cancers with a good prognosis, subsets with differential clinical outcomes might still exist. Because ER+/HER2- and Luminal tumors account for the largest fraction of breast cancer cases, identification of clinically relevant cancer subsets within this subset is valuable.

Elevated cellular proliferation is a hallmark of cancer and aggressive cancer. Identifying molecular biomarkers for cancer proliferation is an important first step in precision medicine. In breast cancer, Ki-67 is an excellent proliferation biomarker due to a minimal expression in non-cancer tissue and elevated levels by subtype. For example, Ki-67 levels and thresholds for clinical decision-making differ between the more treatable ER+/HER2- subtype and the more aggressive triple-negative/basal-like subtype[2–4]. In the discovery cohort, Ki-67 varied systematically across

breast cancer subtypes. On average, Ki-67 was indeed lower in ER+/HER2- while elevate in triple-negative cases. Furthermore, Ki-67 was lower in LumA than other intrinsic subtypes[8].

Despite its prognostic utility, clinically measured Ki-67 is not available in every setting. Nevertheless, transcriptome-wide gene expression is often collected in breast cancer studies to investigate various endpoints. A gene signature for cellular proliferation could be developed using a dataset with both transcriptomics and Ki-67 measurements. In a subtype-specific manner, the present study leveraged elastic net logistic regression supervised by Ki-67 levels to identify a gene signature from transcriptome-wide expression data.

The gene signature identified in this study likely has biologically and clinical relevance. The five genes most positively associated with Ki-67 had known involvement in cancer cell proliferation and were specifically studied in breast cancer. *HSPB8*, *SMARCE1*, and *NET1* had established roles in promoting cellular invasiveness specifically in the breast cancer context[12–14]. *NME1*, the gene with the fourth largest coefficient, targets two genes (*OSGEPL1* and *BRMS1*) whose expression is associated with favorable survival outcomes[15]. The downregulated gene set had more diverse cellular functions. Four genes most negatively associated with Ki-67 (*LEMD1*, *INSL4*, *PROCR*, and *ADRA2C*) are broadly involved in diverse intra- and inter-cellular signal transduction processes[10].

Interestingly, the Fanconi Anemia-family gene *FANCA* involved in homologous recombination DNA repair showed the eighth largest effect size among the 112 downregulated genes. Tumors incapable of homologous recombination cannot repair DNA double-stranded breaks, resulting in genomic instability, a hallmark of aggressive cancers[1,16]. Protein-changing mutations and variants in *FANCA* were associated with an increased risk of breast cancer[16,17]. Although the consequence of *FANCA* loss might not be as high as other related, higher-penetrance genes including *FANCD1* (*BRCA2*) and *FANCO* (*RAD51C*)[17], the observed downregulation in *FANCA* expression could be biologically and clinically relevant. On the one hand, *FANCA* downregulation might hint at the similarity between the highly proliferative cells and genomically unstable, more aggressive tumor cells lacking homologous recombination. On the other hand, *FANCA* downregulation could be an alternative mechanism leading to elevated cellular proliferation potential through genomic instability, another hallmark of cancer[1]. Typically, ER+/HER2- cancers display less extensive genomic perturbations[5,6], but recent evidence suggests some ER+/HER2- cases display genomic signatures similar to homologous recombination-deficient cancers[18,19].

Applying the gene signature to two large-scale breast cancer cohorts identified subpopulations (Clusters 1 and 2). The cluster structure was similar between the application cohorts, revealing the commonality of the gene signature in breast cancer. Cluster 1 subjects showed older age and differential menopause status. The observed differences in clinical features between clusters highlight the importance of the differentially expressed genes associated with cellular proliferation.

Compared to other breast cancer subtypes, the ER+/HER2- clinical and LumA intrinsic subtypes both display good prognosis and survival outcomes[5,6]. However, within these benign tumors, Cluster 1 showed elevated risk of disease recurrence in the univariate and covariate-adjusted multivariate models. There was also evidence of slightly worse overall survival in Cluster 1. This finding highlights the biological heterogeneity within a known cancer subtype.

This study has limitations and observations it cannot explain. First, Ki-67 was assumed to be an accurate reflection of cellular proliferation. This might not always be true because of the intra-tumoral heterogeneity as well as measurement errors and subjectivity of the evaluators[3]. Second, although all samples in the discovery cohort were from treatment-naïve patients, this was not the case for the application cohorts. The diversity in drug, surgical, and/or radiation therapy may impact the expression profile of the gene signature. Finally, the survival outcome of the application cohort 2 did not reach a statistical significance threshold *p*-value of 0.05. This was most likely due to the nearly 30% smaller sample size in the second application cohort especially after restricting to cases with complete survival and potential confounders. There might be alternative explanations, such as the fact that ER+/HER2- and Luminal A subtypes are associated with favorable survival and recurrence in general as well as the known heterogeneity in molecular and clinical profiles across datasets.

CONCLUSION

In this study, a supervised analysis of a published transcriptome-wide gene expression and Ki-67 level identified a 127-feature gene signature specific to ER+/HER2-/LumA IDCs. Based on the putative biological function of the genes, this gene signature may have biological relevance. Application of the identified gene signature to two

independent, population-based breast cancer cohorts shows evidence of differential survival outcomes. These findings revealed the biological and clinical heterogeneity within a benign tumor subset and could be used as a basis for developing future therapeutic strategies.

REFERENCES

1. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022;12(1):31-46. doi:10.1158/2159-8290.CD-21-1059
2. Zhang A, Wang X, Fan C, Mao X. The Role of Ki67 in Evaluating Neoadjuvant Endocrine Therapy of Hormone Receptor-Positive Breast Cancer. *Front Endocrinol (Lausanne)*. 2021;12:687244. doi:10.3389/fendo.2021.687244
3. Nielsen TO, Leung SCY, Rimm DL, et al. Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. *JNCI: Journal of the National Cancer Institute*. 2021;113(7):808-819. doi:10.1093/jnci/djaa201
4. Zhu X, Chen L, Huang B, et al. The prognostic and predictive potential of Ki-67 in triple-negative breast cancer. *Sci Rep*. 2020;10(1):225. doi:10.1038/s41598-019-57094-3
5. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70. doi:10.1038/nature11412
6. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-352. doi:10.1038/nature10983
7. Vieira AF, Schmitt F. An Update on Breast Cancer Multigene Prognostic Tests-Emergent Clinical Biomarkers. *Front Med (Lausanne)*. 2018;5:248. doi:10.3389/fmed.2018.00248
8. Wu SZ, Al-Eryani G, Roden DL, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet*. 2021;53(9):1334-1347. doi:10.1038/s41588-021-00911-1
9. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22.
10. Safran M, Rosen N, Twik M, et al. The GeneCards Suite. In: *Practical Guide to Life Science Databases*. Springer Nature Singapore; 2021:27-56. doi:10.1007/978-981-16-5812-9_2
11. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal*. 2013;6(269):pl1-pl1. doi:10.1126/scisignal.2004088
12. Piccolella M, Crippa V, Cristofani R, et al. The small heat shock protein B8 (HSPB8) modulates proliferation and migration of breast cancer cells. *Oncotarget*. 2017;8(6):10400-10415. doi:10.18632/oncotarget.14422
13. Sokol ES, Feng YX, Jin DX, et al. SMARCE1 is required for the invasive progression of in situ cancers. *Proc Natl Acad Sci U S A*. 2017;114(16):4153-4158. doi:10.1073/pnas.1703931114
14. Carr HS, Zuo Y, Oh W, Frost JA. Regulation of focal adhesion kinase activation, breast cancer cell motility, and amoeboid invasion by the RhoA guanine nucleotide exchange factor Net1. *Mol Cell Biol*. 2013;33(14):2773-2786. doi:10.1128/MCB.00175-13
15. McCorkle JR, Leonard MK, Kraner SD, et al. The metastasis suppressor NME1 regulates expression of genes linked to metastasis and patient outcome in melanoma and breast carcinoma. *Cancer Genomics Proteomics*. 11(4):175-194.
16. Litim N, Labrie Y, Desjardins S, et al. Polymorphic variations in the FANCA gene in high-risk non-BRCA1/2 breast cancer individuals from the French Canadian population. *Mol Oncol*. 2013;7(1):85-100. doi:10.1016/j.molonc.2012.08.002
17. del Valle J, Rofes P, Moreno-Cabrera JM, et al. Exploring the Role of Mutations in Fanconi Anemia Genes in Hereditary Cancer Patients. *Cancers (Basel)*. 2020;12(4). doi:10.3390/cancers12040829
18. Chen Y, Wang Y, Salas LA, et al. Molecular and epigenetic profiles of BRCA1-like hormone-receptor-positive breast tumors identified with development and application of a copy-number-based classifier. *Breast Cancer Res*. 2019;21(1):14. doi:10.1186/s13058-018-1090-z

19. Moore GM, Powell SN, Higginson DS, Khan AJ. Examining the prevalence of homologous recombination repair defects in ER+ breast cancers. *Breast Cancer Res Treat.* 2022;192(3):649-653. doi:10.1007/s10549-022-06529-z