GATA-3 Expression in Breast Cancer Has a Strong Association with Estrogen Receptor but Lacks Independent Prognostic Value

David Voduc,[1,2] Maggie Cheang,[1] and Torsten Nielsen[1,3]

[1]Genetic Pathology Evaluation Centre, University of British Columbia; [2]Department of Radiation Oncology, BC Cancer Agency; and [3]Department of Pathology, Vancouver Coastal Health Research Institute, Vancouver, British Columbia, Canada

Abstract

Background: **GATA-3 is a transcription factor involved in human growth and differentiation. Gene expression profiling has shown that GATA-3 is highly expressed in the Luminal A subtype of breast cancer. A recent study found GATA-3 to be associated with favorable breast cancer pathologic features, including negative lymph node and positive estrogen receptor (ER) status. GATA-3 levels were also found to be an independent prognostic marker, with low expression predicting for breast cancer recurrence.**

Materials and Methods: **Our case series consists of 3,119 cases of invasive breast cancer in which GATA-3 expression was assessed by immunohistochemistry on tissue microarrays. We considered >5% nuclear staining to be a positive result for GATA-3.**

Results: **Thirty-two percent of cases were GATA-3 positive. GATA-3 is almost exclusively expressed in ER+ patients and is also associated with lower tumor grade, older age at diagnosis, and the absence of Her2 overexpression. In univariate analysis, the presence of GATA-3 is a marker of good prognosis and predicted for superior breast cancer–specific survival, relapse-free survival, and overall survival. However, in multivariate models including patient age, tumor size, histologic grade, nodal status, ER status, and Her2 status, GATA-3 was not independently prognostic for these same outcomes. In the subgroups of ER+ patients treated with or without tamoxifen, GATA-3 was again nonprognostic for all outcomes.**

Discussion: **GATA-3 is a molecular marker that is highly associated with ER expression, but it does not seem to have prognostic value independent of ER, nor does it predict for response to tamoxifen among ER-positive patients.** (Cancer Epidemiol Biomarkers Prev 2008;17(2):365–73)

Introduction

Estrogen receptor (ER) status in breast cancer is used to estimate prognosis and to guide systemic treatment. ER positive status is generally associated with a more favorable prognosis (1, 2), but more importantly, it is a predictive marker of response to hormonal therapies such as tamoxifen and aromatase inhibitors (3, 4). However, ER-positive tumors represent a large and heterogeneous subgroup, and treatment decisions are most often based on clinicopathologic features such as tumor grade and lymph node status. Novel biomarkers can be used to further refine prognostic models, and some may be useful to predict response to adjuvant treatment.

Gene expression profiling studies have shown a consistent separation of ER-positive tumors into Luminal A and Luminal B subtypes, with Luminal B tumors having significantly worse outcome (5-7). In addition, from studies of tamoxifen in advanced breast cancer, it is estimated that up to 30% to 50% of ER-positive tumors are not tamoxifen responsive (8, 9). It would be of great clinical significance if Luminal subtype and response to tamoxifen could be determined with inexpensive and readily available biomarker laboratory tests such as immunohistochemistry.

The proteins of the GATA family are zinc-finger transcription factors involved in embryogenesis and development. GATA-3, in particular, plays a role in placental development, hematopoiesis, and adipogenesis (10-12). In breast epithelial cells, it has been proposed that GATA-3 acts to maintain a differentiated state. Aberrant expression of GATA-3 has been reported in breast cancer as well as in pancreatic and cervical cancers (12-14).

Studies using both gene expression profiling and immunohistochemistry have shown that GATA-3 expression in breast cancer is closely associated with the ER (6, 15, 16). In another study, *GATA-3* gene constructs were transfected into breast cancer cell lines, and it was found that many GATA-3–induced genes were also in the luminal gene cluster (17). In a recent tissue microarray–based study [Mehra et al. (16)], GATA-3 protein was found to have independent prognostic value in a cohort of 139 breast cancer patients. More specifically, low expression of GATA-3 predicted for breast cancer

Received 12/22/06; revised 10/15/07; accepted 11/16/07.

Grant support: National Cancer Institute Strategic Partnership to Evaluate Cancer Signatures program and an unrestricted educational grant from Sanofi-Aventis Canada.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article is available at http://www.gpec.ubc.ca/index.php?content=papers/GATA3.php.

Requests for reprints: David Voduc, BC Cancer Agency, 600 West 10th Avenue, Vancouver, British Columbia, Canada V5Z 4E6. Phone: 604-707-0595. E-mail: dvoduc@bccancer.bc.ca

doi:10.1158/1055-9965.EPI-06-1090

relapse. They also concluded that low GATA-3 levels could identify a subgroup of ER-positive tumors with a greater risk of relapse. This work suggests that GATA-3 may be useful in distinguishing Luminal A and Luminal B subtype tumors. Herein we attempt to confirm these findings with the immunohistochemical analysis of GATA-3 in a much larger cohort of breast cancer patients, with adjuvant treatment data and long follow-up.

Materials and Methods

Study Population. Our original study cohort consisted of 4,444 breast cancer patients diagnosed between January 1986 and September 1992. This represents 34% of all patients diagnosed with breast cancer in the province of British Columbia during this time period. This large, well-characterized cohort was derived from a consecutive series of patients who were referred to the BC Cancer Agency for consultation and had tumor samples sent to a central laboratory at the Vancouver General Hospital for ER testing. For all of these patients, we have available detailed demographic and outcome data as well as formalin-fixed paraffin-embedded primary tumor samples for immunohistochemical analysis. Patients with *in situ* disease, recurrent disease, metastatic disease at presentation, and male breast cancer were excluded from analysis. Available clinical information includes age, histology, tumor grade, tumor size, lymph node status, type of local and adjuvant systemic therapy, and dates of first recurrence and death. A portion of this cohort of patients was recently used in a population study validating the on-line breast cancer prognostic calculator ADJUVANT! Online (18). This study was approved by the Clinical Research Ethics Board of the University of British Columbia and the BC Cancer Agency.

Table 1. Summary of clinicopathologic variables

Clinicopathologic variables	Total (%)
Total	3,119 (100)
Age (y)	
<40	234 (7)
40-49	662 (21)
50-65	1,112 (36)
>65	1,111 (36)
Size (cm)	
<2	1,586 (51)
2-5	1,359 (44)
>5	149 (5)
Unknown	25 (<1)
Grade	
1 or 2	1,341 (43)
3	1,649 (53)
Unknown	129 (4)
Nodal status	
Negative	1,741 (56)
Positive	1,369 (44)
Unknown	9 (<1)
ER status	
Negative	618 (20)
Positive	2,421 (78)
Unknown	80 (3)
Her2 status	
Negative	2,627 (84)
Positive	413 (13)
Unknown	79 (3)

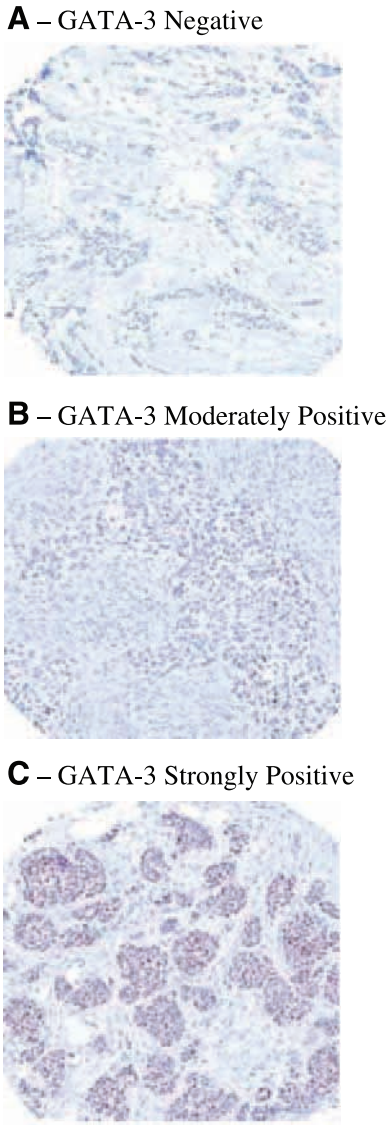

Figure 1. GATA-3 immunostaining was scored semiquantitatively; tumors with <5% nuclei stained were considered GATA-3 negative (**A**), tumors with 5% to 20% nuclei stained were moderately positive (**B**), and tumors with >20% nuclei stained were strongly positive (**C**).

Tissue Microarrays and Immunohistochemistry. The Vancouver Hospital Estrogen Receptor laboratory retained single archival tumor blocks from each case in this patient cohort. The material had been frozen before neutral buffered formalin fixation. H&E-stained slides from these blocks were reviewed by two pathologists to identify areas of invasive breast cancer. Cores of 0.6 mm were extracted from the tumor blocks and used to construct a tissue microarray as previously described (19, 20).

Using a single core per case, 17 tissue microarray blocks were required to represent the series. Four-micrometer sections were stained for GATA-3 using the Ventana Systems Discovery XT automated immunostainer. Slides were deparaffinized and incubated with

Table 2. Comparison of univariate hazard ratios using BCSS, RFS, DRFS, and OS for the standard clinicopathologic variables in this study

Variable	Difference in BCSS at 10 y (%)	Breslow test *P*	Univariate hazard ratio
BCSS			
Age (y)			
40-49 vs <40	75 vs 62	5.6e−04	0.66
50-65 vs <40	75 vs 62		0.68
>65 vs <40	75 vs 62		0.69
Size (cm)			
2-5 vs <2	66 vs 83	8.2e−34	1.94
>5 vs <2	52 vs 83		3.19
Grade			
3 vs 1/2	67 vs 83	5.4e−27	1.98
Nodal status			
Pos. vs neg.	61 vs 84	3.7e−51	2.69
ER status			
Pos. vs neg.	77 vs 63	2.3e−17	0.63
Her2 status			
Pos. vs neg.	60 vs 76	1.0e−15	1.76
RFS			
Age (y)			
40-49 vs <40	62 vs 53	3.4e−05	0.72
50-65 vs <40	64 vs 53		0.68
>65 vs <40	64 vs 53		0.65
Size (cm)			
2-5 vs <2	57 vs 70	6.1e−25	1.59
>5 vs <2	42 vs 70		2.53
Grade			
3 vs 1/2	56 vs 71	3.1e−23	1.70
Nodal status			
Pos. vs neg.	51 vs 73	5.8e−42	2.19
ER status			
Pos. vs neg.	65 vs 56	2.2e−12	0.73
Her2 status			
Pos. vs neg.	51 vs 65	6.5e−13	1.56
DRFS			
Age (y)			
40-49 vs <40	69 vs 57	3.9e−04	0.67
50-65 vs <40	68 vs 57		0.68
>65 vs <40	69 vs 57		0.66
Size (cm)			
2-5 vs <2	60 vs 77	2.6e−33	1.90
>5 vs <2	46 vs 77		2.92
Grade			
3 vs 1/2	61 vs 77	5.8e−26	1.86
Nodal status			
Pos. vs neg.	53 vs 79	5.8e−58	2.71
ER status			
Pos. vs neg.	70 vs 60	3.9e−12	0.70
Her2 status			
Pos. vs neg.	55 vs 70	2.0e−24	1.66
OS			
Age (y)			
40-49 vs <40	73 vs 61	4.9e−29	0.68
50-65 vs <40	68 vs 61		0.94
>65 vs <40	51 vs 61		1.63
Size (cm)			
2-5 vs <2	56 vs 71	4.7e−26	1.60
>5 vs <2	46 vs 71		1.98
Grade			
3 vs 1/2	56 vs 71	1.1e−19	1.48
Nodal status			
Pos. vs neg.	52 vs 71	1.3e−38	1.90
ER status			
Pos. vs neg.	64 vs 56	1.2e−7	0.84
Her2 status			
Pos. vs neg.	51 vs 64	1.7e−9	1.39

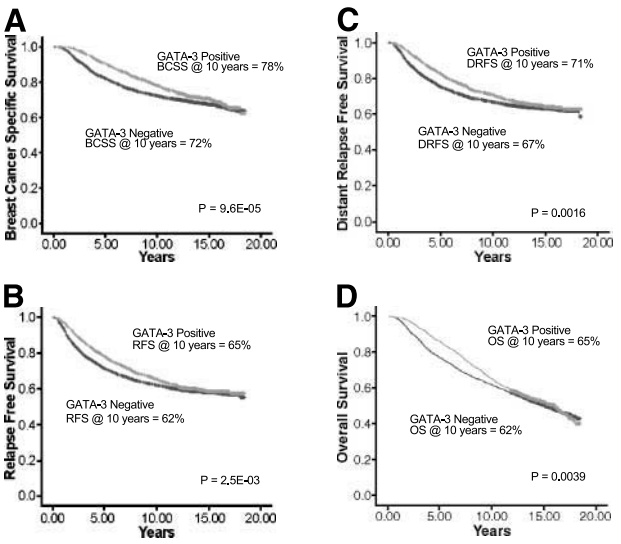

Figure 2. GATA-3–positive versus GATA-3–negative Kaplan-Meier plots in the whole patient cohort, for the outcomes BCSS (**A**), RFS (**B**), DRFS (**C**), and OS (**D**). In the whole patient cohort, and for all outcomes, the presence of GATA-3 is a weaker marker of good prognosis. The difference in 10-y BCSS is 78% versus 72% (P = 9.6e−05; **A**); 10-y RFS, 65% versus 62% (P = 2.5e−03; **B**); 10-y DRFS, 71% versus 67% (P = 0.0016; **B**); and 10-y OS, 65% versus 62% (P = 0.0039; **D**).

EDTA buffer (pH 8.0) for antigen retrieval. The slides were incubated with anti–GATA-3 antibodies for 32 min (Santa Cruz Biotechnology, Inc.; HG3-31 mouse monoclonal antibody, 1:20), and then with secondary antibody (Ventana universal secondary antibody) for an additional 32 min. Slides were counterstained with hematoxylin, rinsed with soap solution, and dehydrated through graded ethanol. The slides were then cleared in xylene and coverslipped with Cytoseal XYL (Richard-Allan Scientific). Stained tissue microarray slides were digitally scanned and linked to a relational database, and the primary image data are available for public review.[4]

Scoring of GATA-3 immunostaining was semiquantitative using digital images (Fig. 1). The scoring system used by Mehra et al. was based on a combination of percent positive nuclei and staining intensity, but cases were subsequently binarized into low and high GATA-3 expression. For ease of reproducibility, we limited interpretation to the percentage of positive nuclei and applied the scoring system published by van de Rijn et al. (21). Staining was considered negative (0) if <5% of nuclei were stained above background, moderate (1+) if 5% to 20% of nuclei were stained, and strong (2+) if >20% were stained. For statistical analysis, we dichotomized GATA-3 staining into negative (0) and positive (1+ or 2+). We also excluded cases for which it was not possible to assign a score to the immunostaining (insufficient invasive tumor in the core, tissue core cut through, or tissue disc lost or folded during sectioning).

[4] http://www.gpecimage.ubc.ca

Statistical Analysis. Statistical analysis was done using SPSS 14.0. In univariate analysis, relapse-free survival (RFS), breast cancer–specific survival (BCSS), distant relapse-free survival (DRFS), and overall survival (OS) were estimated using Kaplan-Meier curves. Significant differences in survival were assessed with Breslow tests. For OS, any death was considered an event, and patients were censored at the time of last follow-up. For RFS, an event was defined as any breast cancer relapse (locoregional or distant), and patients were censored at the date of death or the date of last follow-up. For DRFS, only distant relapses were considered events. For BCSS, an event was defined as a breast cancer death, and patients were censored at the time of a non–breast cancer death or at the date of last follow-up. For all outcomes, survival time was calculated from the date of surgery to the date of an event or date that the patient was censored. Six patients with unknown cause of death were excluded from RFS, BCSS, and DRFS analysis (but they were included in OS analysis). Cox proportional hazards models were used to calculate adjusted hazard ratios accounting for covariates. Pearson χ^2 and the Mann-Whitney tests were used to measure the association of GATA-3 status with common pathologic variables. All statistical tests were two sided. The intent of this study was to validate preexisting hypotheses on GATA-3 using a much larger cohort of patients; in addition, we carried out relatively few statistical tests on a large number of patients with long-term follow-up. Consequently, we did not apply corrections for multiple comparisons. The P value for statistical significance in this study is 0.05.

Missing GATA-3 Data. There were a relatively large proportion of cases with missing GATA-3 data, attributable in part to the practical necessity of using a single core per case in this very large tissue microarray series. Variations in the thickness of the source blocks led to cut-through and loss of some individual cores, and some were also lost during staining and sectioning. To ensure accurate scoring, we did not assign a GATA-3 score for cores containing <50 definite invasive cancer cells. We did carry out GATA-3 staining on a smaller tissue microarray (n = 413) with duplicate cores and found 90% concordance between duplicate cores.

With the exclusions stated previously (*in situ* disease, male breast cancer, and patients presenting with recurrent or metastatic disease), 4,049 patients remained in our patient cohort. Nine hundred thirty (23%) tissue microarray cores were considered uninterpretable for GATA-3. The size of the cohort with available GATA-3 data was 3,119.

Missing GATA-3 data points were not significantly associated with age, lymph node status, ER status, or Her2 status (full analysis available within the online Supplementary data, Section 1). There was a statistically significant association between missing GATA-3 data and grade 1/2 tumors and tumors >5 cm; however, the absolute differences were small. In the whole patient cohort, the proportion of cases with missing GATA-3 data was 23%. In grade 1/2 tumors the proportion was 25%, versus 20% in grade 3 tumors (Pearson χ^2 test, P = 3.0e−04). The proportion of missing GATA-3 data was 32% in tumors >5 cm, 20% in tumors 2 to 5 cm, and 24% in tumors <2 cm (Pearson χ^2 test, P = 3.7 e−05). In

consideration of these results, it is reasonable to conclude that there is a bias in our methods leading to very large tumors being underrepresented. It is possible that areas of necrosis observed in large tumors lead to sampling error during core extraction and tissue microarray construction. However, it should be noted that tumors >5 cm represent only 5.5% of our whole study population. Although these large tumors are associated with an inferior prognosis, we found no difference in survival between patients with missing GATA-3 data and those with GATA-3 data (BCSS at 10 years, 74% versus 74%; Breslow test, $P = 0.57$).

Results

Patient Demographics and Pathologic Data. In our cohort of 3,119 patients, the mean age at diagnosis was 59 years and the median follow-up for BCSS is 12.6 years (follow-up is defined as the time of diagnosis to time of event or last follow-up). The range for follow-up is 1 month to 18.5 years. The median tumor size was 2.5 cm. Fifty-three percent of patients had grade 3 tumors, 44% were node positive, and 78% were ER positive (Table 1). Forty-two percent had breast-conserving surgery and 58% had mastectomy; adjuvant radiation therapy was given to 55%. Fifty-eight percent received adjuvant systemic therapy (either chemotherapy or hormonal therapy). Outcome was last updated June 30, 2004, and to that date there were a total of 914 (29%) breast cancer deaths, 1,202 (39%) relapses, 1,043 (33%) distant relapses, and 1,556 (50%) total deaths.

GATA-3 Immunostaining. Of 3,119 breast cancer tumors examined with interpretable GATA-3 staining, 2,140 (68%) cases were GATA-3 negative, 646 (21%) exhibited moderate staining, and 333 (11%) were strongly positive. Once GATA-3 data were dichotomized, 68% of cases were GATA-3 negative and 32% GATA-3 positive.

Associations with Known Pathologic Factors. In our patient cohort, the common clinicopathologic variables including patient age at diagnosis, tumor size, histologic grade, nodal status, ER status, and Her2 status were all statistically significant predictors of BCSS, RFS, DRFS, and OS (Table 2) on univariate analysis.

There was a strong association between GATA-3 protein expression and positive ERs (Pearson χ^2 test, $P = 2.1e{-}67$). Seventy-eight percent of this tissue microarray consists of ER-positive cases, and among the ER-positive cases 39% were also GATA-3 positive. In contrast, among the ER-negative cases, only 2.6% were GATA-3 positive. Overall, 98% of GATA-3–positive cases were also ER positive (full cross-tabulations with numbers and frequencies available in Supplementary data, Section 2).

There was a linear association between age at diagnosis and GATA-3 expression (Mann-Whitney test, $P = 3.6e{-}05$), with GATA-3 present in a greater proportion of older patients. However, the actual difference was small; the mean age in GATA-3–positive cases was 60 years, versus 58 years in GATA-3–negative cases. There was also a linear association between GATA-3 expression and smaller tumor size (Mann-Whitney test, $P = 0.002$). Again, the actual difference was small. The mean tumor size in GATA-3–positive cases was 2.3 cm, versus 2.6 cm in GATA-3–negative cases.

There was a strong association between GATA-3–positive cases and grade 1/2 tumors (Pearson χ^2 test, $P = 1.1e{-}16$) and absence of Her2 overexpression (Pearson χ^2 test, $P = 9.7e{-}13$). We did not find a significant association between GATA-3 and lymph node status (Pearson χ^2 test, $P = 0.47$).

Univariate Survival Analysis. In this study, we carried out survival analysis for multiple outcomes; however, the results from GATA-3 survival analysis were consistent throughout all outcomes analyzed. In this patient cohort, the presence of GATA-3 was a statistically significant marker of good prognosis for all outcomes including BCSS, RFS, DRFS, and OS (Fig. 2). The BCSS at 10 years was 78% for GATA-3–positive cases versus 72% for negative cases (Breslow test, $P = 9.6e{-}05$). The difference in 10-year RFS, DRFS, and OS was 3%, 4%, and 3% respectively. However, within the ER-positive subgroup, GATA-3 did not have statistically significant prognostic value for any of these outcomes (Fig. 3). The BCSS at 10 years was 78% versus 76% ($P = 0.26$). Our results show a difference in BCSS of +1.9% for GATA-3–positive cases, with the 95% confidence interval between +5.5% and −1.7%. The difference in 10-year RFS, DRFS, and OS in the ER-positive subgroup was 1%, 2%, and 1%, respectively.

We also analyzed the prognostic significance of GATA-3 in the patient subgroup that was ER positive and did not receive adjuvant tamoxifen and in the ER-positive subgroup that was treated with tamoxifen

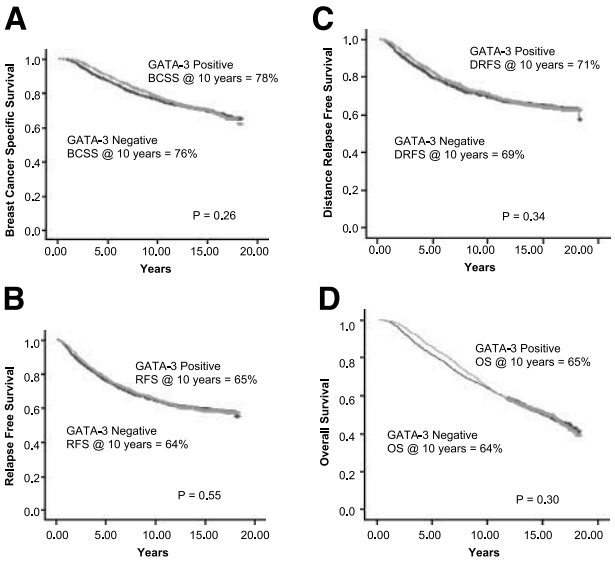

Figure 3. GATA-3–positive versus GATA-3–negative Kaplan-Meier plots in the ER-positive subgroup, for the outcomes BCSS (**A**), RFS (**B**), DRFS (**C**), and OS (**D**). In the ER-positive subgroup, and for all outcomes, GATA-3 is not a marker of prognosis. The difference in 10-y BCSS is 78% versus 76% ($P = 0.26$; **A**); 10-y RFS, 65% versus 64% ($P = 0.55$; **B**); 10-y DRFS, 71% versus 69% ($P = 0.34$; **C**); and 10-y OS, 65% versus 64% ($P = 0.30$; **D**).

Table 3. Cox proportional hazard regression analysis for BCSS, RFS, DRFS, and OS

	Multivariate hazard ratio			
	Whole cohort	ER positive only	ER positive + tamoxifen	ER positive − tamoxifen
BCSS				
Sample size	3,114	2,416	917	954
Age at diagnosis (y)				
40-49 vs <40	0.76 $P = 0.035$	0.72 $P = 0.057$	NA*	1.65 $P = 0.41$
50-65 vs <40	0.84 $P = 0.16$	0.82 $P = 0.23$	NA	2.04 $P = 0.23$
>65 vs <40	0.93 $P = 0.54$	0.88 $P = 0.44$	NA	1.75 $P = 0.35$
Tumor size (cm)				
2 < size ≤5 vs ≤2	1.60 $P = 5.7e{-10}$	1.65 $P = 1.4e{-8}$	1.57 $P = 7.1e{-4}$	1.65 $P = 3.5e{-5}$
>5 vs ≤2	2.25 $P = 2.9e{-8}$	3.13 $P = 2.1e{-12}$	4.58 $P = 9.8e{-9}$	0.75 $P = 0.78$
Grade				
Grade 3 vs grade 1/2	1.54 $P = 2.0e{-9}$	1.50 $P = 1.8e{-6}$	1.52 $P = 1.3e{-3}$	1.44 $P = 0.025$
Nodal status				
Positive vs negative	2.35 $P = 5.6e{-31}$	2.35 $P = 1.4e{-22}$	1.57 $P = 1.1e{-6}$	5.15 $P = 7.0e{-10}$
ER status				
Positive vs negative	0.79 $P = 0.011$	NA	NA	NA
Her2 status				
Positive vs negative	1.36 $P = 0.0010$	1.59 $P = 1.7e{-4}$	1.76 $P = 0.0016$	1.80 $P = 0.023$
GATA-3 status				
Positive vs negative	1.01 $P = 0.90$	1.03 $P = 0.69$	0.90 $P = 0.44$	1.29 $P = 0.12$
RFS				
Sample size	3,114	2,416	917	954
Age at diagnosis (y)				
40-49 vs <40	0.81 $P = 0.075$	0.75 $P = 0.055$	NA*	1.08 $P = 0.84$
50-65 vs <40	0.80 $P = 0.045$	0.73 $P = 0.029$	NA	1.08 $P = 0.83$
>65 vs <40	0.82 $P = 0.082$	0.70 $P = 0.016$	NA	0.89 $P = 0.74$
Tumor size (cm)				
2< size ≤5 vs ≤2	1.37 $P = 1.1e{-6}$	1.40 $P = 7.1e{-6}$	1.49 $P = 8.9e{-4}$	1.58 $P = 5.3e{-4}$
>5 vs ≤2	1.91 $P = 2.3e{-7}$	2.34 $P = 1.5e{-8}$	4.30 $P = 3.8e{-9}$	1.27 $P = 0.68$
Grade				
Grade 3 vs grade 1/2	1.41 $P = 3.4e{-7}$	1.38 $P = 1.0e{-5}$	1.49 $P = 4.8e{-4}$	1.19 $P = 0.17$
Nodal status				
Positive vs negative	1.97 $P = 8.2e{-27}$	1.92 $P = 2.8e{-19}$	2.14 $P = 2.2e{-8}$	4.35 $P = 7.8e{-11}$
ER status				
Positive vs negative	0.90 $P = 0.18$	NA	NA	NA
Her2 status				
Positive vs negative	1.31 $P = 0.0016$	1.40 $P = 2.5e{-3}$	1.69 $P = 0.0015$	1.33 $P = 0.20$
GATA-3 status				
Positive vs negative	1.01 $P = 0.85$	1.04 $P = 0.63$	0.97 $P = 0.81$	1.09 $P = 0.48$
DRFS				
Sample size	3,114	2,416	917	954
Age at diagnosis (y)				
40-49 vs <40	0.78 $P = 0.052$	0.71 $P = 0.036$	NA*	1.06 $P = 0.93$
50-65 vs <40	0.85 $P = 0.17$	0.79 $P = 0.13$	NA	1.30 $P = 0.57$

(*Continued on the following page*)

Table 3. Cox proportional hazard regression analysis for BCSS, RFS, DRFS, and OS (Cont'd)

	Multivariate hazard ratio			
	Whole cohort	ER positive only	ER positive + tamoxifen	ER positive − tamoxifen
>65 vs <40	0.88 P = 0.29	0.78 P = 0.12	NA	1.07 P = 0.88
Tumor size (cm)				
2< size ≤5 vs ≤2	1.58 P = 1.1e−10	1.63 P = 2.7e−9	1.65 P = 5.6e−5	1.90 P = 2.2e−5
>5 vs ≤2	2.13 P = 8.2e−9	2.84 P = 2.0e−11	4.45 P = 2.9e−9	0.59 P = 0.60
Grade				
Grade 3 vs grade 1/2	1.48 P = 5.8e−8	1.46 P = 1.7e−6	1.46 P = 0.0020	1.43 P = 0.017
Nodal status				
Positive vs negative	2.39 P = 4.9e−36	2.34 P = 6.0e−26	2.18 P = 4.6e−8	5.23 P = 2.0e−11
ER status				
Positive vs negative	0.88 P = 0.13	NA	NA	NA
Her2 status				
Positive vs negative	1.34 P = 0.0011	1.44 P = 0.0020	1.68 P = 0.0021	1.53 P = 0.087
GATA-3 status				
Positive vs negative	1.00 P = 0.98	1.02 P = 0.84	0.918 P = 0.48	1.08 P = 0.62
OS				
Sample size	3,119	2,421	919	957
Age at diagnosis (y)				
40-49 vs <40	0.74 P = 0.014	0.69 P = 0.025	NA*	1.17 P = 0.74
50-65 vs <40	1.07 P = 0.53	1.09 P = 0.57	NA	2.36 P = 0.060
>65 vs <40	1.97 P = 9.1e−10	2.00 P = 4.2e−6	NA	4.83 P = 5.2e−4
Tumor size (cm)				
2< size ≤5 vs ≤2	1.42 P = 5.6e−10	1.45 P = 5.8e−9	1.45 P = 8.2e−5	1.60 P = 9.2e−6
>5 vs ≤2	1.81 P = 4.5e−7	2.49 P = 5.6e−11	3.15 P = 2.2e−7	1.25 P = 0.71
Grade				
Grade 3 vs grade 1/2	1.33 P = 6.9e−7	1.29 P = 4.5e−5	1.24 P = 0.019	1.16 P = 0.15
Nodal status				
Positive vs negative	1.77 P = 1.9e−25	1.75 P = 3.1e−19	1.79 P = 3.7e−8	3.25 P = 2.6e−9
ER status				
Positive vs negative	0.88 P = 0.075	NA	NA	NA
Her2 status				
Positive vs negative	1.27 P = 0.0023	1.45 P = 1.3e−4	1.54 P = 0.0026	1.44 P = 0.041
GATA-3 status				
Positive vs negative	0.99 P = 0.82	1.00 P = 0.97	0.92 P = 0.40	1.12 P = 0.29

*Hazard ratio not available for age groups in the ER positive + tamoxifen subgroup because there were no patients diagnosed at age <40 y in this subgroup.

(full results and K-M plots available in Supplementary data, Section 3). In the no adjuvant tamoxifen subgroup, GATA-3 was not prognostic for any outcome (i.e., BCSS at 10 years, 85% versus 85%; P = 0.55). Similarly, in the subgroup treated with adjuvant tamoxifen, GATA-3 was again not prognostic for all outcomes (i.e., BCSS at 10 years, 76% versus 71%; P = 0.10).

Multivariate Survival Analysis. Using a Cox proportional hazards model including all patients, tumor size, histologic grade, Her2 status, and nodal status were all independent predictors of BCSS, RFS, DRFS, and OS (Table 3). In comparison with baseline, certain age groups were prognostic for BCSS, RFS, and OS. ER status was independently prognostic only for BCSS. GATA-3 was not independently prognostic for any of the outcomes measured. We repeated the Cox proportional hazards model for the ER-positive subgroup and the ER-positive subgroups treated with and without adjuvant tamoxifen. The results were similar to those of the

whole cohort, with GATA-3 not independently prognostic in any subgroup, for any outcome.

Discussion

GATA-3 is a highly conserved protein that plays a critical role in development and cellular differentiation (10, 12). Usary et al. carried out mutational analysis of GATA-3 in human breast tumors and proposed that GATA-3 is involved in luminal differentiation. It was also suggested that high expression of GATA-3 in breast luminal cells is ''normal,'' and a loss of GATA-3 expression may contribute to tumorigenesis.

The association between GATA-3 and ER was previously recognized using hierarchal clustering analysis of gene expression data from 34 primary breast carcinomas (22) and subsequently confirmed on a separate study of 78 breast cancer tumors (6). This study from Sorlie et al. found an ER gene cluster including *ERα* and *GATA-3*; also in this gene cluster were *X-box binding protein 1, trefoil factor 3, hepatocyte nuclear factor 3α,* and *LIV-1*. Expression levels of this gene cluster separated the Luminal/ER-positive tumors into subgroups. The good prognosis Luminal A group exhibited the highest expression of the *ER* gene cluster, whereas the Luminal B/C group had low/moderate expression and was associated with a worse prognosis.

Mehra et al. (16) carried out a meta-analysis of four breast cancer microarray expression profile data sets (including the previously mentioned data set from Sorlie et al.) totaling 305 breast tumor samples. *GATA-3* mRNA levels were prognostic for RFS, with higher levels associated with favorable outcome. *GATA-3* mRNA levels were lower in ER-negative and high-grade tumors.

In addition to the meta-analysis, these researchers constructed a tissue microarray from 139 consecutive single institution breast cancer patients and GATA-3 levels were assessed by immunohistochemistry. Their GATA-3 score was based on both intensity of staining and percentage of nuclei stained, but a binarized score was used for analysis. Mehra et al. found that low levels of GATA-3 were associated with larger tumor size, positive lymph node status, higher grade, ER positive status, and Her2 negative status. Overall, high GATA-3 expression was a marker of good prognosis for breast cancer RFS and OS, and in multivariate analysis GATA-3 independently predicted for superior RFS.

Using an easily reproducible system for scoring GATA-3 immunostaining, our study of GATA-3 protein expression in a much larger cohort of breast cancer tumors confirms some of these previously published conclusions. We found that GATA-3 is almost exclusively expressed in association with ER (P = 2.1e−67), with 98% of GATA-3–positive cases also being ER positive. In our study cohort, among ER-positive cases, 39% are GATA-3 positive. This number is similar to the results of Mehra et al., who found that 46% of ER-positive cases were GATA-3 positive. In addition, we also confirm that GATA-3 is associated with favorable prognostic features including older age at diagnosis, lower histologic grade, and Her2 negative status. However, in our study, GATA-3 levels were not significantly associated with lymph node status, and this is discordant with the results of Mehra et al.

In univariate analysis of the entire cohort, we found that the presence of GATA-3 is a relatively weak marker of good prognosis for all outcomes including OS and RFS (the outcomes presented by Mehra et al.). This finding is consistent with the close association between GATA-3 and ER (ER is a stronger marker of good prognosis). However, Mehra et al. reported that high GATA-3 protein expression predicted for superior RFS in univariate analysis of the ER-positive subgroup and in multivariate analysis of their whole cohort. In our study, GATA-3 was not significantly prognostic in the ER-positive subgroup for any of the survival outcomes reported, nor did it have independent prognostic significance in our multivariate models. Because of its lack of prognostic significance within the ER-positive subgroup, immunohistochemical assay of GATA-3 is unlikely to be useful in distinguishing the Luminal A from Luminal B biological subtypes. In the ER patients, our results show a difference in 10-year BCSS of 2% between GATA-3–positive and GATA-3–negative cases, and the 95% confidence interval excludes an absolute difference of >5.5%. Thus, GATA-3 is not a clinically useful prognostic marker in breast cancer given that ER status is already routinely obtained.

It has also been hypothesized that GATA-3 could be a predictive biomarker for tamoxifen responsiveness in ER-positive tumors (23). Parikh et al. carried out GATA-3 immunohistochemistry on 14 tumor samples from patients determined to be hormone unresponsive (based on clinical progression or early recurrence on hormonal therapy). They found that the hormone-unresponsive patients were more likely to have low expression of GATA-3 compared with the 14 hormone-responsive controls. Their conclusion that GATA-3 is predictive of tamoxifen response is not supported by our study. We carried out Kaplan-Meier analyses of GATA-3 expression among ER-positive patients receiving adjuvant tamoxifen and in ER-positive patients not receiving any adjuvant systemic therapy. Although evidence of predictive effect in a biomarker is best shown in the context of a randomized clinical trial, we can make some useful observations in our cohort.

If GATA-3 status were predictive of tamoxifen response, then we would expect to see GATA-3 having a reasonably strong prognostic effect in the ER-positive group treated with tamoxifen. Similarly, this prognostic effect would be reduced, or absent, in the ER-positive group that did not receive tamoxifen. However, in both the tamoxifen-treated and untreated subgroups, GATA-3 did not have prognostic value for any of the outcomes presented in this study. Consequently, there is no evidence from our study that GATA-3 is a predictive biomarker for response to hormonal therapies. Consistent with our finding is the exclusion of GATA-3 from the final 21-gene panel used to calculate the Recurrence Score, a multigene molecular test that is independently prognostic in a large series of node-negative, ER-positive, and tamoxifen-treated breast cancer patients (24).

Conclusion

Using a large, well-annotated tissue microarray series, a clinically practical immunohistochemical assessment,

and strict statistical analysis, we find that GATA-3 is a breast cancer marker almost exclusively expressed among ER-positive tumors. Similar to the ER, it is associated with favorable prognostic features and is a univariate marker of good prognosis across multiple survival outcomes, including relapse, breast cancer death, and OS. GATA-3 does not have independent prognostic significance in multivariate Cox models incorporating the standard clinicopathologic variables. It is not prognostic, for any outcome, within the ER-positive subgroup and does not seem to predict for tamoxifen response in ER-positive patients.

Acknowledgments

We thank Challayne Kenney and Samuel Leung for technical assistance.

References

1. Fisher B, Redmond C, Fisher ER, et al. Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. J Clin Oncol 1988;6:1076–87.
2. Grann VR, Troxel AB, Zojwalla NJ, et al. Hormone receptor status and survival in a population-based cohort of patients with breast carcinoma. Cancer 2005;103:2241–51.
3. Early Breast Cancer Trialists' Collaborative G, Early Breast Cancer Trialists' Collaborative G. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials [see comment]. Lancet 2005;365:1687–717.
4. Grana G, Grana G. Adjuvant aromatase inhibitor therapy for early breast cancer: a review of the most recent data. J Surg Oncol 2006;93:585–92.
5. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
6. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
7. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100:10393–8.
8. Morgan LR, Jr., Schein PS, Woolley PV, et al. Therapeutic use of tamoxifen in advanced breast cancer: correlation with biochemical parameters. Cancer Treat Rep 1976;60:1437–43.
9. Osborne CK, Osborne CK. Tamoxifen in the treatment of breast cancer. N Engl J Med 1998;339:1609–18.
10. LaVoie HA. The role of GATA in mammalian reproduction. Exp Biol Med (Maywood) 2003;228:1282–90.
11. Tong Q, Tsai J, Tan G, et al. Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation. Mol Cell Biol 2005;25:706–15.
12. Usary J, Llaca V, Karaca G, et al. Mutation of GATA3 in human breast tumors. Oncogene 2004;23:7669–78.
13. Gulbinas A, Berberat PO, Dambrauskas Z, et al. Aberrant gata-3 expression in human pancreatic cancer. J Histochem Cytochem 2006;54:161–9.
14. Steenbergen RD, OudeEngberink VE, Kramer D, et al. Down-regulation of GATA-3 expression during human papillomavirus-mediated immortalization and cervical carcinogenesis. Am J Pathol 2002;160:1945–51.
15. Charafe-Jauffret E, Ginestier C, Monville F, et al. Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene 2006;25:2273–84.
16. Mehra R, Varambally S, Ding L, et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. Cancer Res 2005;65:11259–64.
17. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers [see comment]. J Clin Oncol 2006;24:1656–64.
18. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 2005;23:2716–25.
19. Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens [see comment]. Nat Med 1998;4:844–7.
20. Parker RL, Huntsman DG, Lesack DW, et al. Assessment of interlaboratory variation in the immunohistochemical determination of estrogen receptor status using a breast cancer tissue microarray. Am J Clin Pathol 2002;117:723–8.
21. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 2002;161:1991–6.
22. Bertucci F, Houlgatte R, Benziane A, et al. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. Hum Mol Genet 2000;9:2981–91.
23. Parikh P, Palazzo JP, Rose LJ, Daskalakis C, Weigel RJ. GATA-3 expression as a predictor of hormone response in breast cancer. J Am Coll Surg 2005;200:705–10.
24. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer [see comment]. N Engl J Med 2004;351:2817–26.

GATA-3 Expression in Breast Cancer Has a Strong Association with Estrogen Receptor but Lacks Independent Prognostic Value

David Voduc, Maggie Cheang and Torsten Nielsen

Cancer Epidemiol Biomarkers Prev 2008;17:365-373.