Differential expression of prospero homeobox 1 in cancers of the breast.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY 11730

Breast cancer affects women at relatively high frequency¹. We mined published microarray datasets^{2,3} to determine in an unbiased fashion and at the systems level genes most differentially expressed in the primary tumors of patients with breast cancer. We report here significant differential expression of the gene encoding prospero homeobox 1, PROX1, when comparing primary tumors of the breast to the tissue of origin, the normal breast. PROX1 was also differentially expressed in the tumor cells of patients with triple negative breast cancer. PROX1 mRNA was present at significantly lower quantities in tumors of the breast as compared to normal breast tissue. Analysis of human survival data revealed that expression of PROX1 in primary tumors of the breast was correlated with recurrence-free survival in patients with luminal A and luminal B subtype cancer, demonstrating a relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by molecular subtype. PROX1 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, PROX1, prospero homeobox 1, systems biology of breast cancer, targeted therapeutics in breast cancer.

Invasive breast cancer is diagnosed in over a quarter of a million women in the United States each year¹ and in 2018, breast cancer was the leading cause of cancer death in women worldwide⁴. While patients with localized breast cancer are provided a 99% 5-year survival rate, patients with regional breast cancer, cancer that has spread to lymph nodes or nearby structures, are provided an 86% 5-year survival rate^{5,6}. Patients with metastasis to distant sites, like the brain, are provided a 27% 5-year survival rate^{5,6}. Understanding how primary tumors are most transcriptionally different from the tissue from which they originate, the breast, can facilitate development of novel diagnostic and therapeutics to promote early detection and enhanced treatment, and contribute to efforts to prevent progression to metastatic stages. We mined published microarray data^{2,3} to understand at the transcriptome level and in an unbiased fashion genes most differentially expressed in primary tumors of the breast as compared to normal breast tissue. Prospero homeobox 1 emerged as among the most differentially expressed genes in cancer of the female breast.

Methods

We utilized datasets GSE109169² and GSE38959³ for this differential gene expression analysis of female breast cancer in conjunction with GEO2R. GSE109169 was generated using Affymetrix Human Exon 1.0 ST Array technology with $n=25$ normal breast tissue and $n=25$ tumors of the breast; analysis was performed using platform GPL5175. The tissues from this first dataset are paired tissues (25 tumors matching 25 breast tissues from 25 patients). GSE38959 was generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F technology with $n=13$ samples of normal mammary gland ductal cells and $n=30$ samples of tumor cells from patients with triple negative breast cancer; analysis was performed using platform GPL4133. The Benjamini and Hochberg method of p -value adjustment was used for ranking of differential expression but raw p -values were used to assess statistical significance of global differential expression. Log-transformation of data was auto-detected, and the NCBI generated category of platform annotation was used. A statistical test was performed to evaluate whether PROX1 expression was significantly different between primary breast tumors and breast tissue using a two-tailed t -test. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter online tool⁷ for correlation of PROX1 mRNA expression levels with recurrence-free survival (RFS) in $n=846$ patients with basal subtype cancer, $n=2277$ patients with luminal A subtype cancer, $n=1491$ patients with luminal B subtype cancer, and $n=315$ patients with HER2+ cancer.

Results

We performed discovery of genes associated with breast cancer in females by mining two independently published microarray datasets^{2,3}.

PROX1 is differentially expressed in primary tumors of the breast.

Comparison of 25 normal breast tissues to 25 tumors of the breast² from patients with early-onset breast cancer revealed that the gene encoding prospero homeobox 1, PROX1, was among the genes most differentially expressed in tumors of the breast in human breast cancer (Chart 1). When sorting each of the genes expressed in tumors of the breast based on significance of difference as compared to normal breast tissue, PROX1 ranked 236 out of 19076 total transcripts, equating to 98.8% differential expression (Chart 1). Differential expression of PROX1 in female breast cancer was statistically significant (Chart 1; $p=9.69E-13$).

Analysis of a second microarray dataset³, here in tumor cells of patients with triple negative breast cancer as compared to normal mammary gland ductal cells, again revealed significant differential expression of PROX1 in human breast cancer (Chart 2). When sorting each of the genes expressed in

the tumor cells of patients with triple negative breast cancer based on significance of difference as compared to normal mammary gland ductal cells, PROX1 ranked 17435 out of 45015 total transcripts, equating to 61.2% differential expression (Chart 2). Differential expression of PROX1 in the tumor cells of patients with triple negative breast cancer was statistically significant (Chart 2; $p=6.64E-02$). These data suggested that differential expression of PROX1 was not an artifact of a single microarray dataset, nor was it associated with one type of breast cancer, rather a general feature of cancers of the breast.

PROX1 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for PROX1 from the breast and from breast tumors to understand the magnitude and direction of PROX1 expression change. PROX1 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of PROX1 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). PROX1 was expressed at 6.25 ± 0.31 arbitrary units (AU) in normal breast tissue, while it was expressed at 5.46 ± 0.29 AU in tumors of the breast. We calculated a mean fold change of 0.87 in PROX1 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

PROX1 expression correlates with survival outcomes in luminal A and luminal B human breast cancer.

We performed Kaplan-Meier survival analysis in an attempt to correlate PROX1 mRNA expression levels with survival outcomes in patients with breast cancer, evaluating correlations between molecular subtype and recurrence-free survival. We observed a statistically significant correlation between PROX1 expression and recurrence-free survival (RFS) in patients with luminal A subtype breast cancer (Figure 2; log rank p -value: 0.0048 for recurrence-free survival, hazard ratio: 0.79 (0.67-0.93) (Fig. 2)). PROX1 mRNA levels were a negative prognostic indicator in luminal A subtype breast cancer patients. Median OS was 228.85 months for luminal A patients with low tumor expression of PROX1 while median OS was 216.66 months for luminal A patients with high tumor expression of PROX1 (Chart 3). We also observed a statistically significant correlation between PROX1 expression and recurrence-free survival (RFS) in patients with luminal B subtype breast cancer (Figure 2; log rank p -value: 0.0035 for recurrence-free survival, hazard ratio: 0.77 (0.64-0.92) (Fig. 2)). PROX1 mRNA levels were a positive prognostic indicator in luminal B subtype breast cancer patients. Median OS was 37.2 months for luminal B patients with low tumor expression of PROX1 while median OS was 52.32 months for luminal B patients with high tumor expression of PROX1 (Chart 3). PROX1 primary tumor expression was not correlated with recurrence-free survival in basal subtype breast cancer (Figure 2; log rank p -value: 0.6 for OS, hazard ratio: 1.06 (0.85-1.33) (Fig. 2)) or in HER2+ breast cancer (Figure 2; log rank p -value: 0.35 for OS, hazard ratio: 0.85 (0.59-1.2) (Fig. 2)).

Thus, through comparative transcriptome analysis of primary tumors of the breast and normal breast tissue, we found that differential expression and down-regulation of PROX1 was among the most significant transcriptional features in primary tumors from patients with breast cancer. PROX1 expression in primary tumors of the breast was correlated with recurrence-free survival in patients with luminal A and luminal B subtype disease, with mRNA levels of PROX1 a negative prognostic indicator for luminal A breast cancer patients and a positive prognostic indicator for luminal B breast cancer patients.

Discussion

Invasive breast cancer is a medical problem with a 27% 5-year survival rate for women whose disease has spread to distant sites^{5,6}. To facilitate understanding of the basic transcriptional differences between primary tumors of the breast and the tissues from which these tumors originate, normal breast

1 tissues, we performed comparative transcriptome analysis using two independently published microarray
2 datasets^{2,3}, providing evidence here that differential expression of prospero homeobox 1, encoded by
3 PROX1, is a defining transcriptional feature of human breast cancer: in early onset-breast cancer, and in
4 triple negative breast cancer. PROX1 was expressed at significantly lower levels in primary tumors from
5 patients with breast cancer as compared to normal breast tissue. Importantly, in patients with luminal A
6 subtype and luminal B subtype breast cancers, expression of PROX1 was correlated with recurrence-free
7 survival. Prospero homeobox 1 type B and the molecular processes to which it pertains to may be relevant
8 to the initiation or progression of human breast cancer.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

References

1. DeSantis, C.E., Ma, J., Goding Sauer, A., Newman, L.A. and Jemal, A., 2017. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians*, 67(6), pp.439-448.
2. Chang, J.W., Kuo, W.H., Lin, C.M., Chen, W.L., Chan, S.H., Chiu, M.F., Chang, I.S., Jiang, S.S., Tsai, F.Y., Chen, C.H. and Huang, P.H., 2018. Wild-type p53 upregulates an early onset breast cancer-associated gene GAS7 to suppress metastasis via GAS7–CYFIP1- mediated signaling pathway. *Oncogene*, 37(30), pp.4137-4150.
3. Komatsu, M., Yoshimaru, T., Matsuo, T., Kiyotani, K., Miyoshi, Y., Tanahashi, T., Rokutan, K., Yamaguchi, R., Saito, A., Imoto, S. and Miyano, S., 2013. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. *International journal of oncology*, 42(2), pp.478-506.
4. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6), pp.394-424.
5. ACS Cancer Facts & Figures 2019.
<https://www.cancer.net/cancer-types/breast-cancer/metastatic/statistics>.
6. Survival Rates for Breast Cancer. <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html>
7. Györfy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q. and Szallasi, Z., 2010. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*, 123(3), pp.725-731.

Rank: 236
Probe ID: 2379665
p-value: 9.69E-13
t: 9.41
B: 18.823578
Gene: PROX1
Gene name: prospero homeobox 1

Chart 1: PROX1 is differentially expressed in the primary tumors of patients with early-onset human breast cancer.

Rank of differential expression, probe ID, p-value of global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared, gene and gene name are listed in this chart.

Rank: 17435
Probe ID: 35357
p-value: 6.64E-02
t: 1.8811957
B: -5.042958
Gene: PROX1
Gene name: prospero homeobox 1

Chart 2: PROX1 is differentially expressed in the primary tumor cells of patients with triple negative breast cancer.

Rank of differential expression, probe ID, p-value of global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared, gene and gene name are listed in this chart.

Figure 1: PROX1 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of PROX1 in normal breast tissue (left) and in primary tumors of the breast (right) is graphically depicted with the result of a statistical test evaluating significance of difference in PROX1 expression between normal breast tissue and primary tumors of the breast, a *p*-value, is *p*<0.0001.

Figure 2: PROX1 expression correlates with recurrence-free survival in patients with luminal A and luminal B subtype breast cancer, but not in patients with basal subtype or HER2+ cancer.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival (RFS) for $n=846$ patients with basal subtype breast cancer (upper left), $n=315$ patients with HER2+ breast cancer (upper right), $n=2277$ patients with luminal A breast cancer (lower left), and $n=1491$ patients with luminal B breast cancer (lower right), stratified into two groups, based on low or high expression of PROX1 in patient primary tumors. The log rank p-value denoting statistical significance of difference in recurrence-free survival when comparing the two groups, as well as hazard ratio for this comparison is listed above. Listed below is the number of patients at risk (number of patients alive) per interval, after stratification based on PROX1 expression; in the first interval, number at risk is number of patients alive; in each subsequent interval, number at risk is the number at risk less those who have expired or are censored.

Luminal A patients

Low PROX1 expression: 228.85 months

High PROX1 expression: 216.66 months

Luminal B patients

Low PROX1 expression: 37.2 months

High PROX1 expression: 52.32 months

Chart 3: In luminal A subtype breast cancer, median recurrence-free survival is superior in patients with low primary tumor expression of PROX1, but in luminal B subtype breast cancer, median recurrence-free survival is superior in patients with high primary tumor expression of PROX1.

The recurrence-free survival of $n=2277$ luminal A subtype and $n=1491$ luminal B subtype breast cancer patients based on stratification into low or high tumor expression of PROX1 is listed in this chart.