Differential expression of teashirt zinc finger homeobox 2 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 teashirt zinc finger homeobox 2, TSHZ2, when comparing primary tumors of the breast to the tissue of origin, the normal breast. TSHZ2 was also differentially expressed in the tumor cells of patients with triple negative breast cancer. TSHZ2 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 TSHZ2 in primary tumors of the breast was correlated with overall survival in patients with basal subtype cancer, demonstrating a relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by molecular subtype. TSHZ2 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, TSHZ2, teashirt zinc finger homeobox 2, 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. Teashirt zinc finger homeobox 2 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. 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 TSHZ2 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 TSHZ2 mRNA expression levels with overall survival (OS) in $n=278$ patients with basal subtype cancer, $n=377$ patients with basal subtype cancer, $n=177$ patients with luminal B subtype cancer, and $n=111$ patients with luminal B cancer.

Results

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

TSHZ2 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 teashirt zinc finger homeobox 2, encoded by the TSHZ2 gene, 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, TSHZ2 ranked 58 out of 19076 total transcripts, equating to 99.7% differential expression (Chart 1). Differential expression of TSHZ2 in female breast cancer was statistically significant (Chart 1; $p=8.03E-16$).

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 TSHZ2 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, TSHZ2 ranked 1117 out of 45015 total transcripts, equating to 97.5% differential expression (Chart 2). Differential expression of TSHZ2 in the tumor cells of patients with triple negative breast cancer was statistically significant (Chart 2; $p=2.43E-07$). These data suggested that differential expression of TSHZ2 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.

TSHZ2 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for TSHZ2 from the breast and from breast tumors² to understand the magnitude and direction of TSHZ2 expression change. TSHZ2 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of TSHZ2 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). TSHZ2 was expressed at 9.82 ± 0.52 arbitrary units (AU) in normal breast tissue, while it was expressed at 7.70 ± 0.79 AU in tumors of the breast. We calculated a mean fold change of 0.78 in TSHZ2 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

TSHZ2 expression correlates with survival outcomes in basal subtype human breast cancer.

We performed Kaplan-Meier survival analysis in an attempt to correlate TSHZ2 mRNA expression levels with survival outcomes in patients with breast cancer, evaluating correlations between molecular subtype and overall survival. We observed a statistically significant correlation between TSHZ2 expression and overall survival (OS) in patients with basal subtype breast cancer (Figure 2; log rank p -value: 0.024 for overall survival, hazard ratio: 1.72 (1.07-2.78) (Fig. 2)). TSHZ2 mRNA levels were a negative prognostic indicator in basal subtype breast cancer patients. Median OS was, on average, markedly greater for basal subtype patients with low tumor expression of TSHZ2 as compared to basal subtype patients with high tumor expression of TSHZ2 (Chart 3). TSHZ2 primary tumor expression was not correlated with overall survival in luminal A subtype (Figure 2; log rank p -value: 0.89 for OS, hazard ratio: 0.97 (0.6-1.55) (Fig. 2)), HER2+ (Figure 2; log rank p -value: 0.33 for OS, hazard ratio: 0.71 (0.36-1.42) (Fig. 2)), or luminal B subtype breast cancer (Figure 2; log rank p -value: 0.43 for OS, hazard ratio: 0.78 (0.42-1.45) (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 TSHZ2 was among the most significant transcriptional features in primary tumors from patients with breast cancer. TSHZ2 expression in primary tumors of the breast was correlated with overall survival in patients with basal subtype cancers, with TSHZ2 primary tumor mRNA levels a negative prognostic indicator for basal subtype 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 tissues, we performed comparative transcriptome analysis using two independently published microarray datasets^{2,3}, providing evidence here that differential expression of teashirt zinc finger homeobox 2, encoded by TSHZ2, is a defining transcriptional feature of human breast cancer: in early onset-breast cancer, and in triple negative breast cancer. TSHZ2 was expressed at significantly lower levels in primary tumors from patients with breast cancer, as compared to normal breast. In patients with basal primary tumor expression of TSHZ2 was correlated with overall survival. Teashirt zinc finger homeobox 2 and the

molecular processes to which it pertains to may be relevant to the initiation or progression of human breast cancer.

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: 58
Probe ID: 3889624
p-value: 8.03E-16
t: 1.15E+01
B: 25.810306
Gene: TSHZ2
Gene name: teashirt zinc finger homeobox 2

Chart 1: TSHZ2 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: 1117
Probe ID: 10789
p-value: 2.43E-07
t: 6.0702188
B: 6.762579
Gene: TSHZ2
Gene name: teashirt zinc finger homeobox 2

Chart 2: TSHZ2 is differentially expressed in the 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: TSHZ2 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of TSHZ2 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 TSHZ2 expression between normal breast tissue and primary tumors of the breast, a *p*-value, is $p<0.0001$.

Figure 2: TSHZ2 expression correlates with overall survival in patients with basal breast cancer, but not in patients with luminal A, luminal B, or HER2+ cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=278$ patients with basal subtype breast cancer (upper left), $n=111$ patients with HER2+ breast cancer (upper right), $n=377$ patients with luminal A breast cancer (lower left), and $n=177$ patients with luminal B breast cancer (lower right), stratified into two groups, based on low or high expression of TSHZ2 in patient primary tumors. The log rank p-value denoting statistical significance of difference in overall 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 TSHZ2 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.