Differential expression of long intergenic non-coding RNA 01140 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 a long intergenic non-coding RNA known as LINC01140 when comparing primary tumors of the breast to the tissue of origin, the normal breast. LINC01140 was also differentially expressed in lymph node metastasis in human breast cancer. LINC01140 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 LINC01140 in primary tumors of the breast was correlated with overall survival in patients with basal-like breast cancer, demonstrating a relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by PAM50 molecular subtype. LINC01140 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, LINC01140, long intergenic non-coding RNA 01140, 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. long intergenic non-coding RNA 01140 emerged as among the most differentially expressed genes in cancer of the female breast.

Methods

We utilized datasets GSE42568² and GSE124648³ for this global differential gene expression analysis of female breast cancer. GSE42568 was generated using Affymetrix Human Genome U133 Plus 2.0 array technology with $n=17$ normal breast tissue biopsies and $n=104$ primary breast tumor biopsies from patients with breast cancer; analysis was performed using platform GPL570; the majority of patients whose tumors were analyzed were age 50 or older. GSE124648 was generated using Affymetrix Human Genome U133A array technology with $n=10$ normal breast tissues and $n=44$ lymph node metastases from patients with breast cancer; analysis was performed using platform GPL96. 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 LINC01140 expression was significantly different between primary breast tumors and normal breast tissue using a two-tailed t-test. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter online tool⁷ for correlation of LINC01140 mRNA expression levels with overall survival (OS) in $n=283$ patients with basal-like subtype cancer, $n=259$ patients with luminal A subtype cancer, $n=183$ patients with luminal B subtype cancer, $n=218$ patients with HER2+ cancer, and $n=15$ patients with normal-like subtype cancer.

Results

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

LINC01140 is differentially expressed in primary tumors of the breast.

Studying the global gene expression profiles of 104 breast cancers from patients aged 31 to 89 revealed that a long intergenic non-coding RNA, LINC01140, was among the 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, LINC01140 ranked 156 out of 54675 total transcripts, equating to 99.7% differential expression (Chart 1). Differential expression of LINC01140 in female breast cancer was statistically significant (Chart 1; $p=1.42E-28$).

Analysis of a second microarray dataset³, here studying global gene expression patterns in the lymph node metastases of 44 patients with breast cancer, again revealed significant differential expression of LINC01140 in human breast cancer (Chart 2). When sorting each of the genes expressed in the lymph

node metastases of patients with breast cancer based on significance of difference as compared to normal breast tissue, LINC01140 ranked 4619 out of 22283 total transcripts, equating to 79.3% differential expression (Chart 2). Differential expression of LINC01140 in the tumors of patients with breast cancer was statistically significant (Chart 2; $p=4.80E-03$). These data suggested that differential expression of LINC01140 was not an artifact of a single microarray dataset, nor was it strictly associated with primary tumors of the breast, rather a transcriptional feature representing a spectrum of transformation in human cancer.

LINC01140 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for LINC01140 from the breast and from breast tumors to understand the magnitude and direction of LINC01140 expression change. LINC01140 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of LINC01140 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). LINC01140 was expressed at 8.22 ± 2.04 arbitrary units (AU) in normal breast tissue, while it was expressed at 3.51 ± 1.05 AU in tumors of the breast. We calculated a mean fold change of 0.43 in LINC01140 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

LINC01140 expression correlates with survival outcomes in basal subtype human breast cancer.

We performed Kaplan-Meier survival analysis to study relationships between tumor LINC01140 mRNA expression levels and survival outcomes in patients with breast cancer. We observed a correlation between LINC01140 expression and overall survival (OS) in patients with basal-like (basal) subtype breast cancer which approached towards statistical significance (Figure 2; log rank p -value: 0.063 for overall survival, hazard ratio: 0.64 (0.4-1.03) (Fig. 2)). LINC01140 mRNA levels were a positive prognostic indicator in basal subtype breast cancer patients. Median OS was, on average, markedly shorter for basal patients with low tumor expression of LINC01140 as compared to basal patients with high tumor expression of LINC01140 (Chart 3). LINC01140 primary tumor expression was not correlated with overall survival in luminal A breast cancer (Figure 2; log rank p -value: 0.63 for OS, hazard ratio: 0.84 (0.42-1.68) (Fig. 2)), luminal B subtype breast cancer (Figure 2; log rank p -value: 0.84 for OS, hazard ratio: 0.94 (0.51-1.72) (Fig. 2)), in HER2+ breast cancer (Figure 2; log rank p -value: 0.86 for OS, hazard ratio: 0.96 (0.59-1.54) (Fig. 2)) or in patients with normal-like breast cancer (Figure 2; log rank p -value: 0.41 for OS, hazard ratio: 2.03 (0.37-11.29) (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 LINC01140 was among the most significant transcriptional features in primary tumors from patients with breast cancer. LINC01140 expression in primary tumors of the breast was correlated with overall survival in patients with basal subtype disease, with mRNA levels of LINC01140 a positive prognostic indicator for basal 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 tissues, we performed comparative transcriptome analysis using two independently published microarray datasets^{2,3}, providing evidence here that differential expression of long intergenic non-coding RNA 01140 is a defining transcriptional feature of human breast cancer: in patients diagnosed after age 50, and in the lymph node metastases of patients with breast cancer. LINC01140 was expressed at significantly lower

1 levels in primary tumors from patients with breast cancer as compared to normal breast tissue.
2 Importantly, in patients with basal subtype breast cancers, expression of LINC01140 was correlated with
3 overall survival. long intergenic non-coding RNA 01140 and the molecular processes to which it pertains
4 to may be relevant to the initiation or progression of human breast cancer.
5
6
7
8
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. Clarke, C., Madden, S.F., Doolan, P., Aherne, S.T., Joyce, H., O'driscoll, L., Gallagher, W.M., Hennessy, B.T., Moriarty, M., Crown, J. and Kennedy, S., 2013. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis*, 34(10), pp.2300-2308.
3. Sinn, B.V., Fu, C., Lau, R., Litton, J., Tsai, T.H., Murthy, R., Tam, A., Andreopoulou, E., Gong, Y., Murthy, R. and Gould, R., 2019. SET ER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ breast cancer*, 5(1), pp.1-8.
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: 156
Probe ID: 215039_at
p-value: 1.42E-28
t: 1.47E+01
B: 54.4376885
Gene: LINC01140
Gene name: long intergenic non-coding RNA 01140

Chart 1: LINC01140 is differentially expressed in the primary tumors of patients with 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: 4629
Probe ID: 215039_at
p-value: 4.80E-03
t: 2.938426
B: -2.922181
Gene: LINC01140
Gene name: long intergenic non-coding RNA 01140

Chart 2: LINC01140 is differentially expressed in the lymph node metastases of patients with 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.

Figure 1: LINC01140 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of LINC01140 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 LINC01140 expression between normal breast tissue and primary tumors of the breast, a *p*-value, listed above.

Figure 2: LINC01140 expression correlates with overall survival in patients with basal subtype breast cancer, but not in patients with luminal A, luminal B, HER2+ or normal-like subtype cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=283$ patients with basal-like breast cancer (upper left), $n=218$ patients with HER2+ breast cancer (upper right), $n=259$ patients with luminal A breast cancer (middle left), $n=183$ patients with luminal B breast cancer (middle right), and $n=15$ patients with normal-like breast cancer (below) stratified into two groups, based on low or high expression of LINC01140 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 LINC01140 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.