Differential expression of C2orf40 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 chromosome 2 open reading frame 40, C2orf40, when comparing primary tumors of the breast to the tissue of origin, the normal breast. C2orf40 was also differentially expressed in the tumor cells of patients with triple negative breast cancer. C2orf40 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 C2orf40 in primary tumors of the breast was correlated with recurrence-free survival in patients with luminal A subtype cancer, demonstrating a relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by molecular subtype. C2orf40 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, C2orf40, chromosome 2 open reading frame 40, 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. chromosome 2 open reading frame 40 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 C2orf40 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 C2orf40 mRNA expression levels with recurrence-free survival (RFS) in $n=417$ patients with basal subtype cancer, $n=952$ patients with luminal A subtype cancer, $n=465$ patients with luminal B subtype cancer, and $n=198$ 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}.

C2orf40 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 chromosome 2 open reading frame 40, encoded by the C2orf40 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, C2orf40 ranked 13 out of 19076 total transcripts, equating to 99.9% differential expression (Chart 1). Differential expression of C2orf40 in female breast cancer was statistically significant (Chart 1; $p=9.66E-20$).

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 C2orf40 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, C2orf40 ranked 852 out of 45015 total transcripts, equating to 98.1% differential expression (Chart 2). Differential expression of C2orf40 in the tumor cells of patients with triple negative breast cancer was statistically significant (Chart 2; $p=6.22E-08$). These data suggested that differential expression of C2orf40 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.

C2orf40 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for C2orf40 from the breast and from breast tumors² to understand the magnitude and direction of C2orf40 expression change. C2orf40 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of C2orf40 in primary breast tumors was statistically significant (Figure 1; $p<0.0001$). C2orf40 was expressed at 8.12 ± 0.68 arbitrary units (AU) in normal breast tissue, while it was expressed at 6.04 ± 0.28 AU in tumors of the breast. We calculated a mean fold change of 0.74 in C2orf40 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

C2orf40 expression correlates with survival outcomes in luminal A subtype human breast cancer.

We performed Kaplan-Meier survival analysis in an attempt to correlate C2orf40 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 C2orf40 expression and recurrence-free survival (RFS) in patients with luminal A subtype breast cancer (Figure 2; log rank p -value: 0.00021 for recurrence-free survival, hazard ratio: 0.63 (0.49-0.81) (Fig. 2)). C2orf40 mRNA levels were a positive prognostic indicator in luminal A subtype breast cancer patients. Median RFS was 47 months for luminal A patients with low tumor expression of C2orf40 as compared to luminal A patients with high tumor expression of C2orf40, who possessed median RFS of 83 months (Chart 3). C2orf40 primary tumor expression was not correlated with recurrence-free survival in basal subtype (Figure 2; log rank p -value: 0.59 for RFS, hazard ratio: 0.92 (0.68-1.25) (Fig. 2), luminal B subtype (Figure 2; log rank p -value: 0.32 for RFS, hazard ratio: 0.86 (0.64-1.16) (Fig. 2) or in HER2+ breast cancer (Figure 2; log rank p -value: 0.38 for RFS, hazard ratio: 1.22 (0.78-1.89) (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 C2orf40 was among the most significant transcriptional features in primary tumors from patients with breast cancer. C2orf40 expression in primary tumors of the breast was correlated with recurrence-free survival in patients with luminal A subtype cancers, with C2orf40 primary tumor mRNA levels a positive prognostic indicator for luminal A 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 chromosome 2 open reading frame 40, encoded by C2orf40, is a defining transcriptional feature of human breast cancer: in early onset-breast cancer, and in triple negative breast cancer. C2orf40 was expressed at significantly lower levels in primary tumors from patients with breast cancer, as compared to normal breast. In patients with luminal A subtype breast cancer, primary tumor expression of C2orf40 was correlated with recurrence-free survival.

Chromosome 2 open reading frame 40, 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: 13
Probe ID: 2498274
p-value: 9.66E-20
t: 1.45E+01
B: 34.639488
Gene: C2orf40
Gene name: chromosome 2 open reading frame 40

Chart 1: C2orf40 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: 852
Probe ID: 30839
p-value: 6.22E-08
t: 6.4669169
B: 8.088422
Gene: C2orf40
Gene name: chromosome 2 open reading frame 40

Chart 2: C2orf40 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: C2orf40 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of C2orf40 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 C2orf40 expression between normal breast tissue and primary tumors of the breast, a p -value, is $p<0.00001$.

Figure 2: C2orf40 expression correlates with recurrence-free survival in patients with luminal A breast cancer, but not in patients with basal, luminal B or HER2+ cancer.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival (OS) for $n=417$ patients with basal subtype breast cancer (upper left), $n=198$ patients with HER2+ breast cancer (upper right), $n=952$ patients with luminal A breast cancer (lower left), and $n=465$ patients with luminal B breast cancer (lower right), stratified into two groups, based on low or high expression of C2orf40 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 C2orf40 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

Low C2orf40 expression: 47 months

High C2orf40 expression: 83 months

Chart 3: Median recurrence-free survival is superior in luminal A subtype breast cancer patients with high primary tumor expression of C2orf40.

The recurrence-free survival of $n=952$ luminal A subtype breast cancer patients based on stratification into low or high tumor expression of C2orf40 is listed in this chart.