Differential expression of carbonic anhydrase 4 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 carbonic anhydrase 4, CA4, when comparing primary tumors of the breast to the tissue of origin, the normal breast. CA4 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 CA4 in primary tumors of the breast was correlated with distant metastasis-free survival in patients with basal-like subtype cancer, demonstrating a relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by PAM50 molecular subtype. CA4 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, CA4, carbonic anhydrase 4, 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. Carbonic anhydrase 4 emerged as among the most differentially expressed genes in cancer of the female breast.

Methods

We utilized datasets GSE42568² and GSE109169³ 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. 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 whose expression was profiled in this dataset are paired tissues (25 tumors matching 25 breast tissues from 25 patients). 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 CA4 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 CA4 mRNA expression levels with distant metastasis-free survival (DMFS) in $n=630$ patients with basal-like subtype cancer, $n=998$ patients with luminal A subtype cancer, $n=673$ patients with luminal B subtype cancer, $n=401$ patients with HER2+ cancer, and $n=63$ 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}.

CA4 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 the gene encoding carbonic anhydrase 4, CA4, 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, CA4 ranked 41 out of 54675 total transcripts, equating to 99.9% differential expression (Chart 1). Differential expression of CA4 in female breast cancer was statistically significant (Chart 1; $p=4.45E-35$).

Analysis of a second microarray dataset³, here studying global gene expression patterns in the tumors of 25 patients with early-onset breast cancer, again revealed significant differential expression of CA4 in human breast cancer (Chart 2). When sorting each of the genes expressed in the tumors of patients

with breast cancer based on significance of difference as compared to normal breast tissue, CA4 ranked 2068 out of 19076 total transcripts, equating to 89.2% differential expression (Chart 2). Differential expression of CA4 in the tumors of patients with breast cancer was statistically significant (Chart 2; $p=9.21E-07$). These data suggested that differential expression of CA4 was not an artifact of a single microarray dataset, nor was it strictly associated with early-onset breast cancer, rather a general feature of cancers of the breast.

CA4 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for CA4 from the breast and from breast tumors to understand the magnitude and direction of CA4 expression change. CA4 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of CA4 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). CA4 was expressed at 6.32 ± 1.92 arbitrary units (AU) in normal breast tissue, while it was expressed at 2.42 ± 0.48 AU in tumors of the breast. We calculated a mean fold change of 0.39 in CA4 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

CA4 expression correlates with survival outcomes in basal-like subtype human breast cancer.

We performed Kaplan-Meier survival analysis to study relationships between tumor CA4 mRNA expression levels and survival outcomes in patients with breast cancer. We observed a correlation between CA4 expression and distant metastasis-free survival (DMFS) in patients with basal-like subtype breast cancer which was statistically significant (Figure 2; log rank p -value: 0.031 for distant metastasis-free survival, hazard ratio: 1.38 (1.03-1.86) (Fig. 2)). CA4 mRNA levels were a negative prognostic indicator in basal-like subtype breast cancer patients. Median DMFS was 53.45 months for basal-like patients with low tumor expression of CA4 while median DMFS was 25.6 months for basal-like patients with high tumor expression of CA4 (Chart 3). CA4 primary tumor expression was not correlated with distant metastasis-free survival in luminal A breast cancer (Figure 2; log rank p -value: 0.22 for DMFS, hazard ratio: 1.22 (0.88-1.69) (Fig. 2)), luminal B subtype breast cancer (Figure 2; log rank p -value: 0.35 for DMFS, hazard ratio: 1.15 (0.86-1.54) (Fig. 2)), in HER2+ breast cancer (Figure 2; log rank p -value: 0.93 for DMFS, hazard ratio: 1.02 (0.71-1.44) (Fig. 2)) or in patients with normal-like breast cancer (Figure 2; log rank p -value: 0.54 for DMFS, hazard ratio: 1.35 (0.52-3.54) (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 CA4 was among the most significant transcriptional features in primary tumors from patients with breast cancer. CA4 expression in primary tumors of the breast was correlated with distant metastasis-free survival in patients with basal-like subtype disease, with mRNA levels of CA4 a negative prognostic indicator for basal-like 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 carbonic anhydrase 4, encoded by CA4, is a defining transcriptional feature of human breast cancer: in patients diagnosed after age 50, and in early onset-breast cancer. CA4 was expressed at significantly lower levels in primary tumors from patients with breast cancer as compared to normal breast tissue. Importantly, in patients with basal-like subtype breast

1 cancers, expression of CA4 was correlated with distant metastasis-free survival. Carbonic anhydrase 4
2 and the molecular processes to which it pertains to may be relevant to the initiation or progression of
3 human breast cancer.
4
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. 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.
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: 41
Probe ID: 206209_s_at
p-value: 4.45E-35
t: 1.76E+01
B: 69.2261734
Gene: CA4
Gene name: carbonic anhydrase 4

Chart 1: CA4 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: 2068
Probe ID: 3729419
p-value: 9.21E-07
t: 5.58
B: 5.274765
Gene: CA4
Gene name: carbonic anhydrase 4

Chart 2: CA4 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.

Figure 1: CA4 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of CA4 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 CA4 expression between normal breast tissue and primary tumors of the breast, a *p*-value, listed above.

Figure 2: CA4 expression correlates with distant metastasis-free survival in patients with basal-like 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 distant metastasis-free survival (DMFS) for $n=630$ patients with basal-like breast cancer (upper left), $n=401$ patients with HER2+ breast cancer (upper right), $n=998$ patients with luminal A breast cancer (middle left), $n=673$ patients with luminal B breast cancer (middle right), and $n=63$ patients with normal-like breast cancer (below) stratified into two groups, based on low or high expression of CA4 in patient primary tumors. The log rank p-value denoting statistical significance of difference in distant metastasis-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 CA4 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.

Basal-like patients

Low CA4 expression: 53.45 months

High CA4 expression: 25.6 months

Chart 3: In basal-like subtype breast cancer, median distant metastasis-free survival is superior in patients with low primary tumor expression of CA4.

The distant metastasis-free survival of $n=630$ basal-like subtype breast cancer patients based on stratification into low or high tumor expression of CA4 is listed in this chart.