Differential expression of aldehyde dehydrogenase 1 family member A2 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 aldehyde dehydrogenase 1 family member A2, ALDH1A2, when comparing primary tumors of the breast to the tissue of origin, the normal breast. ALDH1A2 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 ALDH1A2 in primary tumors of the breast was correlated with overall survival in patients with luminal A subtype and HER2+ cancer, but in a contrary manner, demonstrating a complex relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by molecular subtype. ALDH1A2 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, ALDH1A2, aldehyde dehydrogenase 1 family member A2, 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. Aldehyde dehydrogenase 1 family member A2 emerged as among the most differentially expressed genes in cancer of the female breast.

Methods

We utilized datasets GSE109169² and GSE42568³ 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). 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. 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 ALDH1A2 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 ALDH1A2 mRNA expression levels with overall survival (OS) in $n=404$ patients with basal subtype cancer, $n=794$ patients with luminal A subtype cancer, $n=515$ patients with luminal B subtype cancer, and $n=166$ 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}.

ALDH1A2 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 aldehyde dehydrogenase 1 family member A2, ALDH1A2, 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, ALDH1A2 ranked 196 out of 19076 total transcripts, equating to 99.0% differential expression (Chart 1). Differential expression of ALDH1A2 in female breast cancer was statistically significant (Chart 1; $p=3.64E-13$).

Analysis of a second microarray dataset³, here in 104 tumors from patients aged 31 to 89 at diagnosis, the majority of which were above age 50, again revealed significant differential expression of ALDH1A2 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, ALDH1A2 ranked 1579 out of 54675 total transcripts, equating to 97.1% differential expression (Chart 2). Differential expression of ALDH1A2 in the tumors of patients with breast cancer was statistically significant (Chart 2; $p=1.40E-13$). These data suggested that differential expression of ALDH1A2 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.

ALDH1A2 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for ALDH1A2 from the breast and from breast tumors to understand the magnitude and direction of ALDH1A2 expression change. ALDH1A2 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of ALDH1A2 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). ALDH1A2 was expressed at 6.62 ± 0.62 arbitrary units (AU) in normal breast tissue, while it was expressed at 5.35 ± 0.25 AU in tumors of the breast. We calculated a mean fold change of 0.81 in ALDH1A2 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

ALDH1A2 expression correlates with survival outcomes in luminal A subtype and HER2+ human breast cancer.

We performed Kaplan-Meier survival analysis in an attempt to correlate ALDH1A2 mRNA expression levels with survival outcomes in patients with breast cancer, evaluating correlations between molecular subtype and overall survival. We observed a correlation between ALDH1A2 expression and overall survival (OS) in patients with luminal A subtype breast cancer which approached statistical significance (Figure 2; log rank p -value: 0.053 for overall survival, hazard ratio: 0.73 (0.53-1) (Fig. 2)). ALDH1A2 mRNA levels were a positive prognostic indicator in luminal A subtype breast cancer patients. Median OS was 123.6 months for luminal A patients with low tumor expression of ALDH1A2 while median OS was 169.2 months for luminal A patients with high tumor expression of ALDH1A2 (Chart 3). We also observed a correlation between ALDH1A2 expression and overall survival (OS) in patients with HER2+ breast cancer which approached statistical significance (Figure 2; log rank p -value: 0.059 for overall survival, hazard ratio: 1.74 (0.97-3.13) (Fig. 2)). ALDH1A2 mRNA levels were a negative prognostic indicator in HER2+ subtype breast cancer patients. Median OS was 70.68 months for HER2+ patients with low tumor expression of ALDH1A2 while median OS was 40.68 months for HER2+ patients with high tumor expression of ALDH1A2 (Chart 3). ALDH1A2 primary tumor expression was not correlated with overall survival in basal subtype (Figure 2; log rank p -value: 0.66 for OS, hazard ratio: 0.92 (0.63-1.34) (Fig. 2)) or luminal B subtype (Figure 2; log rank p -value: 0.79 for OS, hazard ratio: 1.05 (0.74-1.48) (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 ALDH1A2 was among the most significant transcriptional features in primary tumors from patients with breast cancer. ALDH1A2 expression in primary tumors of the breast was correlated with overall survival in patients with luminal A and HER2+ subtype disease, with mRNA levels of ALDH1A2 a positive prognostic indicator for luminal A breast cancer patients but a negative prognostic indicator for HER2+ 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 aldehyde dehydrogenase 1 family
3 member A2, encoded by ALDH1A2, is a defining transcriptional feature of human breast cancer: in early
4 onset-breast cancer, and in patients diagnosed after age 50. ALDH1A2 was expressed at significantly
5 lower levels in primary tumors from patients with breast cancer as compared to normal breast tissue.
6 Importantly, in patients with luminal A subtype and HER2+ breast cancers, expression of ALDH1A2 was
7 correlated with overall survival. Aldehyde dehydrogenase 1 family member A2 and the molecular
8 processes to which it pertains to may be relevant 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. 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.
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. Györffy, 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: 196
Probe ID: 3626312
p-value: 3.64E-13
t: 9.7
B: 19.789855
Gene: ALDH1A2
Gene name: aldehyde dehydrogenase 1 family member A2

Chart 1: ALDH1A2 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: 1579
Probe ID: 207016_s_at
p-value: 1.40E-13
t: 8.35
B: 20.302126
Gene: ALDH1A2
Gene name: aldehyde dehydrogenase 1 family member A2

Chart 2: ALDH1A2 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.

Figure 1: ALDH1A2 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of ALDH1A2 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 ALDH1A2 expression between normal breast tissue and primary tumors of the breast, a p -value, is $p < 0.0001$.

Figure 2: ALDH1A2 expression correlates with overall survival in patients with luminal A subtype and HER2+ breast cancer, but not in patients with basal or luminal B subtype cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=404$ patients with basal subtype breast cancer (upper left), $n=166$ patients with HER2+ breast cancer (upper right), $n=794$ patients with luminal A breast cancer (lower left), and $n=515$ patients with luminal B breast cancer (lower right), stratified into two groups, based on low or high expression of ALDH1A2 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 ALDH1A2 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 ALDH1A2 expression: 123.6 months

High ALDH1A2 expression: 169.2 months

HER2+ patients

Low ALDH1A2 expression: 70.68 months

High ALDH1A2 expression: 40.68 months

Chart 3: In luminal A breast cancer, median overall survival is superior in patients with high primary tumor expression of ALDH1A2, but in HER2+ breast cancer, median overall survival is superior in patients with low primary tumor expression of ALDH1A2.

The overall survival of $n=794$ luminal A subtype breast cancer patients and $n=166$ HER2+ subtype breast cancer patients based on stratification into low or high tumor expression of ALDH1A2 is listed in this chart.