Differential expression of ID4 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 inhibitor of DNA binding 4, HLH protein, ID4, when comparing primary tumors of the breast to the tissue of origin, the normal breast. ID4 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 ID4 in primary tumors of the breast was correlated with overall survival in patients with basal and luminal A subtype cancer, demonstrating a relationship between primary tumor expression of a differentially expressed gene and patient survival outcomes influenced by molecular subtype. ID4 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, ID4, inhibitor of DNA binding 4, HLH protein, 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. Inhibitor of DNA binding 4, HLH protein emerged as among the most differentially expressed genes in cancer of the female breast.

Methods

We utilized datasets GSE109169² and GSE72653³ 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). GSE72653 was generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F with $n=8$ normal epithelium and $n=8$ tumors of the breast from patients with breast cancer; the tissues from this second dataset are paired tissues (8 tumors matching 8 normal breast epithelial tissues from 8 patients); analysis was performed using platform GPL6480. 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 ID4 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 ID4 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}.

ID4 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 inhibitor of DNA binding 4, HLH protein, encoded by the ID4 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, ID4 ranked 74 out of 19076 total transcripts, equating to 99.6% differential expression (Chart 1). Differential expression of ID4 in female breast cancer was statistically significant (Chart 1; $p=3.00E-15$).

To attempt to validate differential transcriptome-wide differential expression of ID4 in human breast cancer, we queried a second microarray dataset³, here comparing normal breast tissues and primary tumors of the breast. Again, we identified ID4 as a differentially expressed gene in the tumor tissues

of patients with breast cancer (Chart 2). When sorting each of the genes expressed in primary tumors of the breast based on significance of difference as compared to normal breast tissues, ID4 ranked 2896 out of 41093 total transcripts (Chart 2), equating to 93.0% differential expression. Differential expression of ID4 in the primary tumors of patients with breast cancer was statistically significant (Chart 2; $p=1.10E-02$). These data indicated that differential expression of ID4 was not an artifact of a single microarray dataset, and perhaps a general feature of cancers of the breast.

ID4 is expressed at significantly lower levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for ID4 from the breast and from breast tumors² to understand the magnitude and direction of ID4 expression change. ID4 was expressed at lower levels in tumors of the breast as compared to normal breast tissue (Figure 1). Decreased expression of ID4 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). ID4 was expressed at 8.22 ± 0.32 arbitrary units (AU) in normal breast tissue, while it was expressed at 7.22 ± 0.32 AU in tumors of the breast. We calculated a mean fold change of 0.88 in ID4 mRNA levels when comparing primary tumors of the breast to normal breast tissues.

ID4 expression correlates with survival outcomes in basal and luminal A subtype human breast cancer.

We performed Kaplan-Meier survival analysis in an attempt to correlate ID4 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 ID4 expression and overall survival (OS) in patients with basal subtype breast cancer (Figure 2; log rank p -value: 0.011 for overall survival, hazard ratio: 1.64 (1.11-2.42) (Fig. 2)). ID4 mRNA levels were a negative prognostic indicator in basal subtype breast cancer patients. Median OS was 118.03 months for basal patients with low tumor expression of ID4 while median OS was 47.67 months for basal patients with high tumor expression of ID4 (Chart 3). We also observed a statistically significant correlation between ID4 expression and overall survival (OS) in patients with luminal A subtype breast cancer (Figure 2; log rank p -value: 0.043 for overall survival, hazard ratio: 0.72 (0.52-0.99) (Fig. 2)). ID4 mRNA levels were a positive prognostic indicator in luminal A subtype breast cancer patients. Median OS was 120 months for luminal A patients with low tumor expression of ID4 while median OS was 175.46 months for luminal A patients with high tumor expression of ID4 (Chart 3). ID4 primary tumor expression was not correlated with overall survival in luminal B breast cancer (Figure 2; log rank p -value: 0.95 for OS, hazard ratio: 0.99 (0.7-1.4) (Fig. 2)) or in HER2+ breast cancer (Figure 2; log rank p -value: 0.69 for OS, hazard ratio: 1.12 (0.64-1.98) (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 ID4 was among the most significant transcriptional features in primary tumors from patients with breast cancer. ID4 expression in primary tumors of the breast was correlated with overall survival in patients with basal and luminal A subtype cancers, with ID4 primary tumor mRNA levels a positive prognostic indicator for luminal A patients but 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

1 datasets^{2,3}, providing evidence here that differential expression of inhibitor of DNA binding 4, HLH
2 protein, encoded by ID4, is a defining transcriptional feature of human breast cancer. ID4 was expressed
3 at significantly lower levels in primary tumors from patients with breast cancer, as compared to normal
4 breast. In patients with luminal A and basal subtype breast cancers, primary tumor expression of ID4 was
5 correlated with overall survival. Inhibitor of DNA binding 4, HLH protein, and the molecular processes to
6 which it pertains to, may be relevant to the initiation or progression of human breast cancer.
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. 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. Abdalla, M., Tran-Thanh, D., Moreno, J., Iakovlev, V., Nair, R., Kanwar, N., Abdalla, M., Lee, J.P., Kwan, J.Y.Y., Cawthorn, T.R. and Warren, K., 2017. Mapping genomic and transcriptomic alterations spatially in epithelial cells adjacent to human breast carcinoma. *Nature communications*, 8(1), pp.1-11.
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: 74
Probe ID: 2897453
p-value: 3.00E-15
t: 1.11E+01
B: 24.514513
Gene: ID4
Gene name: inhibitor of DNA binding 4, HLH protein

Chart 1: ID4 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: 2896
Probe ID: A_23_P59375
p-value: 1.10E-02
t: 2.8555529
B: -2.72945
Gene: ID4
Gene name: inhibitor of DNA binding 4, HLH protein

Chart 2: ID4 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: ID4 is expressed at significantly lower levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of ID4 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 ID4 expression between normal breast tissue and primary tumors of the breast, a p -value, is $p<0.0001$.

Figure 2: ID4 expression correlates with overall survival in patients with basal and luminal A breast cancer, but not in patients with luminal B subtype or HER2+ breast 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 ID4 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 ID4 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 patients

Low ID4 expression: 118.03 months

High ID4 expression: 47.67 months

Luminal A patients

Low ID4 expression: 120 months

High ID4 expression: 175.46 months

Chart 3: In luminal A subtype breast cancer, median overall survival is superior in patients with high primary tumor expression of ID4, but in basal subtype breast cancer, median overall survival is superior in patients with low primary tumor expression of ID4.

The overall survival of $n=404$ basal subtype breast cancer patients and $n=794$ luminal A subtype breast cancer patients based on stratification into low or high tumor expression of ID4 is listed in this chart.