ST3GAL4 expression associates with survival in triple negative breast cancer.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY USA

We mined published microarray data (1) to understand the most significant gene expression differences in the tumors of triple negative breast cancer patients based on survival following treatment: dead or alive. We observed significant transcriptome-wide differential expression of ST3 beta-galactoside alpha-2,3-sialyltransferase 4, encoded by ST3GAL4 when comparing the primary tumors of triple negative breast cancer patients dead or alive. Importantly, ST3GAL4 expression was correlated with distant metastasis-free survival in basal subtype breast cancer, a molecular subtype sharing significant overlap with triple negative breast cancer. ST3GAL4 may be of relevance as a biomarker or as a molecule of interest in understanding the etiology or progression of triple negative breast cancer.

Keywords: triple negative breast cancer, TNBC, systems biology of breast cancer, targeted therapeutics in breast cancer, ST3GAL4, ST3 beta-galactoside alpha-2,3-sialyltransferase 4.

Triple negative breast cancer is defined by lack of expression of the receptors for the hormones progesterone and estrogen (PR negative and ER negative), as well as for the human epidermal growth factor receptor 2 (HER2 negative) (2, 3). Though not completely identical, triple negative breast cancer shares significant overlap with the basal or basal-like molecular subtype of human breast cancer (3). TNBC is more frequently diagnosed in women of African descent (black women) for reasons not understood, and their disease course is similarly more aggressive than in white women with triple negative breast cancer (2, 3). To understand the transcriptional biology of triple negative breast cancer in an unbiased fashion and at the systems level, we mined published microarray data paired with patient survival data (1), performing comparative transcriptome analysis based on survival at time of data collection: dead or alive. We present here a finding from the results of this analysis.

Methods

We used dataset GSE142102 (1) for this differential gene expression analysis in conjunction with GEO2R. GSE142102 was generated using Affymetrix Human Gene 2.1 ST Array technology; in this analysis, we used data from $n=87$ tumors from TNBC patients that were dead at time of analysis and $n=123$ tumors from TNBC patients that were alive at time of analysis. 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 ST3GAL4 expression was significantly different between the primary tumors of TNBC patients that were dead or alive using a two-tailed t-test. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter tool (4) for correlation of ST3GAL4 mRNA expression levels with distant metastasis-free survival (DMFS) in $n=56$ patients with basal subtype breast cancer.

Results

We performed comparative transcriptome analysis using published microarray data (1) to understand in an unbiased fashion the most significant gene expression differences in the primary tumors of patients with triple negative breast cancer based on survival outcomes.

ST3GAL4 is differentially expressed in triple negative breast cancer and negatively correlates with survival.

When comparing primary tumors of 87 live patients with triple negative breast cancer with that of 123 dead TNBC patients we identified ST3 beta-galactoside alpha-2,3-sialyltransferase 4, *ST3GAL4*, as among the most significant transcriptional differences based on survival in human triple negative breast cancer. When ranking each of the transcripts whose expression was measured by microarray based on significance of difference in expression between the tumors of patients dead and alive, ST3GAL4 ranked 18 out of 53617 total transcripts, equivalent to 99.97% differential expression (Chart 1). ST3GAL4 differential expression in the tumors of patients based on survival outcomes was statistically significant (Figure 1; $p=0.00032573$).

ST3GAL4 is expressed at significantly lower levels in the breast tumors of TNBC patients when comparing based on survival.

We obtained exact mRNA expression levels for ST3GAL4 from breast tumors of TNBC patients

1 dead and alive to understand the magnitude and direction of ST3GAL4 expression change based on
2 survival outcomes. ST3GAL4 was expressed at lower levels in tumors of TNBC patients alive as
3 compared to that of TNBC patients dead (Figure 1). Decreased expression of ST3GAL4 in primary breast
4 tumors was statistically significant (Figure 1: $p=0.0006$). ST3GAL4 was expressed at 0.30 ± 1.11
5 arbitrary units (AU) in the primary tumors of TNBC patients dead, while it was expressed at -0.20 ± 0.91
6 AU in the primary tumors of TNBC patients alive.

7 ST3GAL4 expression correlates with distant metastasis-free survival in basal subtype human breast 8 cancer.

9 We referenced a bioinformatics tool containing Kaplan-Meier survival data of human breast cancer
10 patients to understand more thoroughly whether a relationship existed between ST3GAL4 primary tumor
11 expression and patient survival in human breast cancer. We found that ST3GAL4 expression was
12 correlated with distant metastasis-free survival in patients with p53 mutant basal or basal-like human
13 breast cancer, a breast cancer molecular subtype which shares significant overlap with triple negative
14 breast cancer, in the lower quartile; this correlation trended towards statistical significance (Figure 2; log
15 rank p -value: 0.08 for distant metastasis-free survival, hazard ratio: 5.1 (0.67-38.86) (Fig. 2)). ST3GAL4
16 mRNA levels were a negative prognostic indicator in p53 mutant basal subtype breast cancer patients.
17 Median DMFS was, on average, markedly greater for p53 mutant basal subtype patients with low tumor
18 expression of ST3GAL4 as compared to p53 mutant basal subtype patients with high tumor expression of
19 ST3GAL4.

20 Thus, we found through unbiased comparative transcriptome analysis that ST3GAL4 was among
21 the genes whose expression was most significantly different in the primary tumors of triple negative breast
22 cancer patients alive when compared to that of TNBC patients dead, that ST3GAL4 was expressed at
23 significantly lower levels in the primary tumors of TNBC patients alive as compared to TNBC patients
24 dead, and that ST3GAL4 expression was negatively correlated with distant metastasis-free survival in the
25 basal or basal-like molecular subtype, in patients whose tumors contained a mutation in p53.

26 **Discussion**

27 It is not understood why women of African descent (black women) are diagnosed with triple
28 negative breast cancer more frequently than white women, and why this disease is more aggressive in
black women. ST3GAL4 may be of relevance as a biomarker in identifying patients who are more likely
to survive, perhaps as part of a panel of genes used for prognostic stratification, and the molecule itself
may be important to the biology of triple negative breast cancer etiology or more likely its progression.
Blind systems-level analyses, like this, can help shed light on the fundamental molecular characteristics of
a disease whose racial propensities demand further explanation.

References

1. Purrington, K.S., Knight III, J., Dyson, G., Ali-Fehmi, R., Schwartz, A.G., Boerner, J.L. and Bandyopadhyay, S., 2020. CLCA2 expression is associated with survival among African American women with triple negative breast cancer. *PloS one*, 15(4), p.e0231712.
2. Foulkes, W.D., Smith, I.E. and Reis-Filho, J.S., 2010. Triple-negative breast cancer. *New England journal of medicine*, 363(20), pp.1938-1948.
3. Dietze, E.C., Sistrunk, C., Miranda-Carboni, G., O'regan, R. and Seewaldt, V.L., 2015. Triple-negative breast cancer in African-American women: disparities versus biology. *Nature Reviews Cancer*, 15(4), pp.248-254.
4. 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: 18
Probe ID: 16733288
p-value: 0.00032573
t: 3.5939282
B: -2.61
Gene: ST3GAL4
Gene name: ST3 beta-galactoside alpha-2,3-sialyltransferase 4

Chart 1: ST3GAL4 is differentially expressed at the transcriptome-level when comparing the primary tumors of TNBC patients dead and alive.

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: ST3GAL4 is expressed at significantly lower levels in the primary tumors of TNBC patients alive as compared to the primary tumors of TNBC patients dead.

The mRNA expression level of ST3GAL4 in the primary tumors of TNBC patients dead (left) and in primary tumors of TNBC patients alive (right) is graphically depicted with the result of a statistical test evaluating significance of difference in ST3GAL4 expression between the primary tumors of TNBC patients dead and alive, a p -value, listed here: $p=0.0006$.

Figure 2: ST3GAL4 expression correlates with distant metastasis-free survival in p53 mutant basal subtype human breast cancer.

Depicted in this Kaplan-Meier plot is the probability of distant metastasis-free survival (DMFS) for $n=56$ breast cancer patients of basal subtype, based on low or high expression of ST3GAL4 in patient primary tumors, in patients whose tumors contained a mutation in p53. 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 ST3GAL4 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.