TMPRSS7 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 transmembrane protease, serine 7, encoded by TMPRSS7 when comparing the primary tumors of triple negative breast cancer patients dead or alive. Importantly, TMPRSS7 expression was correlated with recurrence-free survival in patients with breast cancer. TMPRSS7 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, TMPRSS7, transmembrane protease, serine 7.

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 TMPRSS7 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 TMPRSS7 mRNA expression levels with recurrence-free survival (RFS) in $n=1090$ patients with 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.

TMPRSS7 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 transmembrane protease, serine 7, *TMPRSS7*, 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, *TMPRSS7* ranked 164 out of 53617 total transcripts, equivalent to 99.7% differential expression (Chart 1). *TMPRSS7* differential expression in the tumors of patients based on survival outcomes was statistically significant (Figure 1; $p=0.00402739$).

TMPRSS7 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 *TMPRSS7* from breast tumors of TNBC patients

1 dead and alive to understand the magnitude and direction of TMPRSS7 expression change based on
2 survival outcomes. TMPRSS7 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 TMPRSS7 in primary breast
4 tumors was statistically significant (Figure 1: $p=0.0050$). TMPRSS7 was expressed at 0.22 ± 1.07
5 arbitrary units (AU) in the primary tumors of TNBC patients alive, while it was expressed at -0.18 ± 0.92
6 AU in the primary tumors of TNBC patients dead.

7 TMPRSS7 expression correlates with recurrence-free survival in human breast cancer.

8 We referenced a bioinformatics tool containing Kaplan-Meier survival data of human breast cancer
9 patients to understand more thoroughly whether a relationship existed between TMPRSS7 primary tumor
10 expression and patient survival in human breast cancer. We found that TMPRSS7 expression was
11 correlated with recurrence-free survival in patients with breast cancer (Figure 2; log rank p -value: 0.0033
12 for recurrence-free survival, hazard ratio: 0.47 (0.28-0.79) (Fig. 2)). TMPRSS7 mRNA levels were a
13 positive prognostic indicator in breast cancer patients. Median RFS was, on average, markedly shorter for
14 patients with low tumor expression of TMPRSS7 as compared to patients with high tumor expression of
15 TMPRSS7.

16 Thus, we found through unbiased comparative transcriptome analysis that TMPRSS7 was among
17 the genes whose expression was most significantly different in the primary tumors of triple negative breast
18 cancer patients alive when compared to that of TNBC patients dead, that TMPRSS7 was expressed at
19 significantly lower levels in the primary tumors of TNBC patients alive as compared to TNBC patients
20 dead, and that TMPRSS7 expression was positively correlated with recurrence-free survival in patients
21 with breast cancer.

22 **Discussion**

23 It is not understood why women of African descent (black women) are diagnosed with triple
24 negative breast cancer more frequently than white women, and why this disease is more aggressive in
25 black women. TMPRSS7 is a type II transmembrane serine protease; a single-nucleotide polymorphism in
26 the TMPRSS7 gene has been associated with prognosis in human breast cancer (5). TMPRSS7 may be of
27 relevance as a biomarker in identifying patients who are more likely to survive, perhaps as part of a panel
28 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. Nagy, Á., Munkácsy, G. and Györffy, B., 2021. Pancancer survival analysis of cancer hallmark genes. *Scientific reports*, 11(1), pp.1-10.
5. Luostari, K., Hartikainen, J.M., Tengström, M., Palvimo, J.J., Kataja, V., Mannermaa, A. and Kosma, V.M., 2014. Type II transmembrane serine protease gene variants associate with breast cancer. *PLoS One*, 9(7), p.e102519.

Rank: 164
Probe ID: 16943892
p-value: 0.00402739
t: 2.8760088
B: -3.4
Gene: TMPRSS7
Gene name: transmembrane protease, serine 7

Chart 1: TMPRSS7 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: TMPRSS7 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 TMPRSS7 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 TMPRSS7 expression between the primary tumors of TNBC patients dead and alive, a p -value, listed here: $p=0.0050$.

Figure 2: TMPRSS7 expression correlates with recurrence-free survival in human breast cancer.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival (RFS) for $n=1090$ breast cancer patients, based on low or high expression of TMPRSS7 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 TMPRSS7 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.