DNAJC28 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 DnaJ (Hsp40) homolog, subfamily C, member 28, encoded by DNAJC28 when comparing the primary tumors of triple negative breast cancer patients dead or alive. Importantly, DNAJC28 expression was correlated with overall survival in patients with breast cancer. DNAJC28 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, DNAJC28, DnaJ (Hsp40) homolog, subfamily C, member 28.

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 DNAJC28 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 DNAJC28 mRNA expression levels with overall survival (OS) 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.

DNAJC28 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 DnaJ (Hsp40) homolog, subfamily C, member 28, *DNAJC28*, 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, DNAJC28 ranked 57 out of 53617 total transcripts, equivalent to 99.9% differential expression (Chart 1). DNAJC28 differential expression in the tumors of patients based on survival outcomes was statistically significant (Figure 1; $p=0.00144841$).

DNAJC28 is expressed at significantly higher levels in the breast tumors of TNBC patients when comparing based on survival.

We obtained exact mRNA expression levels for DNAJC28 from breast tumors of TNBC patients

1 dead and alive to understand the magnitude and direction of DNAJC28 expression change based on
2 survival outcomes. DNAJC28 was expressed at higher levels in tumors of TNBC patients alive as
3 compared to that of TNBC patients dead (Figure 1). Increased expression of DNAJC28 in primary breast
4 tumors was statistically significant (Figure 1: $p=0.0013$). DNAJC28 was expressed at -0.27 ± 0.93
5 arbitrary units (AU) in the primary tumors of TNBC patients dead, while it was expressed at 0.17 ± 1.03
6 AU in the primary tumors of TNBC patients alive.

7 DNAJC28 expression correlates with overall 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 DNAJC28 primary tumor
10 expression and patient survival in human breast cancer. We found that DNAJC28 expression was
11 correlated with overall survival in patients with breast cancer, in the lower tertile (Figure 2; log rank
12 p -value: 0.049 for overall survival, hazard ratio: 0.72 (0.52-1) (Fig. 2)). DNAJC28 mRNA levels were a
13 positive prognostic indicator in breast cancer patients. Median OS was 116.4 months for patients with low
14 tumor expression of DNAJC28 as compared to 131.5 months for patients with high tumor expression of
15 DNAJC28.

16 Thus, we found through unbiased comparative transcriptome analysis that DNAJC28 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 DNAJC28 was expressed at
19 significantly higher levels in the primary tumors of TNBC patients alive as compared to TNBC patients
20 dead, and that DNAJC28 expression was positively correlated with overall survival in patients with breast
21 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. DNAJC28 may be of relevance as a biomarker in identifying patients who are more likely
26 to survive, perhaps as part of a panel of genes used for prognostic stratification, and the molecule itself
27 may be important to the biology of triple negative breast cancer etiology or more likely its progression.
28 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.

Rank: 57
Probe ID: 16925187
p-value: 0.00144841
t: -3.184827
B: -3.08
Gene: DNAJC28
Gene name: DnaJ (Hsp40) homolog, subfamily C, member 28

Chart 1: DNAJC28 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: DNAJC28 is expressed at significantly higher levels in the primary tumors of TNBC patients alive as compared to the primary tumors of TNBC patients dead.

The mRNA expression level of DNAJC28 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 DNAJC28 expression between the primary tumors of TNBC patients dead and alive, a p -value, listed here: $p=0.0013$.

Figure 2: DNAJC28 expression correlates with overall survival in human breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=1090$ breast cancer patients, based on low or high expression of DNAJC28 in patient primary tumors, in the lower tertile. 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 DNAJC28 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.

1 Low DNAJC28 expression: 116.4 months

2 High DNAJC28 expression: 131.5 months

3
4 **Chart 2: Median overall survival is superior in breast cancer patients with high primary tumor expression of DNAJC28.**

5 The median OS (overall survival) of breast cancer patients, with low primary tumor expression of
6 DNAJC28 and high primary tumor expression of DNAJC28 is listed in this chart, in the lower tertile.