CLEC2D 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 at time of analysis: dead or alive. We observed significant transcriptome-wide differential expression of C-type lectin domain family 2, member D, encoded by CLEC2D when comparing the primary tumors of triple negative breast cancer patients dead or alive. Importantly, CLEC2D expression was significantly correlated with overall survival in basal subtype breast cancer, a molecular subtype sharing significant overlap with triple negative breast cancer. CLEC2D 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, CLEC2D, C-type lectin domain family 2, member D.

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 CLEC2D 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 CLEC2D mRNA expression levels with overall survival (OS) in $n=431$ 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.

CLEC2D 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 C-type lectin domain family 2, member D, *CLEC2D*, 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, CLEC2D ranked 21 out of 53617 total transcripts, equivalent to 99.96% differential expression (Chart 1). CLEC2D differential expression in the tumors of patients based on survival outcomes was statistically significant (Figure 1; $p=0.00037025$).

CLEC2D 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 CLEC2D from breast tumors of TNBC patients

1 dead and alive to understand the magnitude and direction of CLEC2D expression change based on
2 survival outcomes. CLEC2D 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 CLEC2D in primary breast
4 tumors was statistically significant (Figure 1: $p=0.0001$). CLEC2D was expressed at -0.26 ± 0.75 arbitrary
5 units (AU) in the primary tumors of TNBC patients alive, while it was expressed at 0.23 ± 1.11 AU in the
6 primary tumors of TNBC patients dead.

7 CLEC2D expression correlates with overall survival in basal subtype 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 CLEC2D primary tumor
10 expression and patient survival in human breast cancer. We found that CLEC2D expression was
11 correlated with overall survival in patients with basal or basal-like human breast cancer, a breast cancer
12 molecular subtype which shares significant overlap with triple negative breast cancer (Figure 2; log rank
13 p -value: 0.048 for overall survival, hazard ratio: 0.68 (0.46-1) (Fig. 2)). CLEC2D mRNA levels were a
14 positive prognostic indicator in basal subtype breast cancer patients. Median OS was 54.6 months for
15 basal subtype patients with low tumor expression of CLEC2D as compared to 90 months for basal subtype
16 patients with high tumor expression of CLEC2D.

17 Thus, we found through unbiased comparative transcriptome analysis that CLEC2D was among the
18 genes whose expression was most significantly different in the primary tumors of triple negative breast
19 cancer patients alive when compared to that of TNBC patients dead, that CLEC2D was expressed at
20 significantly higher levels in the primary tumors of TNBC patients alive as compared to TNBC patients
21 dead, and that CLEC2D expression was positively correlated with overall survival in the basal or
22 basal-like molecular subtype.

23 **Discussion**

24 It is not understood why women of African descent (black women) are diagnosed with triple
25 negative breast cancer more frequently than white women, and why this disease is more aggressive in
26 black women. CLEC2D may be of relevance as a biomarker in identifying patients who are more likely to
27 survive, perhaps as part of a panel of genes used for prognostic stratification, and the molecule itself may
28 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: 21
Probe ID: 16748205
p-value: 0.00037025
t: -3.5604281
B: -2.65
Gene: CLEC2D
Gene name: C-type lectin domain family 2, member D

Chart 1: CLEC2D 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: CLEC2D 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 CLEC2D 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 CLEC2D expression between the primary tumors of TNBC patients dead and alive, a p -value, listed here: $p=0.0001$.

Figure 2: CLEC2D expression correlates with overall survival in basal subtype human breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=431$ breast cancer patients of basal subtype, based on low or high expression of CLEC2D 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 CLEC2D 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 CLEC2D expression: 54.6 months

2 High CLEC2D expression: 90 months

3
4 **Chart 2: Median overall survival is superior in basal subtype breast cancer patients with high primary tumor expression of CLEC2D.**

5 The median OS (overall survival) of basal subtype breast cancer patients, with low primary tumor
6 expression of CLEC2D and high primary tumor expression of CLEC2D is listed in this chart.