Analysis of gene expression patterns in triple negative breast cancer: MERTK and WHSC1L1.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY 11730

Triple negative breast cancer (TNBC) shares overlap with the basal molecular subtype of breast cancer and is more frequently diagnosed in African-American (black) women for reasons not understood (1,2). To understand genes whose expression may be of pertinence to the development or progression of triple negative breast cancer, we mined published microarray data (3,4) comparing global gene expression profiles of TNBC histology groups, identifying genes whose expression changed the least between among TNBCs, suggesting that these genes may be important for TNBC biology. We identified the MER proto-oncogene tyrosine kinase MERTK and the Wolf-Hirschhorn syndrome candidate 1-like 1 WHSC1L1 among the genes whose expression differed the least when comparing TNBC cases and subtypes. In another dataset, MERTK and WHSCL1 were found to be differentially expressed in TNBC when comparing primary tumors of the breast to normal breast tissue. Kaplan-Meier survival analysis revealed that expression levels of MERTK and WHSCL1 correlated with survival outcomes in human breast cancer, and that this correlation differed based on race of the patient. MERTK and WHSCL1 may be of relevance in understanding the etiology or progression of triple negative breast cancer.

Keywords: triple negative breast cancer, breast cancer, MERTK, NSD3, WHSC1L1, systems biology of breast cancer, targeted therapeutics in breast cancer.

It is not understood why African-American women are more frequently diagnosed with triple negative breast cancer, a disease marked by lack of expression of receptors for the hormones estrogen and progesterone, and for the human epidermal growth factor receptor 2, HER2 (1,2). TNBC shares overlap with the basal-like subtype of breast cancer which is characterized by relatively worse patient outcomes as compared to other subtypes (1). We utilized a transcriptome-wide gene expression analysis approach, mining published microarray data (3,4) for the discovery of genes with potential relevance to the biology of TNBC, aiming to utilize these findings as a starting point for further exploration and validation. We report here initial findings of this analysis.

Methods

We utilized microarray datasets GSE76124 (3) and GSE38959 (4) for this differential gene expression analysis of triple negative breast cancer, in conjunction with GEO2R. GSE76124 was generated using Affymetrix Human Genome U133 Plus 2.0 Array technology with TNBC tumors of three histology groups: $n=7$ adenocarcinoma/carcinoma NOS (not otherwise specified) cases; $n=189$ infiltrating ductal carcinoma cases, and $n=2$ “other” breast cancers, which were medullary carcinomas; analysis was performed using platform GPL570. GSE38959 was generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F technology with $n=13$ samples of normal mammary gland ductal cells and $n=30$ samples of tumor cells from patients with triple negative breast cancer; analysis was performed using platform GPL4133. 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. For Kaplan-Meier survival analysis, we utilized the Kaplan-Meier survival analysis online tool (5), comparing breast cancer cases based on race and evaluating correlations with survival outcomes based on expression of MERTK or WHSC1L1 in $n=1090$ breast cancer patients.

Results

We mined published microarray data (3,4) generated using cells and tissues from patients with triple negative breast cancer (TNBC) for the discovery of genes with importance to the biology of TNBC.

MERTK and WHSC1L1 are among the genes whose expression changes the least transcriptome-wide when comparing TNBC cases based on histology group.

We compared the global gene expression profiles of 198 cases of triple negative breast cancer, divided into three groups based on histology group: adenocarcinoma/carcinoma not otherwise-specified (NOS) (7 cases), infiltrating ductal carcinoma (189 cases), and “other” breast cancers (2 cases), using published microarray data (3). Differential gene expression analyses operate under the principle that genes whose expression differs most between groups compared may be of importance to biological description as they inform the investigator as to the identity of the sample. In this case, we were interested in the opposite: what genes were among those whose expression changed the least? This would direct us towards genes whose importance was shared among TNBC cases – genes that inform the investigator as to the identity of TNBC tumors, regardless of histology group. The expression of 54,675 transcripts in total was compared between TNBC histology groups. To further refine our selection approach, we sorted through the 0.5% of transcripts whose expression was least different transcriptome-wide, then asking if any of these genes had previously been reported as possessing oncogenic or proto-oncogenic abilities, or if any of these genes had previously been reported as

describing the identity of a human breast cancer. Using this approach, we identified two genes of interest: MER proto-oncogene tyrosine kinase MERTK and the Wolf-Hirschhorn syndrome candidate 1-like 1 WHSC1L1.

MERTK differential expression ranked 0.36% transcriptome-wide, with differential expression at 54476 out of 54675 total transcripts when triple negative breast cancers were compared based on TNBC histology group. Differential expression of MERTK was not statistically significant (Chart 1; $p=9.99 \times 10^{-1}$), thus indicating that expression of MERTK is not significantly different between TNBC cases. MERTK is a proto-oncogene.

WHSC1L1 differential expression ranked 0.11% transcriptome-wide, with differential expression at 54613 out of 54675 total transcripts when triple negative breast cancers were compared based on TNBC histology group. Differential expression of WHSC1L1 was not statistically significant (Chart 2; $p=9.99 \times 10^{-1}$), thus indicating that expression of WHSC1L1 was not significantly different between TNBC cases. We have previously identified, using published microarray data, the related gene WHSC1 (NSD2/KMT3G) as among the genes whose expression was most significantly different transcriptome-wide when comparing primary tumors from patients with early-onset breast cancer to normal breast tissue (7).

MERTK and WHSC1L1 are differentially expressed in the tumor cells of patients with TNBC.

Analysis of a second microarray dataset (4), here in tumor cells of patients with triple negative breast cancer as compared to normal mammary gland ductal cells, revealed significant differential expression of MERTK in human triple negative breast cancer breast cancer when comparing TNBC tumor cells to normal mammary gland ductal cells (Chart 3). When sorting each of the genes expressed in the tumor cells of patients with triple negative breast cancer based on significance of difference as compared to normal mammary gland ductal cells, MERTK ranked 12763 out of 45015 total transcripts, equating to 71.6% differential expression (Chart 3). Differential expression of MERTK in the tumor cells of patients with triple negative breast cancer was statistically significant (Chart 3; $p=2.49 \times 10^{-2}$).

WHSC1L1 was also significantly differentially expressed in human triple negative breast cancer breast cancer when comparing TNBC tumor cells to normal mammary gland ductal cells (Chart 4). When sorting each of the genes expressed in the tumor cells of patients with triple negative breast cancer based on significance of difference as compared to normal mammary gland ductal cells, WHSC1L1 ranked 11394 out of 45015 total transcripts, equating to 74.7% differential expression (Chart 4). Differential expression of WHSC1L1 in the tumor cells of patients with triple negative breast cancer was statistically significant (Chart 4; $p=1.62 \times 10^{-2}$).

Thus, MERTK and WHSC1L1 expression is shared among triple negative breast cancer cases, and their expression, to a modest degree, defines the TNBC tumor cell as compared to normal mammary gland ductal cells.

MERTK and WHSC1L1 primary tumor expression is correlated with survival outcomes in human breast cancer, and this differs based on race.

We performed Kaplan-Meier survival analysis of $n=1090$ breast cancer patients in total to determine whether primary tumor expression of MERTK or WHSC1L1 were correlated with survival outcomes in human breast cancer, and whether any such correlation differed based on race.

We observed significant correlation of MERTK primary tumor expression with recurrence-free survival in white breast cancer patients but not in black breast cancer patients. (Figure 1; log rank p -value: 0.044 for recurrence-free survival in white breast cancer patients, hazard ratio: 1.72 (1.01-2.93) (Fig. 1)). MERTK mRNA levels were a negative prognostic indicator in white breast cancer patients. MERTK primary tumor expression was not correlated with recurrence-free survival in black breast cancer patients ((Figure 1; log rank p -value: 0.88 for recurrence-free survival in black breast cancer patients, hazard ratio: 1.07 (0.44-2.57) (Fig. 1)).

We also observed correlation of WHSC1L1 primary tumor expression with overall survival in white breast cancer patients, which approached statistical significance, but not in black breast cancer patients (Figure 2; log rank p -value: 0.069 for overall survival in white breast cancer patients, hazard ratio: 1.42 (0.97-2.07) (Fig. 2)). WHSC1L1 mRNA levels were a negative prognostic indicator in white breast cancer patients. Median OS was 219.77 months for white patients with low tumor expression of WHSC1L1 as compared to white patients with high tumor expression of WHSC1L1, who possessed median OS of 124.53 months (Chart 5). WHSC1L1 primary tumor expression was not correlated with overall survival in black breast cancer patients ((Figure 2; log rank p -value: 0.62 for overall survival in black breast cancer patients, hazard ratio: 0.83 (0.4-1.72) (Fig. 2)).

However, further analysis of Kaplan-Meier data based on survival quartile revealed that in the lower quartile, WHSC1L1 primary tumor expression was significantly correlated with overall survival in black breast cancer patients but not in white breast cancer patients. ((Figure 3; log rank p -value: 0.028 for overall survival in black breast cancer patients, hazard ratio: 0.44 (0.21-0.93) (Fig. 3)). Here, WHSC1L1 mRNA levels were a positive prognostic indicator in black breast cancer patients. Median OS was 78.7 months for black patients with low tumor expression of WHSC1L1 as compared to black patients with high tumor expression of WHSC1L1, who possessed median OS of 129.1 months (Chart 6). WHSC1L1 primary tumor expression was not correlated with overall survival in white breast cancer patients, in the lower survival quartile ((Figure 3; log rank p -value: 0.26 for overall survival in white breast cancer patients, hazard ratio: 1.31 (0.82-2.09) (Fig. 3)).

Thus, expression of WHSC1L1 and MERTK is, transcriptome-wide, among the most similar in triple negative breast cancers when comparing cases based on histology group, MERTK and WHSC1L1 are modestly but significantly differentially expressed at the transcriptome-level in triple negative breast cancer when comparing TNBC tumor cells to normal mammary gland ductal cells, and primary tumor expression of MERTK and WHSC1L1 are correlated with survival outcomes in human breast cancer, and this correlation differs based on race of the patient. MERTK and WHSC1L1 are candidate genes of relevance for understanding the biology of triple negative breast cancer.

Discussion

We found here through unbiased, systems-level analysis of published tumor transcriptome data that among the genes whose expression is most similar between cases of triple negative breast cancer when compared by histology group are the proto-oncogene tyrosine kinase Mer, MERTK, and the Wolf-Hirschhorn syndrome candidate 1-like 1, WHSC1L1, also known as the histone lysine N-methyltransferase NSD3. We do not yet understand why, when evaluating survival in the bulk population, MERTK and WHSC1L1 primary tumor expression levels are correlated with survival outcomes in white patients but not in black patients, despite the fact that triple negative breast cancer is more commonly diagnosed in black (African-American) patients. One likely important factor contributing to these findings is that the Kaplan-Meier analysis data considers all breast cancer patients, not strictly patients with triple negative breast cancer. MERTK and WHSC1L1 are candidate genes of interest for understanding the biology of triple negative breast cancer, and understanding sequence variation

(single-nucleotide polymorphism) in MERTK and WHSC1L1 in the black and white population may potentially shed light on why TNBC is more commonly diagnosed in black females.

References

1. 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.
2. 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.
3. Burstein, M.D., Tsimelzon, A., Poage, G.M., Covington, K.R., Contreras, A., Fuqua, S.A., Savage, M.I., Osborne, C.K., Hilsenbeck, S.G., Chang, J.C. and Mills, G.B., 2015. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clinical Cancer Research*, 21(7), pp.1688-1698.
4. Komatsu, M., Yoshimaru, T., Matsuo, T., Kiyotani, K., Miyoshi, Y., Tanahashi, T., Rokutan, K., Yamaguchi, R., Saito, A., Imoto, S. and Miyano, S., 2013. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. *International journal of oncology*, 42(2), pp.478-506.
5. Nagy, Á., Munkácsy, G. and Györffy, B., 2021. Pancancer survival analysis of cancer hallmark genes. *Scientific reports*, 11(1), pp.1-10.
6. Keating, A.K., Salzberg, D.B., Sather, S., Liang, X., Nickoloff, S., Anwar, A., Deryckere, D., Hill, K., Joung, D., Sawczyn, K.K. and Park, J., 2006. Lymphoblastic leukemia/lymphoma in mice overexpressing the Mer (MerTK) receptor tyrosine kinase. *Oncogene*, 25(45), pp.6092-6100.
7. Mamoor, S., 2020. Differential expression of the Wolf-Hirschhorn syndrome candidate 1, WHSC1 (NSD2/KMT3G), in cancers of the breast.

Rank: 54476
Probe ID: 206028_s_at
p-value: 9.99 e-01
E: 0.0048321
Gene: MERTK
Gene name: MER proto-oncogene, tyrosine kinase

Chart 1: At the transcriptome-level, among the genes whose expression is most shared between triple negative breast cancers is MERTK.

Rank of differential expression, probe ID, p-value of global differential expression, the F-statistic, gene and gene name are listed in this chart. In this case, a significant p-value indicates expression is different between groups compared (TNBC cases based on histology group).

Rank: 54613
Probe ID: 225318_at
p-value: 9.99 e-01
F: 0.001302
Gene: WHSC1L1
Gene name: Wolf-Hirschhorn syndrome candidate 1-like 1

Chart 2: At the transcriptome-level, among the genes whose expression is most shared between triple negative breast cancers is WHSC1L1.

Rank of differential expression, probe ID, p-value of global differential expression, the F statistic, gene and gene name are listed in this chart. In this case, a significant p-value indicates expression is different between groups compared (TNBC cases based on histology group).

Rank: 12763
Probe ID: 28863
p-value: 2.49E-02
t: 2.3198289
B: -4.198133
Gene: MERTK
Gene name: MER proto-oncogene, tyrosine kinase

Chart 3: MERTK is differentially expressed in the tumor cells of patients with triple negative 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: 11394
Probe ID: 43166
p-value: 1.62E-02
t: 2.4969012
B: -3.817625
Gene: WHSC1L1
Gene name: Wolf-Hirschhorn syndrome candidate 1-like 1

Chart 4: WHSC1L1 is differentially expressed in the tumor cells of patients with triple negative 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: MERTK expression correlates with recurrence-free survival in white breast cancer patients but not in black breast cancer patients.

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

Figure 2: WHSC1L1 expression correlates with overall survival in white breast cancer patients but not in black breast cancer patients.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=1090$ breast cancer patients of white race (left) or of black race (right), based on low or high expression of WHSC1L1 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 WHSC1L1 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 WHSC1L1 expression: 219.77 months

2 High WHSC1L1 expression: 124.53 months

3 **Chart 5: In white patients, median overall survival is superior in breast cancer patients with low**
4 **primary tumor expression of WHSC1L1.**

5 The overall survival of white breast cancer patients based on stratification into low or high tumor
6 expression of WHSC1L1 is listed in this chart.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Figure 3: WHSC1L1 expression correlates with overall survival in black breast cancer patients but not in white breast cancer patients, in the lower survival quartile.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=1090$ breast cancer patients of white race (left) or of black race (right), based on low or high expression of WHSC1L1 in patient primary tumors, in the lower survival quartile. 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 WHSC1L1 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 WHSC1L1 expression: 78.7 months
2 High WHSC1L1 expression: 129.1 months

3 **Chart 6: In black patients, median overall survival is superior in breast cancer patients with high**
4 **primary tumor expression of WHSC1L1 - in the lower survival quartile.**

5 The overall survival of black breast cancer patients based on stratification into low or high tumor
6 expression of WHSC1L1 is listed in this chart, in the lower survival quartile.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28