Analysis of gene expression patterns in triple negative breast cancer III: the retinoic acid receptor alpha and other TNBC candidate genes.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY 11730

Triple negative breast cancer (TNBC) shares overlap with the basal or basal-like molecular subtype of breast cancer and is more frequently diagnosed in women of African descent (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) comparing global gene expression profiles of TNBC cases, identifying genes whose expression was least different among TNBC cases, indicating conservation of expression patterns suggestive of importance for TNBC biology. We identified the gene encoding the retinoic acid receptor alpha (RARA), a fatty acid elongase (ELOVL1), as well as multiple genes encoding molecules involved in epigenetic functions or with nucleic acid binding or modification properties, including TDRD7, KDM1B, PHF7, TAF5L, as well as the microRNA hsa-miR-605. Kaplan-Meier survival analysis revealed that expression levels of each of these genes correlated with survival outcomes in the basal subtype of human breast cancer, which shares significant overlap with triple negative breast cancer at the level of gene expression (2). RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and hsa-miR-605 may be of relevance in understanding the etiology or progression of triple negative breast cancer. Together with our previous findings, the data allude to a potential pathogenic mechanism involving transcriptional perturbation of epigenetic machinery in triple negative breast cancer (4, 5).

Keywords: triple negative breast cancer, breast cancer, RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L, hsa-miR-605, 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) 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 further findings of this analysis.

Methods

We utilized microarray datasets GSE142102 (3) for this differential gene expression analysis of human triple negative breast cancer, 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. Two comparisons that were made included: (1) the tumors of $n=58$ patients that were alive (patient IDs ending in 626-727) as compared to the tumors of $n=65$ patients that were alive (patient IDs ending in 729-844), as well as (2) the tumors of $n=41$ patients that were labeled as “dead” as compared to the tumors of $n=46$ patients that were labeled as “dead” (patient IDs available upon request). 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 Kaplan-Meier survival analysis tools (6, 7), comparing breast cancer cases based on molecular subtype and evaluating correlations with survival outcomes based on expression of hsa-miR-605 in $n=1090$ breast cancer patients, or based on expression of ELOVL1 or RARA in patients of the basal molecular subtype for recurrence-free survival in $n=846$ patients, or based on expression of TDRD7, KDM1B, PHF7, or TAF5L for overall survival in $n=404$ or $n=278$ patients.

Results

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

RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and hsa-miR-605 are among the genes whose expression changes the least transcriptome-wide when comparing TNBC cases based on TNBC molecular subtype

We compared the global gene expression profiles of 210 cases of triple negative breast cancer, which had been divided into two groups based on survival outcomes: dead or alive, using published microarray data (3). To enhance reproducibility, we compared groups of patients expired (arbitrarily) amongst each other, as well as patients alive (arbitrarily) amongst each other. 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 the individual patient's transcriptional profile, a composite representation of the patient's genetic background. The expression of 53617 transcripts in total was compared between TNBC patient groups. To further refine

our selection approach, we sorted through the 1% of transcripts whose expression was least different transcriptome-wide, then asking if any of these genes had previously been reported as important for breast cancer tumorigenesis or invasiveness, or in such a process in another solid tumor type, or have been found to be genetically or transcriptionally perturbed in some breast cancer subtype or cell line. Using this comparative tumor transcriptome approach, we identified seven genes of interest as potential candidate genes relevant to the etiology or progression of human triple negative breast cancer: RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and hsa-miR-605.

RARA (the retinoic acid receptor alpha) differential expression ranked 0.03% transcriptome-wide, with differential expression at 53598 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group (to avoid identification of genes that were similar or different based on survival outcomes). Differential expression of RARA was not statistically significant (Chart 1; $p=9.98E-01$), thus indicating that among TNBC patients, the expression level of RARA was among those most significantly similar transcriptome-wide. One study demonstrated that RARA was amplified in 13% of HER2 negative breast cancers (8).

ELOVL1 (ELOVL fatty acid elongase 1) differential expression ranked 0.49% transcriptome-wide, with differential expression at 53354 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group. Differential expression of ELOVL1 was not statistically significant (Chart 1; $p=9.90E-01$), thus indicating that among TNBC patients, the expression level of ELOVL1 was among those most significantly similar transcriptome-wide. A study profiling hydrophobic lipid metabolites in triple negative breast cancer cases observed significant differences in fatty acid metabolism, with 8.5-fold up-regulation of ELOVL1 mRNA in triple negative breast cancer tissues as compared to normal breast tissues, in conjunction with more intense immunohistochemical ELOVL1 staining in triple negative breast cancer tissues as compared to ER positive tumors or normal breast tissues (9).

miR-605 differential expression ranked 0% transcriptome-wide, with differential expression at 53617 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group. Differential expression of miR-605 was not statistically significant (Chart 1; $p=1$), thus demonstrating in conjunction with percent differential expression data that among this cohort of TNBC patients, the expression level of miR-605 was the single most significantly conserved transcriptome-wide. miR-605 is part of a regulatory pathway involving the tumor suppressor p53 and mdm2 wherein p53 can activate miR-605 expression while mdm2 can repress it (10; in Li-Fraumeni syndrome, a genetic disorder featuring mutation of p53 and cancer predisposition, variants in miR-605 influence age of onset of disease (11), and one study suggested that miR-605 could function in breast cancer progression by acting as a endogenous competitor of CircDDX17 (12).

TDRD7 (tudor domain containing 7) differential expression ranked 0.64% transcriptome-wide, with differential expression at 53275 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group. Differential expression of TDRD7 was not statistically significant (Chart 1; $p=9.87E-01$), thus indicating that among TNBC patients, the expression level of TDRD7 was among those most significantly similar transcriptome-wide. In a genetic screen which examined the effect of siRNA depletion of 332 chromatin regulators on MCF7 breast cancer cell proliferation, depletion of TDRD7 ranked in the top 10% with respect to effect on cell proliferation (13).

KDM1B (lysine (K)-specific demethylase 1B, also known as LSD2) differential expression ranked 0.49% transcriptome-wide, with differential expression at 53353 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group. Differential expression of KDM1B was not statistically significant (Chart 1; $p=9.90E-01$), thus indicating that among TNBC patients, the expression level of KDM1B was among those most significantly similar transcriptome-wide.

KDM1B expression is significantly elevated in basal-like subtype breast cancer as compared to normal breast tissues or other breast cancer subtypes, and in xenograft studies using the MDA-MB-231 breast cancer cell line in immunocompromised mice, over-expression of KDM1B (also known as LSD2) accelerated tumor growth but reduced the number of lung metastases (14).

PHF7 (PHD finger protein 7) differential expression ranked 0.40% transcriptome-wide, with differential expression at 53400 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group. Differential expression of PHF7 was not statistically significant (Chart 1; $p=9.85E-01$), thus indicating that among TNBC patients, the expression level of RARA was among those most significantly similar transcriptome-wide. High throughput luciferase receptor screening in HeLa cells demonstrated that PHF7 is a repressor of the estrogen receptor alpha (ER α) (15).

TAF5L (TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa) differential expression ranked 0.17% transcriptome-wide, with differential expression at 53527 out of 53617 total transcripts when triple negative breast cancers were compared arbitrarily but within survival group. Differential expression of TAF5L was not statistically significant (Chart 1; $p=9.93E-01$), thus indicating that among TNBC patients, the expression level of RARA was among those most significantly similar transcriptome-wide. TAF5L expression is significantly down-regulated in primary invasive breast cancers as compared to matched benign tissues (16).

Thus, among genes whose expression is most quantitatively similar, transcriptome-wide, when comparing cases of triple negative breast cancers arbitrarily was RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and miR-605.

RARA, ELOVL1, TDRD7, KDM1B, PHF7, and TAF5L primary tumor expression is correlated with patient survival outcomes in the basal subtype of human breast cancer.

We analyzed patient survival data (Kaplan Meier survival analysis) to determine whether the expression of these genes was correlated with patient survival. We found that primary tumor expression of six of seven of these genes correlated with patient survival outcomes, either recurrence-free survival or overall survival, in patients diagnosed with the basal molecular subtype of human breast cancer which shares significant overlap with triple negative breast cancer (2).

RARA primary tumor expression was correlated with recurrence-free survival (RFS) in patients with basal subtype human breast cancer, in the lower survival tertile, and this correlation trended towards statistical significance ((Figure 1; log rank p -value: 0.074 for RFS in basal type breast cancer patients, hazard ratio: 1.25 (0.98-1.61) (Fig. 1)). RARA mRNA levels were a negative prognostic indicator in basal subtype breast cancer patients. Median RFS was 25 months for basal subtype patients with low tumor expression of RARA as compared to basal subtype patients with high tumor expression of RARA, who possessed median RFS of 21 months (Chart 8).

ELOVL1 primary tumor expression was correlated with recurrence-free survival (RFS) in patients with basal subtype human breast cancer, in the upper survival tertile, and this correlation trended towards statistical significance ((Figure 2; log rank p -value: 0.093 for RFS in basal type breast cancer patients, hazard ratio: 1.22 (0.97-1.54) (Fig. 2)). ELOVL1 mRNA levels were a negative prognostic indicator in basal subtype breast cancer patients. Median RFS was 24.07 months for basal subtype patients with low tumor expression of ELOVL1 as compared to basal subtype patients with high tumor expression of ELOVL1, who possessed median RFS of 18.6 months (Chart 9).

TDRD7 primary tumor expression was significantly correlated with overall survival (OS) in patients with basal subtype human breast cancer ((Figure 3; log rank p -value: 5.5e-05 for OS in basal type breast cancer patients, hazard ratio: 0.45 (0.3-0.67) (Fig. 3)). TDRD7 mRNA levels were a positive prognostic indicator in basal subtype breast cancer patients. Median OS was, on average, markedly shorter for basal subtype patients with low tumor expression of TDRD7 as compared to basal subtype patients with high tumor expression of TDRD7 (Figure 3).

KDM1B primary tumor expression was correlated with overall survival (OS) in patients with basal subtype human breast cancer, and this correlation trended towards statistical significance ((Figure 4; log rank p -value: 0.081 for OS in basal type breast cancer patients, hazard ratio: 0.66 (0.41-1.06) (Fig. 4)). KDM1B mRNA levels were a positive prognostic indicator in basal subtype breast cancer patients. Median OS was 39.5 months for basal subtype patients with low tumor expression of KDM1B as compared to basal subtype patients with high tumor expression of KDM1B, who possessed median OS of 64.8 months (Chart 10).

PHF7 primary tumor expression was significantly correlated with overall survival (OS) in patients with basal subtype human breast cancer, in the upper survival quartile ((Figure 5; log rank p -value: 0.05 for OS in basal type breast cancer patients, hazard ratio: 1.49 (1-2.22) (Fig. 5)). PHF7 mRNA levels were a negative prognostic indicator in basal subtype breast cancer patients. Median OS was 63.96 months for basal subtype patients with low tumor expression of PHF7 as compared to basal subtype patients with high tumor expression of PHF7 who possessed median OS of 34.9 months (Chart 11).

TAF5L primary tumor expression was correlated with overall survival (OS) in patients with basal subtype human breast cancer, and this correlation approached statistical significance ((Figure 6; log rank p -value: 0.067 for OS in basal type breast cancer patients, hazard ratio: 0.7 (0.48-1.03) (Fig. 6)). TAF5L mRNA levels were a positive prognostic indicator in basal subtype breast cancer patients. Median OS was 54.6 months for basal subtype patients with low tumor expression of TAF5L as compared to basal subtype patients with high tumor expression of TAF5L, who possessed median OS of 85.2 months (Chart 12).

miR-605 primary tumor expression is correlated with patient survival outcomes in human breast cancer.

Further Kaplan-Meier survival analysis demonstrated that hsa-miR-605 primary tumor expression was significantly correlated with overall survival (OS) in human breast cancer ((Figure 7; log rank p -value: 0.042 for OS in breast cancer patients, hazard ratio: 2.84 (0.99-8.1) (Fig. 7)). miR-605 mRNA levels were a negative prognostic indicator in breast cancer patients. Median OS was 95.13 months for breast cancer patients with low tumor expression of miR-605 as compared to breast cancer patients with high tumor expression of miR-605, who possessed median OS of 51.87 months (Chart 13).

Thus, expression of RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and miR-605 are among the most similar in triple negative breast cancers when comparing cases arbitrarily within survival group (comparing surviving patients with surviving patients and not those with disparate survival outcomes), and primary tumor expression of RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and miR-605 is correlated with survival outcomes in human breast cancer; for nearly all genes, this correlation can be observed in patients diagnosed with the basal-subtype of human breast cancer, which shares significant similarity with triple negative breast cancer, but with respect to diagnostic overlap and at the level of gene expression. RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and miR-605 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 arbitrarily include RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and hsa-miR-605. The identification of these genes as potential contributors at the genetic level to the etiology or progression of triple negative breast cancer is supported by findings in the scientific literature that suggest that each of these factors may play a role in breast cancer tumorigenesis or invasiveness, or in such a process in another solid tumor type, or have been found to be genetically or transcriptionally perturbed in some breast cancer subtype or cell line. RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and hsa-miR-605 are candidate genes of interest for understanding the biology of triple negative breast cancer, and understanding sequence variation (single-nucleotide polymorphism and copy number variation) in RARA, ELOVL1, TDRD7, KDM1B, PHF7, TAF5L and hsa-miR-605 in the black and white population may potentially shed light on why TNBC is more prevalent in black females. We previously demonstrated that among the genes whose expression levels are most similar transcriptome-wide in triple negative breast cancer using alternative comparative tumor transcriptome approaches (based on molecular and histological subtype) are NSD3, MLLT6, and PHC3 (4, 5). Together with KDM1B and PHF7, the data allude to pervasive transcriptional abnormalities of epigenetic machinery in patients with triple negative breast cancer.

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. 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.e023171
4. Mamoor, S., Analysis of gene expression patterns in triple negative breast cancer: MERTK and WHSC1L1.
5. Mamoor, S., 2021. Analysis of gene expression patterns in triple negative breast cancer II: TNBC molecular subtype
6. Nagy, Á., Munkácsy, G. and Györffy, B., 2021. Pancancer survival analysis of cancer hallmark genes. *Scientific reports*, 11(1), pp.1-10.
7. Györffy, 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.
8. Alsafadi, S., Even, C., Falet, C., Goubar, A., Commo, F., Scott, V., Quidville, V., Albiges, L., Dieci, M.V., Guegan, J. and Lazar, V., 2013. Retinoic acid receptor alpha amplifications and retinoic acid sensitivity in breast cancers. *Clinical breast cancer*, 13(5), pp.401-408.
9. Yamashita, Y., Nishiumi, S., Kono, S., Takao, S., Azuma, T. and Yoshida, M., 2017. Differences in elongation of very long chain fatty acids and fatty acid metabolism between triple-negative and hormone receptor-positive breast cancer. *BMC cancer*, 17(1), pp.1-21.
10. Xiao, J., Lin, H., Luo, X., Luo, X. and Wang, Z., 2011. miR-605 joins p53 network to form a p53: miR-605: Mdm2 positive feedback loop in response to stress. *The EMBO journal*, 30(3), pp.524-532.
11. Said, B.I. and Malkin, D., 2015. A functional variant in miR-605 modifies the age of onset in Li-Fraumeni syndrome. *Cancer genetics*, 208(1-2), pp.47-51.
12. Peng, H.H. and Wen, Y.G., 2020. CircDDX17 acts as a competing endogenous RNA for miR-605 in breast cancer progression. *European Review for Medical and Pharmacological Sciences*, 24(12), pp.6794-6801.
13. Serandour, A.A., Mohammed, H., Miremadi, A., Mulder, K.W. and Carroll, J.S., 2018. TRPS1 regulates oestrogen receptor binding and histone acetylation at enhancers. *Oncogene*, 37(39), pp.5281-5291.
14. Chen, L., Vasilatos, S.N., Qin, Y., Katz, T.A., Cao, C., Wu, H., Tasdemir, N., Levine, K.M.,

- 1 Oesterreich, S., Davidson, N.E. and Huang, Y., 2017. Functional characterization of lysine-specific
2 demethylase 2 (LSD2/KDM1B) in breast cancer progression. *Oncotarget*, 8(47), p.81737.
- 3 15. Han, X., Guo, J., Deng, W., Zhang, C., Du, P., Shi, T. and Ma, D., 2008. High-throughput cell-based
4 screening reveals a role for ZNF131 as a repressor of ERalpha signaling. *BMC genomics*, 9(1), pp.1-10.
- 5 16. Shubbar, E., Helou, K., Kovács, A., Nemes, S., Hajizadeh, S., Enerbäck, C. and Einbeigi, Z., 2013.
6 High levels of γ -glutamyl hydrolase (GGH) are associated with poor prognosis and unfavorable clinical
7 outcomes in invasive breast cancer. *BMC cancer*, 13(1), pp.1-12.
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28

Rank: 53598
Probe ID: 16834071
p-value: 9.98E-01
t: -0.002485
B: -6.295013
Gene: RARA
Gene name: retinoic acid receptor, alpha

Chart 1: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is RARA.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Rank: 53354
Probe ID: 16686122
p-value: 9.90E-01
t: -0.0119755
B: -5.79289
Gene: ELOVL1
Gene name: ELOVL fatty acid elongase 1

Chart 2: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is ELOVL1.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Rank: 53617
Probe ID: 16705006
p-value: 1
t: 0.0000118
B: -5.79295
Gene: MIR605
Gene name: microRNA 605

Chart 3: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is miR-605.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Rank: 53275
Probe ID: 17087308
p-value: 9.87E-01
t: 0.0156932
B: -5.79284
Gene: TDRD7
Gene name: tudor domain containing 7

Chart 4: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is TDRD7.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Rank: 53353
Probe ID: 17005169
p-value: 9.90E-01
t: -0.0119764
B: -5.79289
Gene: KDM1B
Gene name: lysine (K)-specific demethylase 1B

Chart 5: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is KDM1B.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Rank: 53400
Probe ID: 16941493
p-value: 9.85E-01
t: -0.019342
B: -6.2948349
Gene: PHF7
Gene name: PHD finger protein 7

Chart 6: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is PHF7.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Rank: 53527

Probe ID: 16700343

p-value: 9.93E-01

t: 0.009183

B: -6.2949752

Gene: TAF5L

Gene name: TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa

Chart 7: At the transcriptome-level, among the genes whose expression is most similar between triple negative breast cancer cases is TAF5L.

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 two groups compared, 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 arbitrarily, but with the same survival outcome).

Figure 1: RARA expression correlates with recurrence-free survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival (RFS) for $n=846$ basal subtype breast cancer patients, based on low or high expression of RARA in patient primary tumors, in the lower survival 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 RARA 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.

Low RARA expression: 25 months
High RARA expression: 21 months

Chart 8: In patients with basal subtype human breast cancer, median recurrence-free survival is superior in breast cancer patients with low primary tumor expression of RARA.

The recurrence-free survival of $n=846$ basal subtype breast cancer patients based on stratification into low or high tumor expression of RARA is listed in this chart, in this lower survival tertile.

Figure 2: ELOVL1 expression correlates with recurrence-free survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival (RFS) for $n=846$ basal subtype breast cancer patients, based on low or high expression of ELOVL1 in patient primary tumors, in the upper survival 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 ELOVL1 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.

Low ELOVL1 expression: 24.07 months

High ELOVL1 expression: 18.6 months

Chart 9: In patients with basal subtype human breast cancer, median recurrence-free survival is superior in breast cancer patients with low primary tumor expression of ELOVL1.

The recurrence-free survival of $n=846$ basal subtype breast cancer patients based on stratification into low or high tumor expression of ELOVL1 is listed in this chart, in the upper survival tertile.

Figure 3: KDM1B expression correlates with overall survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=278$ basal subtype breast cancer patients, based on low or high expression of KDM1B 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 KDM1B 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
2
3 Low KDM1B expression: 39.5 months

4 High KDM1B expression: 64.8 months

5 **Chart 10: In patients with basal subtype human breast cancer, median overall survival is superior in**
6 **breast cancer patients with high primary tumor expression of KDM1B.**

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
The overall survival of $n=278$ basal subtype breast cancer patients based on stratification into low or high tumor expression of KDM1B is listed in this chart.

Figure 4: PHF7 expression correlates with overall survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=404$ basal subtype breast cancer patients, based on low or high expression of PHF7 in patient primary tumors, in the upper 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 PHF7 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.

Low PHF7 expression: 63.96 months

High PHF7 expression: 34.9 months

Chart 11: In patients with basal subtype human breast cancer, median overall survival is superior in breast cancer patients with low primary tumor expression of PHF7.

The overall survival of $n=404$ basal subtype breast cancer patients based on stratification into low or high tumor expression of PHF7 is listed in this chart, in the upper survival quartile.

Figure 5: TDRD7 expression correlates with overall survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=404$ basal subtype breast cancer patients, based on low or high expression of TDRD7 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 TDRD7 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 6: TAF5L expression correlates with overall survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=404$ basal subtype breast cancer patients, based on low or high expression of TAF5L 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 TAF5L 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
2
3 Low TAF5L expression: 54.6 months

4 High TAF5L expression: 85.2 months

5 **Chart 12: In patients with basal subtype human breast cancer, median overall survival is superior in**
6 **breast cancer patients with high primary tumor expression of TAF5L.**

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
The overall survival of $n=404$ basal subtype breast cancer patients based on stratification into low or high tumor expression of TAF5L is listed in this chart.

Figure 7: miR-605 expression correlates with overall survival in basal subtype breast cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=404$ basal subtype breast cancer patients, based on low or high expression of Homo sapiens (hsa) miR-605 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 hsa-miR-605 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.

Low miR-605 expression: 95.13 months

High miR-605 expression: 51.87 months

Chart 13: Median overall survival is superior in breast cancer patients with low primary tumor expression of miR-605.

The overall survival of breast cancer patients based on stratification into low or high tumor expression of miR-605 is listed in this chart.