Low GATA3 expression distinguishes basal-like human breast cancer.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY 11730

Patients diagnosed with basal-like breast cancer face a more aggressive disease course and more dismal prognosis than patients diagnosed with luminal A and luminal B breast cancer molecular subtypes (1-4). We mined published microarray data (5, 6) to understand in an unbiased fashion the most distinguishing transcriptional features of tumors from patients with basal or basal-like subtype breast cancer. We observed transcriptome-wide differential expression of the transcription factor GATA3 when comparing tumors of patients with basal-like breast cancer with that of other PAM50 molecular subtypes. GATA3 mRNA was present at significantly reduced quantities in the tumors of patients with basal-like breast cancer. Analysis of patient survival data revealed that GATA3 primary tumor expression was correlated with distant metastasis-free survival, with low GATA3 expression correlated with inferior survival outcomes. Low GATA3 expression appears to distinguish basal-like human breast cancer from the other molecular subtypes.

Keywords: basal subtype, basal-like breast cancer, BLBC, systems biology of breast cancer, targeted therapeutics in breast cancer, breast cancer.

Classification of disease types and disease subtypes enables development of enhanced diagnostic capabilities, prognostic information, as well as a nuanced understanding of human disease that can support superior medical, surgical and radiological treatment of complex diseases such as cancer. Molecular subtypes in human breast cancer include luminal A, luminal B, basal or basal-like, and HER2+. Systems-level analyses, such as comparative tumor transcriptome analyses, can inform such approaches. We mined published microarray data (5, 6) to identify the most significant transcriptional differences when comparing human breast cancer molecular subtypes. Integrative microarray dataset analysis has provided data illuminating genes whose expression defines basal-like subtype breast cancers, and we present evidence here.

Methods

We utilized datasets GSE87049 (5) and GSE158309 (6) for this global differential gene expression analysis of basal subtype human breast cancer in conjunction with GEO2R. GSE87049 was generated using Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] technology with $n=14$ basal-like breast tumors and $n=109$ breast tumors of other molecular subtypes (luminal A, luminal B, HER2-enriched, and normal-like); analysis was performed using platform GPL6244. GSE158308 was generated using Affymetrix Human Genome U133A Array technology with $n=52$ tumors from patients with basal-like breast breast and $n=410$ tumors from patients with luminal A, luminal B and HER2+ molecular subtypes; analysis was performed using platform GPL96. The data presented in the graph compare $n=52$ tumors from patients with basal-like breast cancer to $n=182$ tumors from patients with luminal B breast cancer. 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 GATA3 gene expression was significantly different between the tumors of patients with basal and non-basal molecular subtype breast cancer using a two-tailed t-test with Welch's correction. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter tool (8) for correlation of GATA3 mRNA expression levels with distant metastasis-free survival in $n=630$ basal subtype breast cancer patients.

Results

GATA3 is among the genes whose expression is most different when comparing basal-like breast cancers with other molecular subtypes.

We used published microarray data (5) to compare the whole tumor transcriptomes of 14 patients with basal-like breast cancer and 109 patients with breast cancer of other molecular subtype - luminal A, luminal B, HER2-enriched, and normal-like - to identify genes whose expression marked the basal-like molecular subtype. We observed transcriptome-wide differential expression of GATA3 transcript when performing this basal-like breast cancer analysis (Table 2). When sorting each of the transcripts whose expression was measured globally, GATA3 expression ranked 40 out of 33297 total transcripts, equating to 99.9% differential expression; this was statistically significant (p -value: $2.27E-18$).

Analysis of a second published microarray dataset, comparing tumor transcriptome data from a separate cohort of patients (tumors from 52 patients with basal subtype breast cancer and 410 patients with breast cancer of other molecular subtypes: luminal A, luminal B and HER2) (6) revealed transcriptome-wide differential expression of a GATA3 transcript in patients with basal subtype human breast cancer (Table 2). When sorting each of the transcripts whose expression was measured globally, GATA3 expression ranked 42 out of 22283 total transcripts, equating to 99.8% differential expression; this was statistically significant (p -value: $3.16E-07$).

GATA3 expression is significantly lower in basal-like breast cancers.

We obtained exact GATA3 mRNA expression data for basal-like breast tumors and tumors from all other molecular subtypes (luminal A, luminal B, HER2+, and normal-like) to determine magnitude and direction of GATA3 gene expression change in basal-like breast cancer. GATA3 was expressed at lower levels in the tumors of patients with basal-like breast cancer as compared to the tumors of patients from all other PAM50 molecular subtypes, and this difference was statistically significant (Figure 1; $p < 0.0001$; Figure 2; $p = 0.0002$). GATA3 was expressed at 8.57 ± 1.13 arbitrary units (A.U.) in basal-like breast cancer and 10.92 ± 0.77 A.U. in all other molecular subtypes (Figure 1).

GATA3 expression correlates with distant metastasis-free survival in basal subtype breast cancer patients.

We asked whether primary tumor expression of GATA3 was correlated with survival outcomes in patients with basal-like breast cancer. We observed a statistically significant correlation between primary tumor expression of GATA3 and distant metastasis-free survival (DMFS) in patients with basal-like breast cancer (Figure 3; p -value: 0.038; hazard ratio: 0.73 (0.54-0.98)). GATA3 primary tumor expression was a positive prognostic indicator with respect to distant metastasis-free survival in patients with basal-like breast cancer. Patients with low tumor expression of GATA3 possessed median DMFS of 28 months, while patients with high tumor expression of GATA3 possessed median DMFS of 57.6 months (Table 3).

Discussion

We find here that GATA3 is among the genes whose expression is most different when comparing basal-like breast cancers (BLBC) to cancers of other molecular subtypes, with significantly less GATA3 mRNA in BLBC. Kaplan-Meier analysis of patient survival data revealed that GATA3 expression was correlated with distant metastasis-free survival in patients with basal-like breast cancer, with lower mRNA expression of GATA3 associated with inferior outcomes. One study has shown correlation between low GATA3 expression and increased tumor size, as well as negative expression of the receptors for the hormones estrogen and progesterone (8). Changes in GATA3 expression may be important in understanding the processes that govern transformation and tumor maintenance in basal or basal-like human breast cancer (9-11). GATA3 may be one factor that specifies tumor subtype identity in basal-like breast cancer.

References

1. Rakha, E.A., Reis-Filho, J.S. and Ellis, I.O., 2008. Basal-like breast cancer: a critical review. *Journal of clinical oncology*, 26(15), pp.2568-2581.
2. Milioli, H.H., Tishchenko, I., Riveros, C., Berretta, R. and Moscato, P., 2017. Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. *BMC medical genomics*, 10(1), pp.1-17.
3. 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.
4. 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.
5. Romero-Cordoba, S.L., Salido-Guadarrama, I., Rebollar-Vega, R., Bautista-Piña, V., Dominguez-Reyes, C., Tenorio-Torres, A., Villegas-Carlos, F., Fernández López, J.C., Uribe-Figueroa, L., Alfaro-Ruiz, L. and Hidalgo-Miranda, A., 2021. Comprehensive omic characterization of breast cancer in Mexican-Hispanic women. *Nature communications*, 12(1), pp.1-19.
6. Heimes, A.S., Härtner, F., Almstedt, K., Krajnak, S., Lebrecht, A., Battista, M.J., Edlund, K., Brenner, W., Hasenburg, A., Sahin, U. and Gehrmann, M., 2020. Prognostic significance of interferon- γ and its signaling pathway in early breast cancer depends on the molecular subtypes. *International Journal of Molecular Sciences*, 21(19), p.7178.
7. 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.
8. Yoon, N.K., Maresh, E.L., Shen, D., Elshimali, Y., Apple, S., Horvath, S., Mah, V., Bose, S., Chia, D., Chang, H.R. and Goodglick, L., 2010. Higher levels of GATA3 predict better survival in women with breast cancer. *Human pathology*, 41(12), pp.1794-1801.
9. Dydensborg, A.B., Rose, A.A.N., Wilson, B.J., Grote, D., Paquet, M., Giguere, V., Siegel, P.M. and Bouchard, M., 2009. GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. *Oncogene*, 28(29), pp.2634-2642.

1 10. Takaku, M., Grimm, S.A., Roberts, J.D., Chrysovergis, K., Bennett, B.D., Myers, P., Perera, L.,
2 Tucker, C.J., Perou, C.M. and Wade, P.A., 2018. GATA3 zinc finger 2 mutations reprogram the breast
3 cancer transcriptional network. *Nature communications*, 9(1), pp.1-14.

4 11. Yan, W., Cao, Q.J., Arenas, R.B., Bentley, B. and Shao, R., 2010. GATA3 inhibits breast cancer
5 metastasis through the reversal of epithelial-mesenchymal transition. *Journal of Biological Chemistry*,
6 285(18), pp.14042-14051.

Rank: 40
probe ID: 7926105
p-value: 2.27E-18
t: -10.299948
B: 31.19348
Gene: GATA3
Gene name: GATA-binding protein 3

Table 1: GATA3 is among the genes whose expression is most different, across the transcriptome, when comparing basal-like breast cancers with luminal A, luminal B, HER2-enriched, and normal-like breast cancers.

The rank of global differential expression, probe ID, the *p*-value with respect to differential expression transcriptome-wide, *t*, a moderated *t*-statistic, and *B*, the log-odds of differential expression between the groups compared are listed in this chart.

Rank: 42
probe ID: 209603_at
p-value: 3.16E-07
t: -5.190107
B: 6.3659527
Gene: GATA3
Gene name: GATA-binding protein 3

Table 2: In a separate cohort of patients with basal subtype breast cancer, GATA3 is again among the genes most differentially expressed transcriptome-wide when compared to the tumors of patients with breast cancer of other molecular subtypes: luminal A, luminal B, and HER2+.

The rank of global differential expression, probe/transcript ID, the *p*-value with respect to differential expression transcriptome-wide, *t*, a moderated *t*-statistic, *B*, the log-odds of differential expression between the groups compared, the gene and gene name are listed in this chart.

Figure 1: GATA3 mRNA is present at significantly reduced quantities in basal-like breast cancers.

Depicted is mRNA expression of GATA3 in tumors from patients with basal-like breast cancer (left) and from patients with cancer of another molecular subtype (right): luminal A, luminal B, HER2+, and normal-like. The horizontal bar corresponds to the mean value, and the result of a statistical test evaluating the significance of difference in expression between the two groups compared, a *p*-value, is listed above.

Figure 2: GATA3 mRNA is present at significantly reduced quantities in basal-like breast cancers.

Depicted is mRNA expression of GATA3 in tumors from patients with basal-like breast cancer (left) and from patients with cancer of another molecular subtype (luminal B, right). The horizontal bar corresponds to the mean value, and the result of a statistical test evaluating the significance of difference in expression between the two groups compared, a *p*-value, is listed above.

Figure 3: Low GATA3 expression is correlated with inferior distant metastasis-free survival in basal-like breast cancer.

Depicted in this Kaplan-Meier plot is the probability of distant metastasis-free survival for $n=630$ basal subtype patients stratified into two groups, based on low or high expression of GATA3 in patient primary tumors. The log rank p -value denoting statistical significance of difference in distant metastasis-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 GATA3 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 GATA3 expression: 28 months
High GATA3 expression: 57.6 months

Table 3: Median DMFS is superior in basal subtype breast cancer patients with high primary tumor expression of GATA3.

The median DMFS (distant metastasis-free survival) of $n=630$ basal subtype breast cancer patients, with low primary tumor expression of GATA3 and high primary tumor expression of GATA3 is listed in this chart.