Elevated CDC20 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 cell division cycle 20, CDC20, when comparing tumors of patients with basal-like breast cancer with that of other PAM50 molecular subtypes. CDC20 mRNA was present at significantly higher quantities in the tumors of patients with basal-like breast cancer. Analysis of patient survival data revealed that CDC20 primary tumor expression was correlated with distant metastasis-free survival in human breast cancer. Elevated CDC20 expression appears to distinguish basal-like human breast cancer from the other molecular subtypes.

Keywords: basal subtype, basal-like breast cancer, CDC20, systems biology of breast cancer, targeted therapeutics in breast cancer, molecular subtype

Classification of disease types and disease subtypes enables development of enhanced diagnostic capabilities, acquisition of prognostic information, as well as an 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 define basal-like subtype breast cancers, and we present evidence here.

Methods

We utilized datasets GSE74667 (5) and GSE87049 (6) for this global differential gene expression analysis of basal subtype human breast cancer in conjunction with GEO2R. GSE74667 was generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F technology with $n=23$ basal-like breast tumors and $n=72$ breast tumors of other molecular subtypes (luminal A, luminal B, HER2-enriched, and normal-like); analysis was performed using platform GPL6480. 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. 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 CDC20 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 (PRISM 9.1.1 (225)). For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter tool (8) for correlation of CDC20 mRNA expression levels with distant metastasis-free survival in $n=2765$ breast cancer patients, $n=630$ basal subtype breast cancer patients, $n=998$ luminal A subtype breast cancer patients, $n=673$ luminal B subtype breast cancer patients, $n=401$ HER2+ subtype breast cancer patients, and $n=63$ normal-like subtype breast cancer patients.

Results

CDC20 is among the genes whose expression is most different when comparing basal-like breast cancers with breast cancers of other molecular subtypes.

We used published microarray data (5) to compare the whole tumor transcriptomes of 23 patients with basal subtype breast cancer and 72 patients with breast cancer of other molecular subtypes (luminal A, luminal B, HER2+ and normal-like) to identify genes whose expression marked the basal-like molecular subtype. We observed transcriptome-wide differential expression of cell division cycle 20, CDC20, when comparing the tumors of patients with basal-like cancer with that of patients with luminal A, luminal B and HER2+ cancers (Table 1). When sorting each of the transcripts whose expression was measured globally, CDC20 expression ranked 13 out of 31157 total transcripts, equating to 99.96% differential expression; this was statistically significant (p -value: $3.10E-15$).

Analysis of a second published microarray dataset, comparing tumor transcriptome data from a separate cohort of patients (14 patients with basal-like breast cancer and 109 patients with breast cancer of other molecular subtype (luminal A, luminal B, HER2+ and normal-like) (5) revealed transcriptome-wide differential expression of CDC20 in patients with basal subtype human breast cancer (Table 2). When sorting each of the transcripts whose expression was measured globally, CDC20 expression ranked 119 out

of 33297 total transcripts, equating to 99.6% differential expression; this was statistically significant (p -value: $3.86\text{E-}14$). Thus, the data suggested differential expression of CDC20 was a fundamental feature of basal-like breast cancer in humans rather than an artifact of a single microarray dataset or research group, or a finding derived from cell culture or an animal model that was not relevant to human breast cancer.

CDC20 expression is significantly higher in basal-like breast cancers.

We obtained exact CDC20 mRNA expression data for basal-like breast tumors and tumors from other molecular subtypes (luminal A, luminal B, HER2+ and normal-like) to determine magnitude and direction of CDC20 gene expression change in basal-like breast cancer. CDC20 was expressed at higher levels in the tumors of patients with basal-like breast cancer as compared to the tumors of patients from other PAM50 molecular subtypes, and this difference was statistically significant (Figure 1; $p < 0.0001$). CDC20 was expressed at 4.938 ± 1.910 arbitrary units (A.U.) in basal-like breast cancer and 1.772 ± 1.471 A.U. in all other molecular subtypes (Figure 1).

CDC20 expression correlates with distant metastasis-free survival in human breast cancer.

We asked whether primary tumor expression of CDC20 was correlated with survival outcomes in patients with breast cancer when analyzing outcomes based on PAM50 molecular subtype. We observed a correlation between primary tumor expression of CDC20 and distant metastasis-free survival (DMFS) in patients with breast cancer, irrespective of molecular subtype, which approached statistical significance (Figure 2; p -value: $3.5\text{e-}15$; hazard ratio: 1.87 (1.6-2.19)). CDC20 primary tumor expression was a negative prognostic indicator in patients with breast cancer. Patients with low tumor expression of CDC20 possessed median DMFS of 137 months, while patients with high tumor expression of CDC20 possessed median DMFS of 41.64 months (Table 3).

CDC20 primary tumor expression was not correlated with distant metastasis-free survival in basal subtype breast cancer (Figure 2; p -value: 0.26; hazard ratio: 0.85 (0.63-1.13)), in luminal A subtype breast cancer (Figure 2; p -value: 0.29; hazard ratio: 1.19 (0.86-1.66)), in luminal B subtype breast cancer (Figure 2; p -value: 0.36; hazard ratio: 1.15 (0.85-1.54)), in HER2+ breast cancer (Figure 2; p -value: 0.17; hazard ratio: 1.28 (0.9-1.82)), or in normal-like breast cancer (Figure 2; p -value: 0.42; hazard ratio: 1.48 (0.57-3.85)).

Discussion

We find here that CDC20 is among the genes whose expression is most different when comparing basal-like breast cancers (BLBC) to cancers of other molecular subtypes, with significantly more CDC20 mRNA in BLBC. Kaplan-Meier analysis of patient survival data revealed that CDC20 expression was correlated with distant metastasis-free survival in patients with breast cancer. CDC20 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. She, Q.B., Gruvberger-Saal, S.K., Maurer, M., Chen, Y., Jumppanen, M., Su, T., Dendy, M., Lau, Y.K.I., Memeo, L., Horlings, H.M. and van de Vijver, M.J., 2016. Integrated molecular pathway analysis informs a synergistic combination therapy targeting PTEN/PI3K and EGFR pathways for basal-like breast cancer. *BMC cancer*, 16(1), pp.1-16.
6. 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.
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.

Rank: 13
probe ID: A_23_P149200
p-value: 3.10E-15
t: 9.333136
B: 24.12106
Gene: CDC20
Gene name: cell division cycle 20

Table 1: CDC20 is among the genes whose expression is most different, across the transcriptome, when comparing basal-like breast cancers with luminal A, luminal B, and HER2+ 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: 119
probe ID: 7900699
 p -value: 3.86E-14
t: 8.545521
B: 21.71868
Gene: CDC20
Gene name: cell division cycle 20

Table 2: In a separate cohort of patients with basal subtype breast cancer, CDC20 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+ and normal-like.

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: CDC20 expression is significantly higher in basal-like breast cancers.

Depicted is mRNA expression of CDC20 in tumors from patients with basal-like breast cancer (left) and from patients with cancer of all other molecular subtypes (luminal A, luminal B, HER2+ and normal-like) (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 2: CDC20 primary tumor expression is correlated with distant metastasis-free survival in patients with breast cancer.

Depicted in this Kaplan-Meier plot is the probability of distant metastasis-free survival for $n=2765$ patients (above, left), $n=630$ basal subtype patients (above, middle), $n=401$ HER2+ subtype patients (above, right), $n=998$ luminal A subtype patients (below, left), $n=673$ luminal B subtype patients (below, middle), and $n=63$ normal-like subtype patients (below, right) stratified into two groups, based on low or high expression of CDC20 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 CDC20 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 CDC20 expression: 137 months
High CDC20 expression: 41.64 months

Table 3: Median DMFS is superior in breast cancer patients with low primary tumor expression of CDC20.

The median DMFS (distant metastasis-free survival) of $n=2765$ breast cancer patients, with low primary tumor expression of CDC20 and high primary tumor expression of CDC20 is listed in this chart.