Differential expression of centromere protein K in triple negative breast cancer.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY 11730

Women diagnosed with triple negative breast cancer can benefit neither from endocrine therapy nor from HER2-targeted therapies¹. We mined published microarray datasets^{2,3} to determine in an unbiased fashion and at the systems level genes most differentially expressed in the primary tumors of patients with breast cancer. We report here significant differential expression of the gene encoding centromere protein K, CENPK, when comparing the tumor cells of patients with triple negative breast cancer to normal mammary ductal cells². CENPK was also differentially expressed in bulk tumor in human breast cancer³. CENPK mRNA was present at significantly increased quantities in TNBC tumor cells relative to normal mammary ductal cells. Analysis of human survival data revealed that expression of CENPK in primary tumors of the breast was correlated with recurrence-free survival in patients with basal-like and luminal A type cancer, while within triple negative breast cancer, primary tumor expression of CENPK was correlated with overall survival in patients with basal-like 2 subtype disease. CENPK may be of relevance to initiation, maintenance or progression of triple negative breast cancers.

Keywords: breast cancer, CENPK, centromere protein K, systems biology of breast cancer, targeted therapeutics in breast cancer.

Patients with hormone receptor positive disease can benefit from endocrine therapy⁴: aromatase inhibition such as that provided by exemestane or letrozole, or estrogen receptor antagonism such as that provided by tamoxifen or fulvestrant. Patients whose tumors overexpress the human epidermal growth factor 2 (HER2) can be treated with HER2-targeted therapies: HER2 inhibitors, immunoglobulin⁵- or small molecule-based⁶. As triple negative breast cancer (TNBC) is defined by its lack of HER2 expression as well as by its lack of expression of the hormone receptor estrogen, patients diagnosed with triple negative breast cancer will not benefit from either of these treatment strategies.

Understanding the transcriptional composition of triple negative breast cancers in humans can facilitate development of novel therapeutics and contribute to efforts to prevent progression to metastatic stages. We mined published microarray data^{2,3} to understand at the transcriptome level and in an unbiased fashion genes most differentially expressed in the tumor cells of patients with triple negative breast cancer.

Methods

We utilized datasets GSE38959² and GSE109169³ for this differential gene expression analysis of female breast cancer in conjunction with GEO2R. 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. GSE109169 was generated using Affymetrix Human Exon 1.0 ST Array technology with $n=25$ normal breast tissue and $n=25$ tumors of the breast; analysis was performed using platform GPL5175. The tissues from this dataset are paired tissues (25 tumors matching 25 normal breast tissues from 25 patients). 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 CENPK expression was significantly different between primary breast tumors and breast tissue using a two-tailed t-test.

For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter tool⁷ for correlation of CENPK mRNA expression levels with recurrence-free survival (RFS) in $n=296$ patients with basal subtype cancer, $n=222$ patients with luminal A subtype cancer, $n=200$ patients with luminal B subtype cancer, $n=198$ patients with HER2+ cancer, and $n=27$ patients with normal-like subtype cancer. In patients with triple negative breast cancer⁷, we analyzed overall survival (OS) for $n=73$ patients with basal-like 1 subtype TNBC, $n=45$ patients with basal-like 2 subtype TNBC, $n=90$ patients with immunomodulatory subtype TNBC, $n=62$ patients with luminal androgen receptor subtype TNBC, $n=78$ patients with mesenchymal subtype TNBC, and $n=33$ patients with mesenchymal stem-like subtype TNBC.

Results

We performed discovery of genes associated with triple negative breast cancer in humans by mining two independently published microarray datasets^{2,3}.

CENPK is differentially expressed in primary tumors and tumor cells of the breast.

Comparison of the tumor cells of patients with triple negative breast cancer to normal mammary ductal cells² revealed centromere protein K, encoded by CENPK, as among the genes whose expression was most significantly different in human TNBC as compared to the normal breast (Chart 1). 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 ductal cells, CENPK ranked 241 out of 45015 total transcripts, equating to 99.5% differential expression (Chart 1). Differential expression of CENPK in the tumor cells of patients with triple negative breast cancer was statistically significant (Chart 1; $p=7.45E-11$).

Analysis of a second microarray dataset³, here comparing 25 normal breast tissues to 25 tumors of the breast revealed that CENPK was also among the genes most differentially expressed in human breast cancer, here in the tumors of women diagnosed with early-onset breast cancer (Chart 2). When sorting each of the genes expressed in tumors of the breast based on significance of difference as compared to normal breast tissue, CENPK ranked 277 out of 19076 total transcripts, equating to 98.5% differential expression (Chart 2). Differential expression of CENPK in female breast cancer was statistically significant (Chart 2; $p=2.45E-12$).

These data indicated that differential expression of CENPK was not an artifact of a single microarray dataset, was not an artifact of FACS isolation, and suggested that perturbed expression of CENPK was likely a transcriptional feature of triple negative breast cancers and early-onset human breast cancers.

CENPK is expressed at significantly higher levels in triple negative tumor cells as compared to normal cells of the breast.

We obtained exact mRNA expression levels for CENPK - from normal mammary ductal cells of the breast and from tumor cells of patients with triple negative breast cancer - to understand the magnitude and direction of CENPK expression change in TNBC. CENPK mRNA transcript was present at higher levels in triple negative tumor cells of the breast as compared to normal mammary ductal cells (Figure 1). Increased expression of CENPK in primary tumor-isolated triple negative breast cancer cells was statistically significant (Figure 1: $p<0.0001$). CENPK was expressed at 7.70 ± 6.44 arbitrary units (A.U.) in normal mammary ductal cells, while it was expressed at 48.73 ± 25.92 A.U. in tumor cells of patients with triple negative breast cancer. We calculated a mean fold change of 6.33 when comparing triple negative breast cancer cells to normal mammary ductal cells.

CENPK expression correlates with survival outcomes in basal-like and luminal A human breast cancer.

We performed Kaplan-Meier survival analysis to evaluate potential relationships between primary tumor CENPK mRNA expression levels and survival outcomes in patients with breast cancer, when analyzing outcomes based on the molecular subtype of the patient (PAM50 classification): luminal A, luminal B, HER2+, normal-like, and basal-like. We observed a correlation between CENPK expression and recurrence-free survival (RFS) in patients with basal-like subtype breast cancer which trended towards statistical significance (Figure 2; log rank p -value: 0.089 for recurrence-free survival, hazard ratio: 0.67 (0.42-1.07) (Fig. 2)). CENPK mRNA levels were a positive prognostic indicator in basal-like breast cancer patients. RFS was, on average, markedly shorter for basal-like patients with low tumor expression of CENPK as compared to basal-like patients with high tumor expression of CENPK (Chart 3).

We also observed a correlation between CENPK expression and recurrence-free survival (RFS) in patients with luminal A subtype breast cancer which trended towards statistical significance (Figure 2; log rank p -value: 0.063 for recurrence-free survival, hazard ratio: 2.06 (0.95-4.47) (Fig. 2)). CENPK mRNA levels were a negative prognostic indicator in luminal A breast cancer patients. RFS was, on average, markedly greater for luminal A patients with low tumor expression of CENPK as compared to luminal A patients with high tumor expression of CENPK (Chart 3).

CENPK primary tumor expression was not correlated with recurrence-free survival in luminal B subtype (Figure 2; log rank p -value: 0.35 for RFS, hazard ratio: 0.77 (0.44-1.34) (Fig. 2)), in HER2+ breast cancer (Figure 2; log rank p -value: 0.46 for RFS, hazard ratio: 1.23 (0.71-2.13) (Fig. 2)), or in the normal-like subtype (Figure 2; log rank p -value: 0.85 for RFS, hazard ratio: 0.88 (0.23-3.33) (Fig. 2)).

Influence of primary tumor CENPK expression on survival outcomes based on TNBC molecular subtype.

We also performed Kaplan-Meier survival analysis to examine relationships between CEMRNA expression levels with survival strictly in patients with triple negative breast cancer, analyzing outcomes based on the TNBC molecular subtype⁸ of the patient: basal-like 1, basal-like 2, immunomodulatory, mesenchymal, mesenchymal stem-like, and luminal androgen receptor. We observed a statistically significant correlation between CENPK expression and overall survival (OS) in TNBC patients with basal-like 2 subtype disease (Figure 3; log rank p -value: 0.012 for overall survival, hazard ratio: 5.64 (1.23-25.87) (Fig. 3)). CENPK mRNA levels were a negative prognostic indicator in patients with basal-like 2 TNBC. OS was, on average, markedly greater for basal-like 2 TNBC patients with low tumor expression of CENPK as compared to basal-like 2 TNBC patients with high tumor expression of CENPK (Fig. 3).

CENPK primary tumor expression was not correlated with overall survival in basal-like 1 subtype (Figure 3; log rank p -value: 0.92 for OS, hazard ratio: 0.95 (0.38-2.4) (Fig. 3)), immunomodulatory subtype (Figure 3; log rank p -value: 0.8 for OS, hazard ratio: 0.86 (0.26-2.81) (Fig. 3)), in the mesenchymal subtype (Figure 3; log rank p -value: 0.25 for OS, hazard ratio: 0.64 (0.3-1.38) (Fig. 3)), in the mesenchymal stem-like subtype (Figure 3; log rank p -value: 0.61 for OS, hazard ratio: 0.76 (0.26-2.21) (Fig. 3)), or in the luminal androgen receptor subtype (Figure 3; log rank p -value: 0.32 for OS, hazard ratio: 0.66 (0.28-1.52) (Fig. 3)).

Discussion

We performed comparative transcriptome analysis using two independently published microarray datasets^{2,3} to define and catalog transcriptional differences in human triple negative breast cancer. We provided evidence here that differential expression of centromere protein K, encoded by the CENPK gene, is a defining transcriptional feature of triple negative breast cancer. CENPK was also differentially expressed in the bulk tumors of patients with early-onset breast cancer. Expression of CENPK was correlated with recurrence-free survival in patients with basal-like and luminal A subtype breast cancer, and within triple negative breast cancer patients, primary tumor expression of CENPK was correlated with overall survival in patients with basal-like 2 subtype disease. Centromere protein K and the molecular processes to which it pertains to may be relevant to the initiation or progression of triple negative breast cancer in humans.

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. 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.
3. Chang, J.W., Kuo, W.H., Lin, C.M., Chen, W.L., Chan, S.H., Chiu, M.F., Chang, I.S., Jiang, S.S., Tsai, F.Y., Chen, C.H. and Huang, P.H., 2018. Wild-type p53 upregulates an early onset breast cancer-associated gene GAS7 to suppress metastasis via GAS7–CYFIP1- mediated signaling pathway. *Oncogene*, 37(30), pp.4137-4150.
4. Spring, L.M., Gupta, A., Reynolds, K.L., Gadd, M.A., Ellisen, L.W., Isakoff, S.J., Moy, B. and Bardia, A., 2016. Neoadjuvant endocrine therapy for estrogen receptor–positive breast cancer: a systematic review and meta-analysis. *JAMA oncology*, 2(11), pp.1477-1486.
5. Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M. and Pinter, T., 2011. Adjuvant trastuzumab in HER2-positive breast cancer. *New England journal of medicine*, 365(14), pp.1273-1283.
6. Park, J.W., Liu, M.C., Yee, D., Yau, C., van't Veer, L.J., Symmans, W.F., Paoloni, M., Perlmutter, J., Hylton, N.M., Hogarth, M. and DeMichele, A., 2016. Adaptive randomization of neratinib in early breast cancer. *New England Journal of Medicine*, 375(1), pp.11-22.
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. Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y. and Pietenpol, J.A., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation*, 121(7), pp.2750-2767.

Rank: 241
Probe ID: 39562
p-value: 7.45E-11
t: -8.4517517
B: 14.654739
Gene: CENPK
Gene name: centromere protein K

Chart 1: CENPK is differentially expressed in 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: 277
Probe ID: 2859667
p-value: 2.45E-12
t: -9.15
B: 17.909297
Gene: CENPK
Gene name: centromere protein K

Chart 2: CENPK is differentially expressed in early-onset human 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: CENPK is expressed at significantly higher levels in primary TNBC tumor cells as compared to normal cells of the breast.

The mRNA expression level of CENPK in normal mammary ductal cells (left) and in the isolated primary tumor cells of patients with triple negative breast cancer (right) is graphically depicted with the result of a statistical test evaluating significance of difference in CENPK expression between normal cells of the breast and TNBC tumor cells, a *p*-value, listed above.

Figure 2: CENPK expression correlates with recurrence-free survival in patients with basal-like and luminal A breast cancer, but not in patients with luminal B, HER2+, or normal-like subtype cancer.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival (RFS) for $n=296$ patients with basal subtype breast cancer (upper left), $n=198$ patients with HER2+ breast cancer (upper right), $n=222$ patients with luminal A breast cancer (middle left), $n=200$ patients with luminal B breast cancer (middle right), and $n=27$ patients with normal-like breast cancer (below) stratified into two groups, based on low or high expression of

CENPK 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 CENPK 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 3: CENPK expression correlates with overall survival in triple negative breast cancer patients whose tumors are of the basal-like 2 molecular subtype.

Depicted in this Kaplan-Meier plot is the probability of overall survival (OS) for $n=73$ patients with basal-like 1 subtype TNBC (upper left), $n=45$ patients with basal-like 2 subtype TNBC (upper right), $n=90$ patients with immunomodulatory subtype TNBC (middle left), $n=62$ patients with luminal androgen receptor subtype TNBC (middle right), $n=78$ patients with mesenchymal subtype TNBC (lower left), and $n=33$ patients with mesenchymal stem-like subtype TNBC (lower right), stratified into two groups, based on low or high expression of CENPK in the primary tumors of patients with TNBC. 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 CENPK 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.