Over-expression of centromere protein F in human endometrial cancer.

Shahan Mamoor, MS¹
¹shahanmamoor@gmail.com
East Islip, NY USA

Gynecologic cancers including cancers of the endometrium are a clinical problem¹⁻⁴. We mined published microarray data^{5,6} to discover genes associated with endometrial cancers by comparing transcriptomes of the normal endometrium and endometrial tumors from humans. We identified centromere protein F, encoded by CENPF, as among the most differentially expressed genes, transcriptome-wide, in cancers of the endometrium. CENPF was expressed at significantly higher levels in endometrial tumor tissues as compared to the endometrium. Importantly, primary tumor expression of CENPF was correlated with overall survival in patients with endometrial cancer. CENPF may be a molecule of interest in understanding the etiology or progression of human endometrial cancer.

Keywords: endometrial cancer, gynecologic cancers, endometrium, CENPF, centromere protein F, systems biology of endometrial cancer, targeted therapeutics in endometrial cancer.

Endometrial cancer is the most common gynecologic cancer in the developed world¹. Over the last three decades, the incidence of endometrial cancer has increased 21%⁴ and the death rate has increased 100%³. We harnessed the power of independently published microarray datasets^{5,6} to determine in an unbiased fashion and at the systems-level genes most differentially expressed in endometrial tumors. We report here the differential and increased expression of the centromere protein F (CENPF) in human endometrial cancer.

Methods

We utilized datasets GSE63678⁵ and GSE17025⁶ for this global differential gene expression analysis of human endometrial cancer in conjunction with GEO2R. GSE63678 was generated using Affymetrix Human Genome U133A 2.0 Array technology with $n=5$ control endometrial tissues (including $n=4$ uterine myomas and $n=1$ benign cyst) and $n=7$ endometrial cancers (including $n=2$ endometrial adenocarcinomas, $n=3$ mixed endometrioid adenocarcinomas, and $n=2$ adenocarcinomas with squamous differentiation); analysis was performed using platform GPL571. GSE17025 was generated using Affymetrix Human Genome U133 Plus 2.0 Array technology with $n=5$ normal endometrial tissues (inactive endometrium) and $n=91$ endometrial cancers; analysis was performed using platform GPL570. 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 CENPF gene expression was significantly different between control endometrial tissue and endometrial tumor tissue in humans using a two-tailed t-test. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter tool⁷ for correlation of CENPF mRNA expression levels with overall survival in $n=543$ endometrial cancer patients.

Results

We harnessed the power of blind comparative transcriptome analysis using published microarray data^{5,6} to discover in an unbiased fashion genes associated with endometrial cancer in humans.

CENPF is differentially expressed in endometrial cancer.

We identified centromere protein F, encoded by CENPF, as among the genes most differentially expressed in cancers of the endometrium when compared to benign endometrial tissues (Chart 1). When sorting each of the genes expressed in endometrial tumor tissue based on significance of change in expression as compared to benign endometrial tissue, CENPF ranked 52 out of 22273 transcripts, equating to 99.8% differential expression (Chart 1). Differential expression of CENPF in human endometrial cancers was statistically significant (Chart

1; $p=3.78E-05$).

We queried a second microarray data to validate differential expression of CENPF in endometrial cancer. Again, we observed differential expression of CENPF when comparing endometrial tumor tissue to benign endometrial tissue (Chart 2). When sorting each of the genes expressed in endometrial tumor tissue based on significance of change in expression as compared to benign endometrial tissue, CENPF ranked 63 out of 54675 transcripts, equating to 99.9% differential expression (Chart 2). Differential expression of CENPF in human endometrial cancers was statistically significant (Chart 2; $p=2.36E-08$).

CENPF is expressed at significantly higher levels in endometrial cancers as compared to benign endometrial tissue.

We obtained exact mRNA expression levels for CENPF in endometrial tumor tissues and from benign endometrial tissue to evaluate direction and statistical significance of change in expression of CENPF in human endometrial cancer. CENPF was expressed at higher levels in endometrial tissue as compared to normal endometrial tissue, and this difference was statistically significant (Figure 1; $p<0.0001$). We calculated a mean fold change of 1.34 in CENPF mRNA levels in human endometrial cancer, as CENPF was expressed at 6.57 ± 0.28 arbitrary units (A.U.) in control endometrial tissue but at 8.77 ± 0.80 A.U. in endometrial tumor tissue.

CENPF expression is correlated with patient survival outcomes in endometrial cancer.

We performed Kaplan-Meier survival analysis to evaluate correlation between CENPF primary tumor expression and survival outcomes in 543 patients with endometrial cancer. We observed a correlation between primary tumor expression of CENPF and overall survival (OS) in patients with endometrial cancer (Figure 2). Though median OS data suggested that CENPF mRNA levels were a positive prognostic indicator, CENPF mRNA levels were otherwise a negative prognostic indicator in patients with endometrial cancer (Figure 2; note the shift in outcomes at probability=0.5). Patients whose primary tumors expressed low levels of CENPF possessed median OS of 111.63 months as compared to a median OS of 114.1 months for patients whose tumors expressed high levels of CENPF (Figure 2). This difference in OS based on CENPF tumor expression in patients with endometrial cancer trended towards statistical significance (Figure 2, Chart 3; logrank p -value: 0.084; hazard ratio: 1.44 (0.95-2.2)).

Thus, by mining published microarray data^{5,6} in an unbiased and systematic fashion, we identified centromere protein F, encoded by CENPF, as among the genes whose expression was most different, transcriptome-wide, in the endometrial tumor tissue of patients with endometrial cancer when compared to benign endometrial tissue; we observed significantly increased expression of CENPF in endometrial tumor tissue as compared to benign endometrial tissue. Further, we found a correlation between CENPF expression and patient survival outcomes in human endometrial cancer, as overall survival was generally inferior in patients whose tumors

expressed higher levels of CENPF as compared to patients whose tumors expressed lower levels of CENPF.

Discussion

We provided evidence here that centromere protein F is among the genes most differentially expressed in human endometrial cancer, that mRNA for CENPF is present at significantly increased quantity in endometrial tumor tissue as compared to benign endometrium, and that CENPF primary tumor expression is correlated with overall survival in patients with endometrial cancer. These data suggest CENPF may be of importance to fundamental biological processes that underlie the initiation, progression or maintenance of human endometrial cancer.

References

1. Amant, F., Moerman, P., Neven, P., Timmerman, D., Van Limbergen, E. and Vergote, I., 2005. Endometrial cancer. *The Lancet*, 366(9484), pp.491-505.
2. Sorosky, J.I., 2008. Endometrial cancer. *Obstetrics & Gynecology*, 111(2), pp.436-447.
3. Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N. and Darai, E., 2016. Endometrial cancer. *The Lancet*, 387(10023), pp.1094-1108.
4. Sorosky, J.I., 2012. Endometrial cancer. *Obstetrics & Gynecology*, 120(2 Part 1), pp.383-397.
5. Pappa, K.I., Polyzos, A., Jacob-Hirsch, J., Amariglio, N., Vlachos, G.D., Loutradis, D. and Anagnou, N.P., 2015. Profiling of discrete gynecological cancers reveals novel transcriptional modules and common features shared by other cancer types and embryonic stem cells. *PLoS One*, 10(11), p.e0142229.
6. Day, R.S., McDade, K.K., Chandran, U.R., Lisovich, A., Conrads, T.P., Hood, B.L., Kolli, V.K., Kirchner, D., Litzi, T. and Maxwell, G.L., 2011. Identifier mapping performance for integrating transcriptomics and proteomics experimental results. *BMC bioinformatics*, 12(1), p.213.
7. Nagy, A., Munkacsy, G. and Gyorffy, B., 2020. Pancancer survival analysis of cancer hallmark genes. *bioRxiv*.

Rank: 52
Probe ID: 207828_s_at
p-value: 3.78E-05
t: -6.2884816
B: 2.54438
Gene: CENPF
Gene name: centromere protein F

Chart 1: CENPF is differentially expressed in endometrial cancer when comparing primary endometrial tumors to benign endometrial tissue.

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.

Rank: 63
probe ID: 209172_s_at
p-value: 2.36E-08
t: -6.0734409
B: 8.88718
Gene: CENPF
Gene name: centromere protein F

Chart 2: CENPF is differentially expressed in endometrial cancer when comparing primary endometrial tumors to benign endometrial tissue.

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: CENPF is expressed at significantly higher levels in the endometrial tumors of patients with endometrial cancer when compared to benign endometrium.

The mRNA expression level of CENPF in benign endometrial tissue (left) and in primary tumors of the endometrium (right) is graphically depicted; the result of a statistical test evaluating significance of difference in CENPF expression between benign endometrial tissue and primary tumors of the endometrial tissue is $p < 0.0001$.

Figure 2: Correlation between CENPF primary tumor expression and overall survival in patients with endometrial cancer.

Depicted in this Kaplan-Meier plot is the probability of overall survival for $n=543$ total endometrial cancer patients stratified into two groups, based on low or high expression of CENPF 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 CENPF 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 CENPF expression: 111.63 months

High CENPF expression: 114.1 months

Chart 3: Median overall survival is superior in endometrial cancer patients with high primary tumor expression of CENPF, though overall survival is otherwise generally superior in endometrial cancer patients with low primary tumor expression of CENPF.

The median OS (overall survival) of patients with low primary tumor expression of CENPF and high primary tumor expression of CENPF is listed in this chart, for $n=543$ endometrial cancer patients.