Over-expression of CUB domain containing protein 1 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 CUB domain containing protein 1, encoded by CDCP1, as among the most differentially expressed genes, transcriptome-wide, in cancers of the endometrium. CDCP1 was expressed at significantly higher levels in endometrial tumor tissues as compared to the endometrium. Importantly, in human endometrial cancer, primary tumor expression of CDCP1 was correlated with overall survival in endometrial cancer patients with Grade 2 tumors. CDCP1 may be a molecule of interest in understanding the etiology or progression of human endometrial cancer.

Keywords: endometrial cancer, gynecologic cancers, endometrium, CDCP1, CUB domain containing protein 1, 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 CUB domain containing protein 1 (CDCP1) 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 CDCP1 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 CDCP1 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.

CDCP1 is differentially expressed in endometrial cancer.

We identified CUB domain containing protein 1, encoded by CDCP1, 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, CDCP1 ranked 173 out of 22273 transcripts, equating to 99.2% differential expression (Chart 1). Differential expression of CDCP1 in human endometrial cancers was statistically significant (Chart 1; $p=2.06E-04$).

We queried a second microarray data to validate differential expression of CDCP1 in endometrial cancer. Again, we observed differential expression of CDCP1 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, CDCP1 ranked 263 out of 54675 transcripts, equating to 99.5% differential expression (Chart 2). Differential expression of CDCP1 in human endometrial cancers was statistically significant (Chart 2; $p=2.05E-06$).

CDCP1 is expressed at significantly higher levels in endometrial cancers as compared to benign endometrial tissue.

We obtained exact mRNA expression levels for CDCP1 in endometrial tumor tissues and from benign endometrial tissue to evaluate direction and statistical significance of change in expression of CDCP1 in human endometrial cancer. CDCP1 was expressed at higher levels in endometrial tissue as compared to normal endometrial tissue, and this difference was statistically significant (Figure 1; $p=0.0009$). We calculated a mean fold change of 1.27 in CDCP1 mRNA levels in human endometrial cancer, as CDCP1 was expressed at 7.10 ± 0.65 arbitrary units (A.U.) in control endometrial tissue but at 8.98 ± 0.68 A.U. in endometrial tumor tissue.

CDCP1 expression is correlated with patient survival outcomes in endometrial cancer.

We performed Kaplan-Meier survival analysis to evaluate correlation between CDCP1 primary tumor expression and survival outcomes in 543 patients with endometrial cancer based on differences in tumor grade. We observed a correlation between primary tumor expression of CDCP1 and overall survival in patients with endometrial cancer, in patients with Grade 2 tumors burden (Figure 2). CDCP1 primary tumor mRNA levels were a negative prognostic indicator in endometrial cancer patients with Grade 2 tumors. Patients with Grade 2 tumors whose primary tumors expressed low levels of CDCP1 possessed, on average, markedly greater median OS as compared to patients with Grade 2 tumors whose tumors expressed high levels of CDCP1. This difference in OS based on CDCP1 tumor expression in patients with endometrial cancer with Grade 2 tumors was statistically significant (Figure 2, Chart 3; logrank p -value: 0.049; hazard ratio: 3.03 (0.95-9.68)). CDCP1 primary endometrial tumor expression was not correlated with overall survival in endometrial cancer patients with Grade 1 tumors (Figure 2, Chart 3; logrank p -value: 0.14; hazard ratio: 643643313.17 (0 - ∞)), nor in endometrial cancer patients with Grade 3 tumors ((Figure 2, Chart 3; logrank p -value: 0.99; hazard ratio: 1 (0.62-1.6)).

Thus, by mining published microarray data^{5,6} in an unbiased and systematic fashion, we identified CUB domain containing protein 1, encoded by CDCP1, 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 CDCP1 in endometrial tumor tissue as compared to benign

1 endometrial tissue. Further, we found a correlation between CDCP1 expression and patient
2 survival outcomes in human endometrial cancer, as overall survival was inferior in patients
3 whose tumors expressed higher levels of CDCP1 as compared to patients whose tumors
4 expressed lower levels of CDCP1, in endometrial patients with Grade 2 tumors, but not in
endometrial cancer patients with Grade 1 or Grade 3 tumors.

5 **Discussion**

6 We provided evidence here that CUB domain containing protein 1 is among the genes most
7 differentially expressed in human endometrial cancer, that mRNA for CDCP1 is present at
8 significantly increased quantity in endometrial tumor tissue as compared to benign endometrium,
9 and that CDCP1 primary tumor expression is correlated with overall survival in endometrial
10 cancer patients with Grade 2 tumors. These data suggest CDCP1 may be of importance to
11 fundamental biological processes that underlie the initiation, progression or maintenance of
12 human endometrial cancer.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

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., Litz, 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: 173
Probe ID: 218451_at
p-value: 2.06E-04
t: -5.2176614
B: 0.96618
Gene: CDCP1
Gene name: CUB domain containing protein 1

Chart 1: CDCP1 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: 263
probe ID: 234932_s_at
p-value: 2.05E-06
t: -5.047823
B: 4.72845
Gene: CDCP1
Gene name: CUB domain containing protein 1

Chart 2: CDCP1 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: CDCP1 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 CDCP1 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 CDCP1 expression between benign endometrial tissue and primary tumors of the endometrial tissue is $p=0.0009$.

Figure 2: Correlation between CDCP1 primary tumor expression and overall survival in endometrial cancer, in patients with Grade 2 tumors, but not in patients with Grade 1 or Grade 3 tumors.

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 CDCP1 in patient primary tumors, in patients with Grade 1 tumors (top left), in patients with Grade 2 tumors (top right), and in patients with Grade 3 tumors (bottom). 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 CDCP1 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.