JOURNAL OF GYNECOLOGIC ONCOLOGY

Original Article

Check for updates

Correspondence to
Taesung Park
Department of Statistics, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea.
Email: tspark@stats.snu.ac.kr

Hyojin Kim
Department of Pathology, Seoul National University Bundang Hospital, 82, Gumi-ro 173beon-gil, Bundang-gu, Seongnam, Korea.
Email: hyojinkim7137@gmail.com

†Kidong Kim and Suhyun Hwangbo contributed equally to the manuscript as co-first authors.

ORCID iDs
Kidong Kim
https://orcid.org/0000-0001-9254-6024
Suhyun Hwangbo
https://orcid.org/0000-0001-9294-1518

Clinicopathologic and protein markers distinguishing the "polymerase epsilon exonuclease" from the "copy number low" subtype of endometrial cancer

Kidong Kim [iD],[1,†] **Suhyun Hwangbo** [iD],[2,†] **Hyojin Kim** [iD],[3] **Yong Beom Kim** [iD],[1] **Jae Hong No** [iD],[1] **Dong Hoon Suh** [iD],[1] **Taesung Park** [iD] [4]

[1]Department of Obstetrics and Gynecology, Seoul National University Bundang Hospital, Seongnam, Korea
[2]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Korea
[3]Department of Pathology, Seoul National University Bundang Hospital, Seongnam, Korea
[4]Department of Statistics, Seoul National University, Seoul, Korea

ABSTRACT

Objective: The need to perform genetic sequencing to diagnose the polymerase epsilon exonuclease (*POLE*) subtype of endometrial cancer (EC) hinders the adoption of molecular classification. We investigated clinicopathologic and protein markers that distinguish the *POLE* from the copy number (CN)-low subtype in EC.

Methods: Ninety-one samples (15 *POLE*, 76 CN-low) were selected from The Cancer Genome Atlas EC dataset. Clinicopathologic and normalized reverse phase protein array expression data were analyzed for associations with the subtypes. A logistic model including selected markers was constructed by stepwise selection using area under the curve (AUC) from 5-fold cross-validation (CV). The selected markers were validated using immunohistochemistry (IHC) in a separate cohort.

Results: Body mass index (BMI) and tumor grade were significantly associated with the *POLE* subtype. With BMI and tumor grade as covariates, 5 proteins were associated with the EC subtypes. The stepwise selection method identified BMI, cyclin B1, caspase 8, and X-box binding protein 1 (XBP1) as markers distinguishing the *POLE* from the CN-low subtype. The mean of CV AUC, sensitivity, specificity, and balanced accuracy of the selected model were 0.97, 0.91, 0.87, and 0.89, respectively. IHC validation showed that cyclin B1 expression was significantly higher in the *POLE* than in the CN-low subtype and receiver operating characteristic curve of cyclin B1 expression in IHC revealed AUC of 0.683.

Conclusion: BMI and expression of cyclin B1, caspase 8, and XBP1 are candidate markers distinguishing the *POLE* from the CN-low subtype. Cyclin B1 IHC may replace *POLE* sequencing in molecular classification of EC.

Keywords: Endometrial Neoplasms; Classification; Cyclin B1; Caspase 8; X-Box Binding Protein 1; Immunohistochemistry

Generated by xmlinkpress

Hyojin Kim
https://orcid.org/0000-0001-9201-8328
Yong Beom Kim
https://orcid.org/0000-0003-1196-369X
Jae Hong No
https://orcid.org/0000-0002-2389-6757
Dong Hoon Suh
https://orcid.org/0000-0002-4312-966X
Taesung Park
https://orcid.org/0000-0002-8294-590X

Synopsis

Body mass index and cyclin B1, caspase 8, and X-box binding protein 1 are candidate markers distinguishing between the polymerase epsilon exonuclease (*POLE*) and copy number (CN)-low subtypes of endometrial cancer. Cyclin B1 immunohistochemistry expression was significantly higher in the *POLE* than in the CN-low subtype and may substitute *POLE* sequencing.

INTRODUCTION

Endometrial cancer (EC) is the second most common gynecologic cancer worldwide [1], and the age-standardized incidence and prevalence of this disease are increasing globally [2]. In the past, ECs have typically been divided into histological subtypes that fall into 2 broad categories, namely type 1 (tumors are relatively slow growing and generally confined to the uterus) and type 2 (tumors are more aggressive and more likely to metastasize). More recently, molecular characterization has revealed new subtype classification strategies based on genomic features. According to a study by The Cancer Genome Atlas (TCGA) program, EC can be divided into several molecular subtypes characterized by the following abnormalities: polymerase epsilon exonuclease (*POLE*) mutations (ultramutated), microsatellite instability (MSI) (hypermutated), and somatic copy number alterations (SCNAs) featuring copy number (CN)-low (endometrioid) and CN-high (serous-like) subtypes. Each subtype exhibited a distinctive pattern of progression-free survival [3]. However, the implementation of TCGA classification in routine practice has been limited by the expense involved in genomic profiling of individual patients. Therefore, more pragmatic classifications systems using immunohistochemistry (IHC) markers were developed. For example, the Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) classification system developed by Talhouk et al. [4] assesses the presence or absence of mismatch repair (MMR) proteins and wildtype or abnormal p53 by IHC in place of genomic testing for MSI and SCNAs. The subtype, as defined by ProMisE classification, was associated with overall survival in young women with EC [5] and recurrence-free survival in women with grade 3 EC [6].

The IHC-based approach has simplified the process for the molecular classification of EC. However, gene sequencing is still required to identify the *POLE* mutant subtype, which has hindered the widespread adoption of molecular classification for this cancer [7]. A search for surrogates for *POLE* sequencing revealed that the *POLE* subtype is characterized by young age at diagnosis, lower body mass index (BMI), high proportion of grade 3 tumors, lymphovascular space invasion, endometrioid type, and low tumor stage. However, these variables overlap with other subtypes, making it difficult to categorically classify this EC type [8].

We hypothesize that identifying markers that distinguish the *POLE* from the CN-low subtype will make it possible to identify all 4 subtypes without gene sequencing, since the MSI and CN-high subtypes can be characterized by MMR proteins and p53 IHC. In this study, we conducted a statistical analysis of data associated with the *POLE* and CN-low subtypes to identify clinicopathologic and protein markers capable of distinguishing between these subtypes in EC and validated the markers using IHC in a separate cohort.

MATERIALS AND METHODS

1. Train data

After the study protocol was approved by the institutional review board (IRB No. X-2005-613-903), we identified 107 subjects with *POLE* or CN-low EC subtypes from the supplementary data published by TCGA study that originally outlined the molecular classification system for EC [3]. We extracted data regarding age, race, BMI, stage, histology, grade, and molecular subtype, and using the Firehose website (gdac.broadinstitute.org), we downloaded the protein expression data of TCGA EC cohort (filename: mda_rppa_core-protein_normalization). According to TCGA, protein expression levels were measured by reverse phase protein arrays (RPPAs) and were normalized as follows: 1) the median for each protein across all the samples was calculated; 2) the median (from step 1) was subtracted from values within each protein; 3) the median for each sample across all proteins was calculated; and 4) the median (from step 3) was subtracted from values within each sample [9]. The protein expression data contained the normalized expression levels of 208 protein markers from 440 samples. By inner joining the clinicopathologic and protein expression datasets, we created a train cohort that included 91 subjects (*POLE*: n = 15, CN-low: n = 76). Two missing values for BMI and one for stage were imputed using the multiple imputation by chained equations (MICE) technique [10].

2. Analysis and model

The association of the subtypes with age and BMI was examined using a 2-sample t-test. The association of the subtypes with race, stage, histology, and grade was evaluated using the χ^2 test or Fisher's exact test. A p-value <0.05 was considered to be statistically significant.

Protein markers were subsequently selected using a logistic regression model with the significant clinicopathologic markers as adjusting covariates. For each protein marker, we performed a likelihood ratio test by comparing the deviances of the full model ($logit(p(POLE))= \beta_0+\beta_1 BMI+\beta_2 Grade+\beta_3 Protein$) with the reduced model ($logit(p(POLE))=\beta_0+\beta_1 BMI+\beta_2 Grade$) and the q-values of the false discovery rate for multiple comparisons <0.05 were used to determine the significance. For the selected protein markers, we performed a Wilcoxon rank-sum test to compare expression levels between the 2 subtypes.

Finally, the logistic regression models encompassing the selected markers (clinicopathologic and protein markers) were constructed by a stepwise selection method using the area under the curve (AUC) from 5-fold cross-validation (CV). To determine the optimal threshold value for the final model, the sensitivity, specificity, and balanced accuracy of each threshold value were evaluated. We selected a threshold value of 0.2 for the maximum balanced accuracy. All analyses were conducted using R (version 3.6.1, R core team, 2019; R Foundation, Vienna, Austria).

3. Validation

We verified the selected protein markers by IHC staining using resected EC tissue. Specifically, we identified 24 *POLE* and 131 CN-low subtype EC tissues from 240 patients who underwent surgery at our institute from 2006 to 2013. For molecular classification, we performed digital droplet polymerase chain reaction for the hotspot of the *POLE* gene and IHC for hMLH1, hMSH2, hMSH6, PMS2 and p53 using formalin-fixed, paraffin-embedded specimens. The institutional review board approved the molecular classification and IHC of protein markers of 240 patients as a separate study (IRB No. B-2008/628-304) and waived the requirement for informed consent.

For 24 *POLE* and 131 CN-low subtype EC tissues, IHC staining for selected protein markers (cyclin B1, caspase-8, and X-box binding protein 1 [XBP1]) was performed using the Ventana BenchMark XT autostainer with primary monoclonal antibodies against cyclin B1 (RBT-B1, 1:50; LSBio, Seattle, WA, USA) and caspase-8 (90A992, 1:8,000; Thermo Fisher Scientific, Waltham, MA, USA) and polyclonal XBP1 antibody (1:1,500; LSBio).

The IHC expression levels of protein markers were evaluated by a gynecologic pathologist (K.H.). Specifically, the percentage of positive cells was measured and categorized on a scale of 0 to 4 (0=0%; 1=up to 1%; 2=1%–10%; 3=10%–50%; 4=more than 50%). In accordance with a previous study [11], intensity of staining was judged on an arbitrary scale of 0 to 4: no staining (0), weakly positive staining (1), moderately positive staining (2), strongly positive staining (3), and very strongly positive staining (4). The IHC expression level was scored by multiplying the category of percentage of positive cells (0–4) with staining intensity (0–4). Cytoplasmic and nuclear/cytoplasmic expression levels were evaluated for cyclin B1 and caspase-8, respectively. XBP1 IHC failed due to poor staining quality.

RESULTS

1. Clinicopathologic markers
Clinicopathologic data grouped according to subtype are summarized in **Table 1**. BMI and tumor grade were found to be significantly associated with EC subtype (p=0.004 and p=0.001, respectively). However, age, race, stage, and histology were not significantly associated with either subtype.

2. Protein markers and final model
Using a logistic regression model with BMI and tumor grade as adjusting covariate variables, we examined the protein expression data for correlating protein markers. Five protein markers with the lowest p-value in the univariable analysis were selected (**Table S1**). The expression levels of 5 selected protein markers (cyclin B1, p62 LCK ligand, caspase 8, FoxM1, and XBP1) according to subtype are shown in **Fig. 1**. Cyclin B1, p62 LCK ligand, and FoxM1 levels were all elevated, whereas the expression levels of caspase 8 and XBP1 were decreased, in the *POLE* compared with those in the CN-low subtype.

Table 1. Clinicopathologic markers according to subtype

Characteristic	*POLE* subtype (n=15)	CN-low subtype (n=76)	p-value
Age	58.4±13.9	60.1±11.6	0.672
Race			0.444
White	12 (80)	66 (86.8)	
Non-white	3 (20)	10 (13.2)	
BMI	29.6±6.9	36.3±9.6	0.004
Stage			0.444
1, 2	12 (80)	66 (86.8)	
3, 4	3 (20)	10 (13.2)	
Histology			1.000
Endometrioid	15 (100)	74 (97.4)	
Non-endometrioid	0 (0)	2 (2.6)	
Grade			0.001
1, 2	8 (53.3)	70 (92.1)	
3	7 (46.7)	6 (7.9)	

Values are presented as number (%) or mean ± standard deviation.
BMI, body mass index; CN, copy number; *POLE*, polymerase epsilon exonuclease.
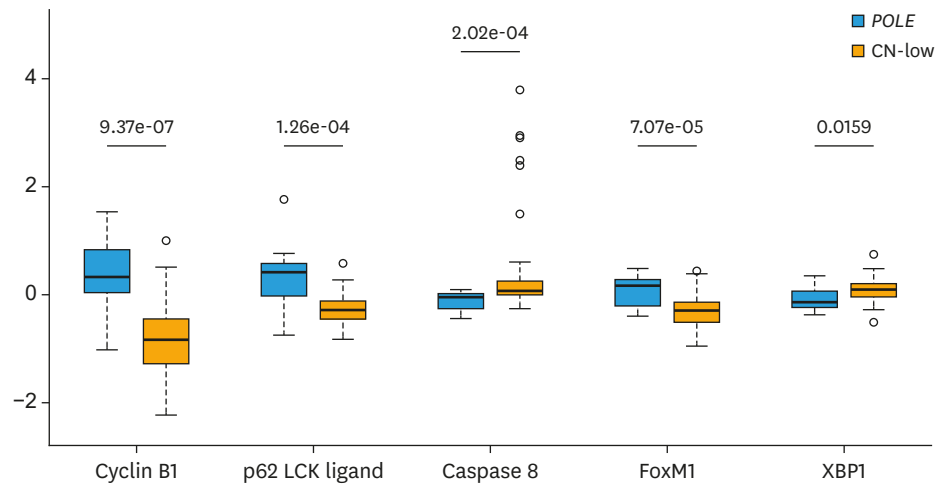
Fig. 1. Expression levels of 5 selected protein markers, according to subtype.
CN, copy number; *POLE*, polymerase epsilon exonuclease; XBP1, X-box binding protein 1.

In the final model, BMI, cyclin B1, caspase 8, and XBP1 were selected; the fitted results are summarized in **Table S2**. With the *POLE* subtype used as a case, the mean values of validation AUC, sensitivity, specificity, and balanced accuracy were 0.97, 0.91, 0.87, and 0.89, respectively.

$$\text{logit}(p(POLE)) \sim BMI + Cyclin\ B1 + Caspase\ 8 + XBP1$$

3. Validation

Cyclin B1 was expressed in 127 of 155 cases (81.9%) (**Fig. 2**); caspase-8 cytoplasmic and nuclear expression was observed in 98 cases (63.2%) and 92 cases (59.3%), respectively. Cyclin B1 IHC expression in the *POLE* subtype was significantly higher than that in the CN-
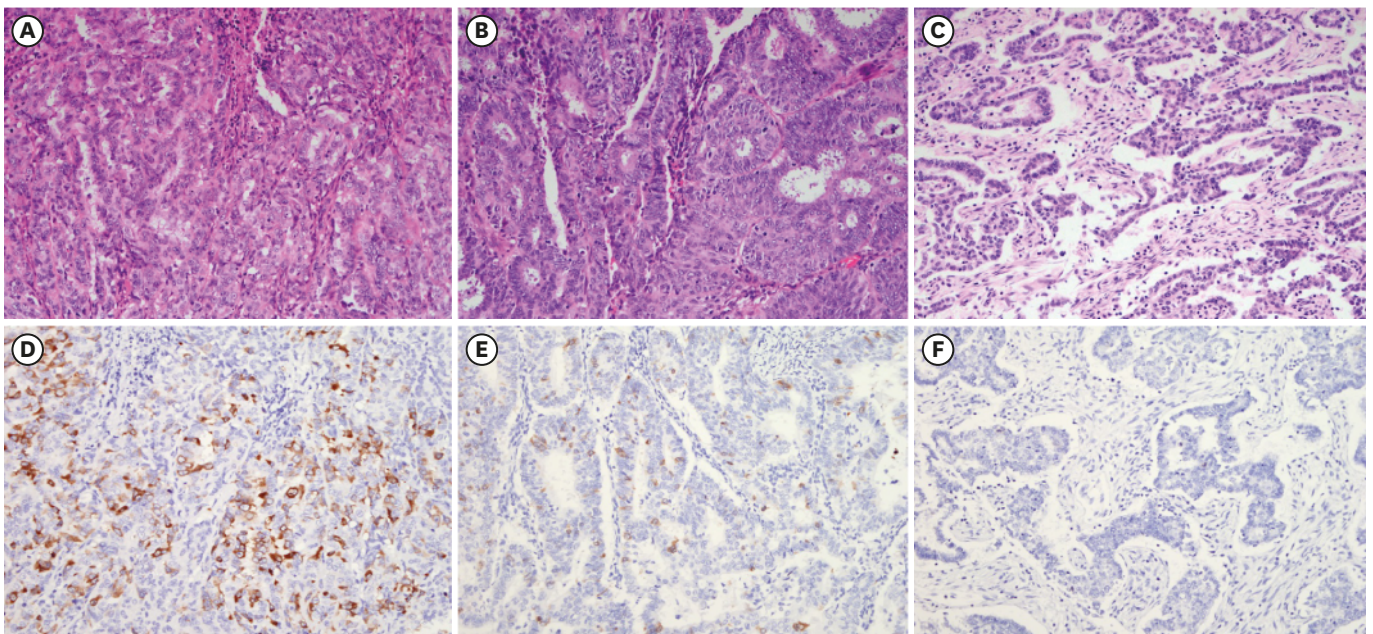

Fig. 2. Representative images of histology (A-C) and immunohistochemistry for cyclin B1 (D-F). Cyclin B1 showed cytoplasmic expression with variable intensity and proportion. (D) Score 6, (E) score 2, and (F) score 0 (×200 magnification).
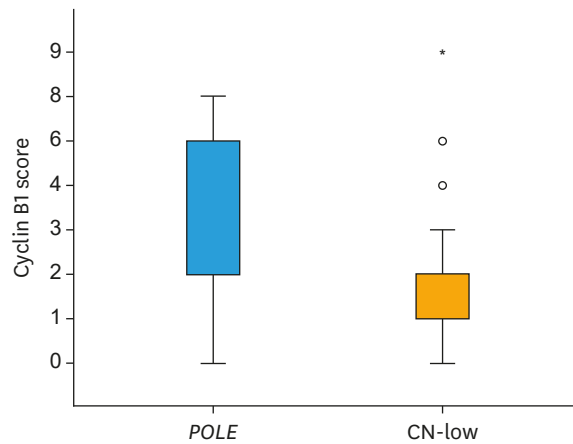
Fig. 3. Cyclin B1 expression in the *POLE* and CN-low subtypes.
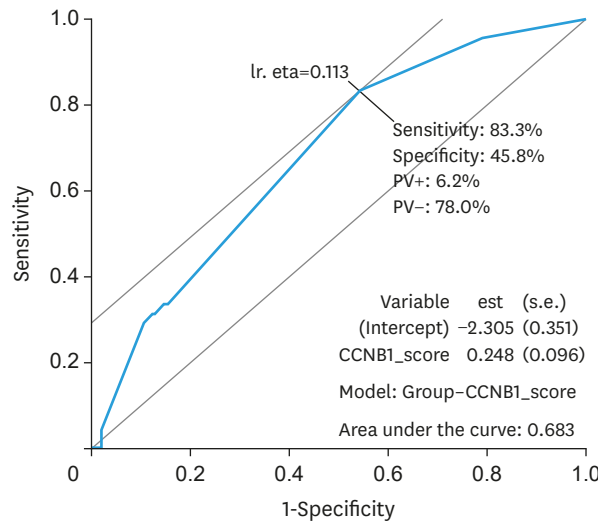CN, copy number; *POLE*, polymerase epsilon exonuclease.

Fig. 4. Receiver operating characteristic curve of cyclin B1 immunohistochemistry expression.

low subtype (**Fig. 3**) and receiver operating characteristic curve of cyclin B1 IHC expression showed AUC of 0.683 (**Fig. 4**).

There was no difference in cytoplasmic caspase-8 expression between the 2 groups; however, nuclear caspase-8 expression (score >2) was slightly lower in the *POLE* group than in the CN-low group; however, the difference was not statistically significant (data not shown).

DISCUSSION

The molecular classification of EC can be a powerful tool for tailored treatment; however, the widespread use of this system has been hindered by the need to perform costly genomic sequencing of samples. Our data are the first to suggest a combination of clinicopathologic and protein markers that potentially replaces *POLE* sequencing for the molecular classification of this EC subtype. The final model makes use of 4 markers (BMI, cyclin B1,

caspase 8, and XBP1) that exhibited excellent results in CV (AUC, 0.97). Specifically, cyclin B1 IHC expression was significantly higher in the *POLE* than in the CN-low subtype; therefore, cyclin B1 IHC may substitute *POLE* sequencing. Identification of all 4 molecular subtypes without gene sequencing could allow for the molecular classification of EC tumors to be routinely implemented in the treatment of EC.

The molecular mechanisms by which the selected protein markers are associated with EC subtype remain unclear. Caspase 8 is a protease involved in apoptosis and is mutated in 10% of ECs [12]. Both caspase 8 and *POLE* mutations are associated with immune signatures and better survival in EC [12]. In a study using large-scale genomic data sets of tumor biopsies, both caspase 8 mutations and programmed death-ligand 1/2 (*PD-L1/2*) amplification were associated with high cytolytic activity of the local immune infiltrate [13]. Furthermore, *PD-L1* expression was reported to be upregulated in the *POLE* subtype [14]. These findings suggested that *POLE*, caspase 8, and *PD-L1* share an immune signature. Cyclin B1 is a regulatory protein involved in mitosis and is associated with a less differentiated phenotype [15] and tumorigenesis in EC [16]. XBP1 is a transcription factor and is known to enable cancer cells to survive under hostile conditions, such as hypoxia and starvation [17].

IHC is a useful technique for the analysis of tumors in daily practice. However, it is subject to staining and observer variation [18] and is not the optimum method to use for marker screening. We, therefore, used existing RPPA data to conduct the analyses for this study. Previous studies have examined the association between IHC intensity and expression levels of proteins assessed by RPPAs. In a study including 37 diffuse large B-cell lymphoma tissues, the normalized expression levels of CD5, CD10, BCL6, MUM1, BCL2, Ki-67, and C-MYC identified by RPPAs correctly discriminated IHC-positive from -negative cases (receiver operating characteristic AUC ranges 0.563–0.874) [19]. Similarly, a study of 69 colon cancer tissues showed that the normalized expression level of vascular endothelial growth factor using RPPAs was correlated with the IHC score (staining intensity multiplied by the percentage of tumor cells) (r=0.283, p=0.018) [20].

The protein expression data of TCGA EC cohort contain information on the expression levels of 208 genes. Our findings are, therefore, limited by the markers included in TCGA dataset and, hypothetically, may have excluded markers that could be better at distinguishing between the *POLE* and CN-low subtypes. While it is unclear how the markers were selected for inclusion in TCGA study, it was suggested that these markers represent the major cellular signaling pathways [21]. Therefore, we believe that markers outside TCGA dataset are less likely to be good markers.

This study has several limitations. First, the small sample size for each subtype renders this study vulnerable to biases. However, to our knowledge, TCGA contains the largest database of information, including protein expression and subtype data, on EC. Second, because only normalized RPPA data were available, we were not able to separately normalize a training and validation protein expression set of data during CV. Therefore, there is a risk of data leakage. Third, the antibody clone and cut-off used in this study are not standardized methods for detecting CyclinB1 expression by IHC. Fourth, the selected markers based on RPPA showed poor performance in the validation set. Specifically, cyclin B1 showed only a modest AUC (0.683) and caspase 8 failed to show differential expression in IHC in validation set. IHC for XBP1 failed due to poor staining quality. The poor performance could be due to the different methods of protein expression measurement between the train and validation sets (RPPA

in the train set but IHC in the validation set). Therefore, model development and validation using protein expression level in IHC should be attempted in future studies.

In conclusion, BMI and the expression levels of cyclin B1, caspase 8, and XBP1 are candidate markers distinguishing between the *POLE* and CN-low subtypes of EC. Among these, cyclin B1 IHC expression was significantly higher in the *POLE* than in the CN-low subtype; therefore, cyclin B1 IHC may substitute *POLE* sequencing. A further validation study using immunohistochemical staining of these markers is necessary to facilitate the adoption of a molecular classification system in daily practice.

SUPPLEMENTARY MATERIALS

Table S1
Univariable analysis results for top 5 protein markers

Click here to view

Table S2
Fitted results of the multiple logistic regression model for the final model

Click here to view

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
   PUBMED | CROSSREF

2. Zhang S, Gong TT, Liu FH, Jiang YT, Sun H, Ma XX, et al. Global, regional, and national burden of endometrial cancer, 1990-2017: results from the Global Burden of Disease study, 2017. Front Oncol 2019;9:1440.
   PUBMED | CROSSREF

3. Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. Nature 2013;497:67-73.
   PUBMED | CROSSREF

4. Talhouk A, McConechy MK, Leung S, Yang W, Lum A, Senz J, et al. Confirmation of ProMisE: a simple, genomics-based clinical classifier for endometrial cancer. Cancer 2017;123:802-13.
   PUBMED | CROSSREF

5. Britton H, Huang L, Lum A, Leung S, Shum K, Kale M, et al. Molecular classification defines outcomes and opportunities in young women with endometrial carcinoma. Gynecol Oncol 2019;153:487-95.
   PUBMED | CROSSREF

6. Bosse T, Nout RA, McAlpine JN, McConechy MK, Britton H, Hussein YR, et al. Molecular classification of grade 3 endometrioid endometrial cancers identifies distinct prognostic subgroups. Am J Surg Pathol 2018;42:561-8.
   PUBMED | CROSSREF

7. Akhtar M, Al Hyassat S, Elaiwy O, Rashid S, Al-Nabet AD. Classification of endometrial carcinoma: new perspectives beyond morphology. Adv Anat Pathol 2019;26:421-7.
   PUBMED | CROSSREF

8. Talhouk A, McAlpine JN. New classification of endometrial cancers: the development and potential applications of genomic-based classification in research and clinical care. Gynecol Oncol Res Pract 2016;3:14.
   PUBMED | CROSSREF

9. MD Anderson UTHealth Graduate School of Biomedical Sciences. The Cancer Proteome Atlas [Internet]. Houston, TX: MD Anderson UTHealth Graduate School of Biomedical Sciences; 2016 [cited 2020 Feb 13]. Available from: https://bioinformatics.mdanderson.org/public-software/tcpa/.

10. Royston P, White IR. Multiple imputation by chained equations (MICE): implementation in Stata. J Stat Softw 2011;45:1-20.
CROSSREF

11. Zhuang L, Lee CS, Scolyer RA, McCarthy SW, Zhang XD, Thompson JF, et al. Mcl-1, Bcl-XL and Stat3 expression are associated with progression of melanoma whereas Bcl-2, AP-2 and MITF levels decrease during progression of melanoma. Mod Pathol 2007;20:416-26.
PUBMED | CROSSREF

12. Ghanekar Y, Sadasivam S. In silico analysis reveals a shared immune signature in CASP8-mutated carcinomas with varying correlations to prognosis. PeerJ 2019;7:e6402.
PUBMED | CROSSREF

13. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 2015;160:48-61.
PUBMED | CROSSREF

14. van Gool IC, Eggink FA, Freeman-Mills L, Stelloo E, Marchi E, de Bruyn M, et al. POLE proofreading mutations elicit an antitumor immune response in endometrial cancer. Clin Cancer Res 2015;21:3347-55.
PUBMED | CROSSREF

15. Milde-Langosch K, Bamberger AM, Goemann C, Rössing E, Rieck G, Kelp B, et al. Expression of cell-cycle regulatory proteins in endometrial carcinomas: correlations with hormone receptor status and clinicopathologic parameters. J Cancer Res Clin Oncol 2001;127:537-44.
PUBMED | CROSSREF

16. Lv S, Xu X, Wu Z. Identification of key candidate genes and pathways in endometrial cancer: evidence from bioinformatics analysis. Oncol Lett 2019;18:6679-89.
PUBMED | CROSSREF

17. Cubillos-Ruiz JR, Bettigole SE, Glimcher LH. Molecular pathways: immunosuppressive roles of IRE1α-XBP1 signaling in dendritic cells of the tumor microenvironment. Clin Cancer Res 2016;22:2121-6.
PUBMED | CROSSREF

18. Kirkegaard T, Edwards J, Tovey S, McGlynn LM, Krishna SN, Mukherjee R, et al. Observer variation in immunohistochemical analysis of protein expression, time for a change? Histopathology 2006;48:787-94.
PUBMED | CROSSREF

19. Suzuki M, Muroi A, Nojima M, Numata A, Takasaki H, Sakai R, et al. Utility of a reverse phase protein array to evaluate multiple biomarkers in diffuse large B-cell lymphoma. Proteomics Clin Appl 2020;14:e1900091.
PUBMED | CROSSREF

20. Chung JY, Braunschweig T, Hong SM, Kwon DS, Eo SH, Cho H, et al. Assessment of vascular endothelial growth factor in formalin fixed, paraffin embedded colon cancer specimens by means of a well-based reverse phase protein array. Proteome Sci 2014;12:27.
PUBMED | CROSSREF

21. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods 2013;10:1046-7.
PUBMED | CROSSREF