Integrative Protein-Based Prognostic Model for Early Stage Endometrioid Endometrial Cancer

Ji-Yeon Yang 1,2, Henrica M.J. Werner 3,4, Jie Li 5, Shannon N. Westin 6, Yiling Lu 5, Mari K. Halle 3,4, Jone Trovik 3,4, Helga B. Salvesen 3,4, Gordon B. Mills 5 and Han Liang 1,5

1Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
2Department of Applied Mathematics, Kumoh National Institute of Technology, Korea
3Centre for Cancer Biomarkers, Department of Clinical Science, The University of Bergen, Bergen, Norway
4Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway
5Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
6Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Corresponding Authors: Han Liang, Ph.D., Department of Bioinformatics and Computational Biology, Unit 1410, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX 77030-3722, USA; hliang1@mdanderson.org
Gordon B. Mills, MD, PhD, Department of Systems Biology, Unit 950, 1515 Holcombe Blvd, Houston, TX 77030-3722, USA; gmills@mdanderson.org

Running title: Prognostic model for early-stage endometrioid carcinoma

Keywords: prognostic modeling, functional proteomics, endometrial cancer, model construction, biomarker

Category: Personalized Medicine

Financial support: This study was supported by the National Institutes of Health through grant number CA143883 to G.B.M.; CA088084 K12 Calabresi Scholar Award to S.N.W; CA175486 to H.L.; CA098258 SPORE in Uterine Cancer to G.B.M., S.N.W., H.L. and CCSG grant CA016672 to MDACC. Additional support was received from Kumoh National Institute of Technology to J.Y.Y.; Helse Vest, Research Council of Norway, The Norwegian Cancer Society (Harald Andersens legat) and University of Bergen to H.B.S.; the Lorraine Dell Program in Bioinformatics for Personalization of Cancer Medicine, one Lee Clark Fellow Award from the Jeanne F. Shelby Scholarship Fund and a grant from the Cancer Prevention and Research Institute of Texas (RP140462) to H.L.

Word count: 5833 words; total number of figures and tables: 6
Abstract

Purpose

Endometrioid endometrial carcinoma (EEC) is the major histological type of endometrial cancer, the most prevalent gynecologic malignancy in USA. EEC recurrence or metastasis is associated with a poor prognosis. Early-stage EEC is generally curable, but a subset has high risk of recurrence or metastasis. Prognosis estimation for early-stage EEC mainly relies on clinicopathological characteristics, but is unreliable. We aimed to identify patients with high-risk early-stage EEC who are most likely to benefit from more extensive surgery and adjuvant therapy by building a prognostic model that integrates clinical variables and protein markers.

Experimental Design

We employed two large, independent early-stage EEC datasets as training ($n = 183$) and validation cohorts ($n = 333$), and generated the levels of 186 proteins and phosphoproteins using reverse-phase protein arrays. By applying an initial filtering and the elastic net to the training samples, we developed a prognostic model for overall survival containing two clinical variables and 18 protein markers and optimized the risk group classification.

Results

Kaplan-Meier survival analyses in the validation cohort confirmed an improved discriminating power of our prognostic model for patients with early-stage EEC over key clinical variables (log-rank test $p$-value = 0.565 for disease stage, 0.567 for tumor grade, $1.3 \times 10^{-4}$ for the integrative model). Compared with clinical variables (stage, grade, and patient age), only the risk groups defined by the integrative model were consistently significant in both univariate and multivariate analyses across both cohorts.

Conclusions

Our prognostic model is potentially of high clinical value for stratifying patients with early-stage EEC and improving their treatment strategies.
Statement of Translational Relevance

A subset of endometrioid endometrial carcinoma (EEC) is associated with recurrence or metastasis, leading to a poor prognosis. Currently, prognosis estimation for early-stage EEC mainly relies on clinicopathological characteristics, but is unreliable. Here we have developed and validated an effective post-operative prognostic model for the overall survival of patients with early-stage EEC, which integrates clinical variables and protein markers. To the best of our knowledge, this is the first quantitative protein-based model that represents a novel, improved prognostic approach for this patient population. The risk score calculated from this model will help identify patients with high-risk early-stage EEC who are most likely to benefit from more extensive surgery and adjuvant therapy.
Introduction

Endometrial cancer is the most prevalent gynecologic malignancy in the United States, with an estimated 54,870 new cases and 10,170 deaths occurring in 2015 (1). In contrast to progress observed for most cancer types, the incidence and mortality from endometrial cancer has increased over the past several decades (2). The most common histological type of endometrial cancer (~80%), endometrioid endometrial carcinoma (EEC), is linked to estrogen excess, obesity, and hormone-receptor positivity. Non-endometrioid tumors include serous carcinoma (~10%) and several other rare types (e.g., clear cell carcinomas and carcinosarcomas) (3), which are typically associated with advanced stage at diagnosis and poor prognosis. These tumors are usually treated with more aggressive therapy, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, and post-surgical radiation therapy and chemotherapy (4, 5).

The reported 5-year survival rate is relatively good in early-stage EEC (stage I, II, 70-90%) but it falls below 50% in late-stage EEC (40-50% for stage III, 15-20% for stage IV) (6). Indeed, patients with late-stage EEC or recurrent disease have a poor prognosis, with a median overall survival of less than 12 months. Their treatment options are similar to those offered to patients with non-endometrioid carcinoma (7). Fortunately, most women with EEC are diagnosed at an early stage of the disease, when the cancer can generally be cured by surgery. However, not all early-stage EEC is clinically indolent: a subset has poor prognosis due to metastasis or disease recurrence following surgery (8). This clinical heterogeneity for early-stage EEC warrants efforts to identify high-quality prognostic markers that can be used to stratify patients and to determine their prognoses, thereby implementing evidence-based individualized therapeutic approaches. Currently, the prediction of prognosis for patients with early-stage EEC is mainly based on disease stage and tumor grade. Other prognostic factors include
lymphovascular space invasion, obesity, age, myometrial invasion depth and lymph node status (9, 10), but none of these factors can reliably predict patient survival. Therefore, it is critical to develop an effective strategy to identify patients with high-risk, early-stage EEC since they are likely to benefit from more extensive surgery and aggressive adjuvant therapies. Furthermore, this may spare patients who have low-risk EEC from potentially toxic interventions.

The detailed characterization of endometrial tumors at the molecular level may provide information that can be used to increase the power of prognostic models beyond that offered by clinicopathological characteristics alone, thereby improving surveillance and treatment strategies. In this study, we aimed to develop an effective prognostic model for patients with early-stage EEC by integrating clinical characteristics with proteomic data. As proteins are major functional units in various biological processes, protein markers have the potential to reflect key processes underlying tumor development and progression that may not be captured by other types of molecular markers (11, 12). To identify and validate potential biomarkers, we used functional proteome-based reverse-phase protein array (RPPA), which is an antibody-based technology that allows for high-throughput measurements of the expression levels of proteins and phosphoproteins in a large number of samples in a cost-effective and sensitive manner (13, 14). The technical reliability and the utility of this platform have been well documented in the literature, including in our previous studies (15-17). A typical coefficient of variation for RPPA is 5-7%, demonstrating high reproducibility (15, 18). In our recent pan-cancer study, the correlation analysis among RPPA protein markers successfully captured known signaling pathways in different tumor contexts (11). Importantly, through a systematic assessment on prognostic utility, we found that RPPA-based protein expression is the most informative data among various types of molecular data surveyed (i.e., somatic copy-number alterations, DNA
methylation, mRNA, microRNA and protein expression)(19). Therefore, we focused on the RPPA-based protein markers in this study.

Here we employed two large, independent patient sets as the training and validation cohorts, respectively. Based on their survival data availability/quality and also because we are more interested in patients’ long-term outcome, we considered overall survival as the endpoint in our analysis. We first built a prognostic model for patients with early-stage EEC from the training cohort, and determined a risk group classification scheme. Our prognostic model that integrates both protein markers and clinical characteristics showed better performance than conventional clinical variables in both the training and validation cohorts. Thus, the integrative prognostic model we have developed may represent a valuable tool for improving the surveillance and treatment strategies of patients with early-stage EEC.

**Materials and Methods**

*Patient sample collection*

In this study, fresh-frozen samples were collected from patients newly diagnosed with EEC at Haukeland University Hospital (Bergen, Norway), the University of Texas MD Anderson Cancer Center (MDACC, Houston, Texas, USA) and The Cancer Genome Atlas (TCGA). Bergen samples were approved by the Norwegian Data Inspectorate (961478-2), the Norwegian Social Science Data Services (15501) and the local Institutional Review Board (REKIII nr. 052.01). The MDACC samples were approved by the Institutional Review Board of MDACC (Lab08-0580). Treatment plans followed the established standards at the respective institutions in accordance with NCCN guidelines and as previously reported (20). In general, early-stage EEC was treated with surgery alone with or without adjuvant radiotherapy; late-stage EEC was treated
with surgery followed by chemotherapy with or without volume-directed radiotherapy. Tumor content, histological classification, grade, and disease stage were reviewed by independent pathologists. All the patients provided written informed consent for the collection of samples and subsequent analyses.

For prognostic modeling, we used 209 EEC samples from the Bergen set as the training cohort because of the large sample size and long follow-up time associated with this data set. We combined the MDACC and TCGA samples as the validation cohort (n = 427) to increase the total sample size and statistical power. The clinical information for TCGA samples was obtained from the TCGA consortium paper for endometrial cancer (21). We had limited clinical data, especially for MDACC and Bergen samples. Among potential affecting factors, patient age, grade, disease stage, BMI and ethnicity were available for MDACC samples, and patient age, grade and disease stage for Bergen samples. Therefore, we used three clinical factors (patient age, disease stage and tumor grade) that were available for all patients in the three sample sets. We capped the overall survival time at 60 months, as described previously (22). The individual clinical data of the patients included in this study are reported in Supplemental Tables 1 and 2.

Reverse-phase protein array profiling and data normalization

We performed RPPA profiling for all samples at the MD Anderson RPPA core facility, as previously described (21). Briefly, proteins were extracted from tumor tissues, denatured by SDS, and printed on nitrocellulose-coated slides, which was followed by antibody probing. Quantification of protein expressions was performed through “supercurve fitting,” as described previously (23, 24), and which is a method that has been extensively validated for both cell line and patient samples (15, 17). We obtained quantitative protein expression profiles of 186
proteins and phosphoproteins for both the training and validation samples (Supplemental Table 3 for the list of 186 antibodies). These proteins were selected to broadly represent the important pathways in cancer, and their RPPA data have been widely used in various TCGA-related analyses, providing deep insights into the molecular mechanisms of various cancer types. All the antibodies were validated by Western blot analysis. The validation process and assessment results on the antibodies for RPPA were detailed previously (15).

All the samples used in this study were run in three different RPPA batches. Since there is a potential problem of batch effects when combining the batches of protein expression data, we used Replicate-Based Normalization (RBN) method as described in our recent study (11), which utilizes replicate samples run across multiple batches. In one “anchor” batch, we ran many replicate samples that were common with the other two batches (200 and 30, respectively). Those samples were all TCGA validation samples and non-control samples. RBN was performed by adjusting each data point in the non-anchor batches so that the mean and variance of the common samples for each antibody are identical to those in the anchor batch. The normalized RPPA data for the 209 training samples and 427 validation samples are provided in Supplemental Tables 1 and 2.

Development of the prognostic model for early-stage EEC

We used data from 183 patients with stage I or II EEC from the Bergen set as the training cohort. As a preprocessing step to reduce feature dimensionality, we filtered out noisy features by applying univariate Cox regressions to protein expression and clinical features (186 protein markers and 3 clinical factors). To avoid ruling out potentially relevant features, we used a cutoff
of 0.15 for the \( p \)-values, and retained only features with a \( p \)-value smaller than 0.15 for model development.

Then, we employed the elastic net (25) to identify markers associated with overall survival and to use in training the final model for prediction with the selected features. The elastic net simultaneously conducts automatic variable selection and group selection of the correlated variables. The explicit object function and the algorithm for estimating the solution of the elastic net has been described previously (25, 26). We used the implementation in the R package “glmnet” (27). We used leave-one-out cross-validation to select the tuning parameter, and determined the elastic net mixing parameter in order to find a parsimonious model while maintaining a modest discriminating accuracy based on the concordance index (28).

The final model is a linear combination of features selected by the elastic net, weighted by the corresponding elastic net coefficients. The weights are a rough estimate of the contribution of the information content of each marker to the overall risk score. Specifically,

\[
\text{Risk Score} = 0.048 \times \text{Age} + 0.118 \times \text{Grade} - 1.032 \times \text{EGFR} - 0.659 \times \text{phospho Myosin IIa} \\
- 0.496 \times \text{AR} - 0.462 \times \text{cMyc} - 0.273 \times \text{phospho STAT3} - 0.131 \times \text{phospho p38 MAPK} \\
- 0.055 \times \text{phospho mTOR} - 0.018 \times \text{Hsp70} + 0.071 \times \text{FASN} \\
+ 0.179 \times \text{Acetyl-\(\alpha\)-Tubulin (Lys40)} + 0.210 \times \text{ACC1} + 0.249 \times \text{Ku80} \\
+ 0.418 \times \text{phospho NF-\(\kappa\)B p65} + 0.420 \times \text{Bim} + 0.504 \times \text{Chk2} + 0.573 \times \text{cKIT} \\
+ 1.527 \times \text{SMAD4}.
\]  

(Equation 1)

Because the model includes both protein markers and clinical variables from samples of early-stage EEC, we refer to the model as the integrative prognostic model for early-stage EEC. The exponential of an elastic net coefficient gives the hazard ratio (HR) of death associated with each marker. However, there is no consensus on a statistically valid method of estimating the
standard error of a coefficient estimate with shrinkage methods including the elastic net, so the standard errors or confidence intervals for elastic net estimates were not reported here.

To facilitate clinical application, we divided the training samples into two risk groups according to each patient's risk score based on Equation 1. We determined the cutoff value for the risk scores to ensure that the two risk groups would have similar numbers of events, as described previously (29). Statistically, it is known that analysis of groups with unbalanced numbers of events may introduce a bias in the parameter estimation and the bias will be substantial when the number of events is small, that is, in case of rare events (30, 31). In EEC (especially in early-stage EEC), mortality is not so high and thus the number of events is relatively small. So we employed a simple but widely used approach of splitting samples into risk groups based on equal/similar number of events, in order to reduce the estimation bias and to ensure the similar standard errors of the parameter estimates across risk groups.

Validation of the integrative prognostic model for early-stage EEC

We used 333 samples of stages I and II EEC from the validation cohort to validate the prognostic model (Equation 1). We computed risk scores based on Equation 1 for the validation samples, and then classified them into low- or high-risk groups, with the cutoff determined in the training set as described above. We used univariate and multivariate Cox regressions to evaluate the patient risk classification.

Development and validation of the prognostic model for late-stage EEC

In a similar way, we constructed and validated a prognostic model for patients with late-stage EEC. We trained the model using samples of stages III and IV EEC from the Bergen set (n = 26)
by initially filtering the data through a univariate Cox regression, with a $p$-value of 0.15 as the
cutoff, and selecting the features through the elastic net, with the tuning parameter selected by
leave-one-out cross-validation. The final model is as follows:

Risk Score = 0.019$\times$Age + 0.203$\times$Grade - 1.511$\times$SMAD4 - 0.479$\times$phospho Shc
- 0.421$\times$cKIT - 0.270$\times$phospho ERK/MAPK - 0.207$\times$PKC$\alpha$ - 0.180$\times$Bim
- 0.099$\times$Bcl2 + 0.113$\times$TFRC + 0.179$\times$ASNS + 0.336$\times$VEGFR-2
+ 0.364$\times$PAI1 + 0.392$\times$phospho HER2 + 0.446$\times$phospho 4E-BP1
+ 1.311$\times$EGFR                                                                       (Equation 2)

This model also contains both protein markers and clinical features, so we refer to it as
the integrative prognostic model for late-stage EEC. We determined the risk score cutoff for the
classification of the training samples such that the two risk groups would have similar numbers
of events.

We computed risk scores for the samples of late-stage EEC (stages III and IV, $n = 94$) in
the validation cohort, and divided the samples into the two risk groups according to the same
cutoff used for the training samples.

Comparison with clinical-variable-only models

For the practical utility of the integrative models, we considered the patient’s classification as
being in the low- or high-risk group instead of risk scores. To evaluate the performance of our
integrative models relative to prognoses based on only clinical variables, we considered
univariate and multivariate Cox proportional hazards models, with the following characteristics
as covariates:

[1] Disease stage
2 Tumor grade
3 Patient age
4 Risk group index based on the integrative model
5 Stage + Grade + Age + Risk group index based on the integrative model

For the disease stage, we compared two levels: stage I vs. stage II in the early-stage EEC model and stage III vs. stage IV in the late-stage EEC model. In the Cox regressions, we treated the tumor grade as a continuous variable, with natural ordering so as to compute only one regression coefficient for this variable and to examine the overall difference across tumor grades. We dichotomized the patient’s age as two categories, younger than 60 years of age or 60 years of age and older, in order to evaluate the efficacy of a clinical practice guideline: patients 60 years of age and older are generally advised to receive adjuvant therapy (32). Furthermore, based on a previous study (10), we considered three categories of patient age (<50, [50, 70) and ≥70), but found no events in the group of <50 years. Instead, we considered an alternative, similar three-age classification (<60, [60, 80) and ≥80]. In addition, we computed concordance indexes (C-indexes) to compare the discriminatory power of clinical factors and the integrative models.

We evaluated the additional prognostic value of the integrative models over the clinical factors using a multivariate regression analysis with Model 5: Stage + Grade + Age + Risk group index based on the integrative model. Although our integrative models include the two clinical factors of patient age and tumor grade, multicollinearity is not an issue in Model 5 because we used the risk group index rather than the risk score as a covariate. In addition, we used log-rank tests to examine the differences in survival between the risk groups as stratified by the integrative models, between disease stages and among tumor grades. In order to assess the
robustness of our models to different cutoff values, we tried different cutoffs that made the sample numbers in risk groups equal to those numbers with different stages (or different grades). We also considered another different cutoff value, the 25th percentile of risk scores (which approximately corresponded to the value of 0.5 in early-stage samples and 2.5 in late-stage samples).

**Results**

*Patient characteristics and the prognostic power of disease stage and tumor grade*

For robust prognostic modeling, we employed two independent data sets of EEC as training and validation cohorts. The patients’ characteristics are summarized in Table 1. Our training cohort was obtained from Haukeland University Hospital, Bergen, Norway (Supplemental Table 1), and contained 183 samples of early-stage EEC (FIGO 2009 stages I and II) and 26 samples of late-stage EEC (stages III and IV). The validation samples were obtained from MDACC (Houston, Texas, USA) and TCGA (Supplemental Table 2). As there was no significant difference in survival between the MDACC and TCGA data sets for either early-stage or late-stage EEC (log-rank test *p*-value = 0.36 for early-stage EEC and 0.85 for late-stage EEC), we combined them as one validation cohort to increase the sample size and boost the statistical power. In total, the validation cohort contained samples from 333 patients with early-stage EEC and 94 patients with late-stage EEC. We generated the expression profiles of 186 proteins and phosphoproteins using RPPA. The RPPA data for the training and validation samples are presented in Supplemental Tables 1 and 2, and information about the 186 antibodies is provided in Supplemental Table 3. The proteomic profiling and quality control followed the well-established procedures in the TCGA project (12, 21). To remove batch effects in RPPA data, we used replicate-based normalization, as previously described (11).
Patients represented in the training cohort were slightly older than those in the validation cohort (early-stage EEC: mean age in training 64.9 years vs. mean age in validation 61.4 years, t-test, \( p \)-value \( 1.4 \times 10^{-3} \); late-stage EEC: mean age in training 67.9 years vs. mean age in validation 62.1 years, \( p \)-value = 0.029, Table 1). No significant difference in the stage distributions was found between the training and validation cohorts, using chi-squared tests (\( p \)-value: 0.22 for early-stage EEC and 0.69 for late-stage EEC). We found a significant difference in the grade distributions between the training and validation cohorts for the patients with early-stage EEC (\( p \)-value = \( 5.3 \times 10^{-4} \)): samples of early-stage EEC in the validation cohort contained a larger proportion of high-grade tumors, which may reflect the strategy to enrich for large tumors in the TCGA compared to the population-based approach for the Bergen cohort (33). There was no statistically significant difference in the overall survival time between the training and validation cohorts for the patients with early-stage EEC (log-rank test, \( p \)-value = 0.48); however, the patients with late-stage EEC who were represented in the training cohort showed a significantly worse survival time than those in the validation cohort (log-rank test, \( p \)-value = 0.014, Supplemental Figure 1).

We evaluated the discriminating power of disease stage and tumor grade for the early-stage and late-stage EEC samples. For the early-stage EEC samples, there was no difference in overall survival when the patients were stratified by disease stage or tumor grade: log-rank test \( p \)-values were 0.40 and 0.57 for the respective training and validation cohorts when split by disease stage (Figure 1A and B; the results also held true when sub-stages were considered); and the log-rank test \( p \)-values were 0.28 and 0.57 for the respective training and validation cohorts when split by tumor grade (Figure 1C and D).
For the late-stage EEC samples, there was a significant survival difference in the training cohort when split by disease stage (log-rank test, \( p \)-value = 0.02) and in the validation cohort when split by tumor grade (log-rank test, \( p \)-value = 0.02). Supplemental Figure 2 shows Kaplan-Meier curves for late-stage samples when patients were stratified according to disease stage or tumor grade. Thus, no significant discriminating power of stage or grade was observed for early-stage EEC, which highlights a pressing need for effective prognostic models for this patient population. Based on the markedly different marker patterns and clinical outcomes, we performed prognostic modeling on early-stage EEC samples and late-stage EEC samples separately.

An effective prognostic model for patients with early-stage EEC

The flow chart in Figure 2 shows the overall procedure of developing and validating an integrative prognostic model for patients with early-stage EEC. We developed a prognostic model after applying an initial filtering and the elastic net to the early-stage samples in the training cohort (see Materials and Methods). The integrative model (Equation 1 in Materials and Methods) contains two clinical factors (patient age and tumor grade) and 18 protein markers (listed in Supplemental Table 4). Based on this model, we computed risk scores for the training samples as a weighted sum of the selected features. As shown in Supplemental Figure 3A, the distribution of risk scores was unimodal with very similar peaks between early-stage training and validation samples. To facilitate clinical application, we chose a risk score cutoff to classify patients with early-stage EEC in the training cohort into low- and high-risk groups. The cutoff was determined to ensure that the two risk groups would have similar numbers of events; and this scheme resulted in a cutoff point of approximately the 90th percentile of risk scores in the
training cohort, as illustrated in the second column of Supplemental Figure 3A. As expected, survival was much worse among patients in the high-risk group compared to those in the low-risk group (log-rank test, \( p \)-value = \(2.3 \times 10^{-14}\), Figure 3A). From the clinical point of view, this risk score cutoff point is of potential utility because it helps identify a subset of patients with early-stage EEC (10%) who are most likely to benefit from additional treatment despite added toxicity.

To evaluate the discriminating power of our integrative model, we locked the model and applied it to samples of early-stage EEC in the validation cohort. The training and validation samples showed similar ranges and distributions of risk scores. Using the same cutoff value, we split the validation samples into two risk groups with similar numbers of events as described above. Kaplan-Meier survival analysis showed a significant survival difference between the two risk groups (log-rank test \( p \)-value = \(1.3 \times 10^{-4}\), Figure 3B).

We used a data-driven cutoff (i.e., the 90th percentile of risk scores of the training samples) as described previously (29, 34); the determination of this cutoff point is a part of our modeling process. To assess the robustness of our model to different cutoff values, we considered cutoffs that gave equal sample sizes based on the same disease stage (or tumor grade). In comparison to the power of the disease stage (stage I vs. II), the 183 early-stage EEC training samples were split into two risk groups, with \( n = 170 \) and \( n = 13 \), which were the numbers of training samples in stage I and II, respectively. Similarly, the 297 validation samples with the lowest risk scores were classified as low risk and the remaining 36 validation samples were classified as high risk. We observed a significant survival difference between these two risk groups (log-rank test, \( p \)-value = \(1.0 \times 10^{-14}\) for the training samples; and \( p \)-value = \(2.5 \times 10^{-3}\) for the validation samples, Figure 3C and D). When using cutoffs that gave the same sample sizes based
on the tumor grades (grades 1, 2, and 3), we also observed a statistically (or marginally) significant difference in overall survival among the risk groups (log-rank test, $p$-value $= 4.0 \times 10^{-10}$ for the training samples; and $p$-value $= 0.066$ for the validation samples, Figure 3E and F). Furthermore, using the 25th percentile of risk scores, which was approximately equal to 2.5, there was statistically significant (or marginal significance) discrimination of patients with regard to survival in early stage training and validation samples, as shown in Supplemental 4A and B (log-rank test $p$-values are 0.01 and 0.076). These separations, however, were not as distinct as those for which the cutoff was driven by the data or the cutoff ensured equal sample sizes for the different stages. In addition, we considered patient stratification according to the median split, and observed significantly improved survival in the low-risk group for both training and validation samples. Taken together, these results indicate that patient stratification based on our integrative model is effective and robust across multiple cutoff points.

We further investigated the additional discriminating power of this integrative model over the clinical factors using univariate and multivariate Cox proportional hazards models. We considered three clinical factors (disease stage, tumor grade and patient age) that were available for both the training and validation cohorts in the analysis. Age was dichotomized at a threshold of 60 years because adjuvant therapy is generally recommended for patients who are 60 years of age or older (32). Age was significant in the univariate model for the validation samples (Wald’s test $p$-value $= 0.044$), but was not significant in the univariate model for the training samples, nor in the multivariate models. In line with a previous study (10), we also examined the effect of a three-age-group classification, and obtained very similar results if age was categorized as $<60$, $[60, 80)$ and $\geq 80$ years. The risk group index (based on our integrative model) was the only factor consistently significant in the univariate and multivariate models across the different data
sets (Table 2). The integrative model showed the highest value of C-index when compared to the other clinical factors as presented in Table 2.

**Prognostic modeling for patients with late-stage EEC**

In parallel, we performed prognostic modeling for patients with late-stage EEC; however, (i) a relatively small sample size may limit our capacity to obtain a reliable model, and (ii) it may not be of great utility as most patients with this diagnosis are treated aggressively. Based on 26 samples of late-stage EEC (stages III and IV) in the training cohort, we developed an integrative model through an initial filtering and the elastic net (see Equation 2 in Materials and Methods), which includes two clinical factors (tumor grade and patient age) and 14 protein markers (Supplemental Table 5). Interestingly, a subset of protein markers had coefficients with signs that were the opposite of those for the early-stage EEC model. This may be due to differences in the context of the markers in tumors associated with good outcomes versus those associated with poor outcomes. The distribution of risk scores was quite similar between late-stage training and validation samples, as shown in Supplemental Figure 3B. To classify low- and high-risk groups, we used the 75th percentile of the risk scores in the training samples as a risk score cutoff to ensure a similar number of events in each risk group, as shown in the second column of Supplemental Figure 3B.

We next applied the model to samples of late-stage EEC in the validation cohort and classified them into two risk groups using the 75th percentile as a cutoff value. The Kaplan-Meier curves for the two risk groups showed a clearer separation in survival as in Supplemental Figure 5B than when the patients were stratified by disease stage as in Supplemental Figure 2B; however, the separation in survival was not as clear as when patients were stratified by tumor
grade as in Supplemental Figure 2D regardless of the p-values. We observed similar patterns with varying cutoff points while ensuring equal sample sizes between the risk groups split by disease stage or tumor grade, as we did for the analysis of early-stage EEC. Kaplan-Meier curves for validation samples are displayed in Supplemental Figure 5D and F, when patients were grouped by the sample numbers of disease stage or tumor grade. Furthermore, we observed similar results using the 25th percentile of risk scores, which was approximately equal to 0.5 (Supplemental 4C and D). Overall, among patients with late-stage EEC, the model based on only the tumor grade outperformed the integrative model in terms of prognostic accuracy. In multivariate analysis for the validation samples, we found that tumor grade, patient age and the risk group index defined by the integrative model were all significant (Supplemental Table 6). Thus, the additional prognostic value obtained from the late-stage EEC integrative model compared to using clinical factors alone seemed modest, and was mainly due to the small number of training samples and intrinsic differences in survival between the patients represented by the training and validation samples, as observed in Supplemental Figure 1B.

Discussion

In this study, we aimed to improve prognosis estimation of EEC patients by incorporating high-throughput protein expression data. We did not intend to invent a new and alternative strategy completely independent from the current staging system. Rather, we utilized the clinical stage as an initial stratification parameter (by classifying patients into early vs. advanced stages: I, II vs. III, IV) and further incorporated protein-based markers designed to provide informative content that the disease stage may not capture. Therefore, we performed prognostic modeling for patients with early- and late-stage EEC, respectively. Importantly, we developed and validated an effective prognostic model for patients with early-stage EEC, which, to the best of our
knowledge, is the first quantitative protein-marker-driven model for this patient population. Although the patients represented in the training and validation cohorts may have some intrinsic differences (e.g., a high rate of obesity among the patients in the United States and different age distributions between the cohorts), the risk group index defined by our model consistently shows an improved discriminating power across the two patient cohorts using different cutoff values, which highlights the robustness of our model. In contrast, major conventional prognostic factors (e.g., disease stage and tumor grade) appear to have little power for this patient population. Therefore, our integrative model for early-stage EEC is of potential clinical utility for identifying/prioritizing the patients with early-stage EEC who have a high risk of disease recurrence and death.

Our integrative model for early-stage EEC contains two clinical features and 18 protein markers. As expected, older age and higher tumor grade are associated with a worse prognosis. As shown in Figure 4, there are two distinct groups of protein markers in regard to prognosis. The proteins for which the expression levels are associated with a good prognosis include EGFR, myosin IIa, AR, c-Myc, STAT3, p38 MAPK, mTOR, fibronectin and HSP70. Among these markers, STAT3, MAPK and mTOR are downstream targets of EGFR signaling. The prognostic effect of EGFR expression in cancer has been controversial in the literature; and it often depends on the specific tumor context, which may be due to the complexities of the downstream effects of EGFR signaling. Our model indicates that a high EGFR level in early-stage EEC is associated with a good prognosis. Indeed, a dual role of EGFR has been reported in endometrial cancer: high EGFR expression in well-differentiated EEC is associated with a low tumor grade and a favorable outcome; whereas high EGFR expression appears to promote disease progression in more aggressive, undifferentiated non-endometrioid tumors (35). In bladder cancer, a high level
of EGFR expression has been reported to be a favorable prognostic factor (36); while in ovarian cancer, a high level of EGFR expression has been reported to be associated with a poor disease-free survival (37). A high level of expression of the androgen receptor has been reported to be a good prognostic marker for breast cancer and for all breast tumors and the triple-negative subtype (38, 39). This observation is potentially related to the tumor differentiation status. The protein markers for which high expression levels are associated with poor prognoses include FASN, acetyl-α-tubulin, ACC1, Ku80, NF-κB, Bim, Chk2, c-Kit and SMAD4. Among them, FASN and ACC1 are both related to fatty acid synthesis, and FASN has been shown to be associated with poor patient survival in urothelial and colorectal cancers (40, 41). Ku80 and Chk2 are DNA damage sensors that may reflect the underlying mechanisms that drive tumors associated with poor outcomes; BIM is regulated by EGFR, which is a marker of good prognosis in our model and may reflect decreased EGFR signaling in the patients with poor prognoses.

In terms of clinical applications, this model can be rapidly translated to the clinic using the RPPA platform which has been made CLIA compliant in several centers and others are moving to make it CLIA compliant. With the relatively small number of 18 protein markers, it is also possible to develop a Luminex assay that would allow rapid transition to the clinic. Finally, a number of multiplex IHC approach are now in place. Thus, the predictors can be readily transitioned to the clinic. Further evaluation using other independent cohorts is essential to establish the ultimate utility of our integrative model.

Patients diagnosed with late-stage EEC (stages III and IV) are usually prescribed aggressive treatment options such as chemotherapy. A potential utility of prognostic modeling for these patients is to identify the individuals with low-risk disease who may not need additional treatment. However, our integrative model did not provide additional prognostic power for this
determination, which may be primarily due to the limited sample size. Another issue is the unbalanced event rates between training and validation sets. Larger patient cohorts will be needed to assess the clinical utility of the integrative model for late-stage EEC.

Our study has several limitations. First, we focused on overall survival as the primary endpoint in our analysis. Although overall survival is often used in the field (42), cancer-specific survival would be a more appropriate endpoint for such a purpose. Unfortunately, cancer-specific survival data are not available in our validation cohort. To minimize the limitation of overall survival (due to cancer-unrelated deaths), survival times were capped at 5 years as in the previous study (22). It would be of high clinical interest to further develop prognostic and predictive models for cancer-specific, disease free and metastasis-free survival based on our patient cohorts when such survival data are available. Second, our analysis only included a limited number of clinical variables due to incomplete clinical information of our patient samples. An important extension would be to integrate other types of clinical variables such as myometrial invasion and lymph node status. Third, we did not include treatment effects in our analysis. Adjuvant therapy (e.g., chemotherapy and/or radiation) may impact survival and thus confound the impact of prognostic factors or markers on survival. Unfortunately, treatment information in our patient cohorts is very incomplete, which prevented us from including these data in the analysis. However, in general, EEC patients with the same or similar stage/grade received similar (adjuvant) therapies. Thus, since (i) our modeling was carried out only for patients with similar disease stages (i.e., early-stage or late-stage) and (ii) the grade effect has been included in our model, this confounding effect would be relatively limited. In future, it is critical to validate our model based on the patient cohorts with high-quality and complete clinical information (such as those from clinical trials). Last but not the least, the current RPPA platform covers only about
200 proteins and therefore we may miss some important protein markers. We are in the process of extending our platform to many more proteins, which may further improve the prognostic power of the model. It remains unclear how several protein markers in our early-stage model such as myosin IIa, fibronectin and HSP70 are related to tumor progression of EEC and affect the clinical outcomes. Further efforts are required to elucidate the role of these proteins in the context of early-stage EEC. However, this mechanistic investigation can be pursued independent of the prognostic values conferred by these markers. Currently, only RPPA data are available for both training and validation cohorts, and we will include molecular features such as somatic mutations and copy number alterations, when the related data are available.

Acknowledgments

We gratefully acknowledge contributions from the Endometrial Cancer Working Group of TCGA Research Network. We thank Kadri Madissoo and Britt Edvardsen (the University of Bergen) for technical support and LeeAnn Chastain (MDACC) for manuscript editing.

Conflict of Interest

The authors declare no competing financial interests.
References


## Tables

Table 1. Clinical characteristics of the training and validation cohorts

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Validation</th>
<th>( P^1 )</th>
<th>( P^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early-stage</td>
<td>Late-stage</td>
<td>Early-stage</td>
<td>Late-stage</td>
</tr>
<tr>
<td># of patients</td>
<td>183</td>
<td>26</td>
<td>333</td>
<td>94</td>
</tr>
<tr>
<td>Age: mean (range)</td>
<td>64.94 (35-93)</td>
<td>67.92 (43-89)</td>
<td>61.38 (24-90)</td>
<td>62.06 (32-89)</td>
</tr>
<tr>
<td>Disease Stage(^3): N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>170</td>
<td>297</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(92.9)</td>
<td></td>
<td>(89.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>13</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7.1)</td>
<td></td>
<td>(10.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>18</td>
<td>71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(69.2)</td>
<td></td>
<td>(75.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>8</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(30.8)</td>
<td></td>
<td>(24.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor Grade: N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>74</td>
<td>3</td>
<td>84</td>
<td>7</td>
</tr>
<tr>
<td>(40.4)</td>
<td></td>
<td>(11.5)</td>
<td>(25.2)</td>
<td>(7.5)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>82</td>
<td>11</td>
<td>167</td>
<td>41</td>
</tr>
<tr>
<td>(44.8)</td>
<td></td>
<td>(42.3)</td>
<td>(50.2)</td>
<td>(43.6)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>27</td>
<td>12</td>
<td>82</td>
<td>46</td>
</tr>
<tr>
<td>(14.8)</td>
<td></td>
<td>(46.2)</td>
<td>(24.6)</td>
<td>(48.9)</td>
</tr>
<tr>
<td>Deaths(^4): N (%)</td>
<td>17</td>
<td>16</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>(9.3)</td>
<td></td>
<td>(61.5)</td>
<td>(6.3)</td>
<td>(21.3)</td>
</tr>
<tr>
<td>Recurrences(^5): N (%)</td>
<td>27</td>
<td>18</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>(14.8)</td>
<td></td>
<td>(69.2)</td>
<td>(10.8)</td>
<td>(49.4)</td>
</tr>
</tbody>
</table>

\(^1\)P-values comparing the patients with early-stage EEC in the training cohort with those in the validation cohort. For age, the two-sample t-test was used. For stage and grade, the chi-squared test was used.

\(^2\)P-values comparing the patients with late-stage EEC in the training cohort with those in the validation cohort. For age, a two-sample t-test was used. For stage and grade, the chi-squared test was used.

\(^3\)FIGO 2009 Criteria

\(^4\)Survival times were capped at 60 months.

\(^5\)Data on recurrence were not available for 27 early-stage validation samples and 7 late-stage validation samples. Recurrence time was defined as the time from initial surgery to the first documented progression or recurrence or the last follow-up in the absence of progressive disease. Recurrence times were capped at 60 months.
Table 2. Univariate and multivariate Cox proportional hazards model analysis for the patients with early-stage EEC

<table>
<thead>
<tr>
<th></th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR(^1) (90% CI)</td>
<td>Wald (P^2)</td>
</tr>
<tr>
<td><strong>Training Set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease Stage: II (vs. I)</td>
<td>1.88 (0.55, 6.47)</td>
<td>0.40</td>
</tr>
<tr>
<td>Tumor Grade(^4)</td>
<td>1.71 (0.98, 2.98)</td>
<td>0.12</td>
</tr>
<tr>
<td>Patient Age(^5): ≥60 yrs (vs.&lt;60 yrs)</td>
<td>2.38 (0.84, 6.76)</td>
<td>0.17</td>
</tr>
<tr>
<td>Integrative Model: High risk (vs. low risk)</td>
<td>16.72 (7.39, 37.80)</td>
<td>(1.5\times10^8)</td>
</tr>
<tr>
<td><strong>Validation Set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease Stage: II (vs. I)</td>
<td>1.43 (0.51, 3.98)</td>
<td>0.57</td>
</tr>
<tr>
<td>Tumor Grade(^4)</td>
<td>1.38 (0.83, 2.30)</td>
<td>0.30</td>
</tr>
<tr>
<td>Patient Age(^5): ≥60 yrs (vs. &lt;60 yrs)</td>
<td>3.05 (1.23, 7.61)</td>
<td>0.044</td>
</tr>
<tr>
<td>Integrative Model: High risk (vs. low risk)</td>
<td>5.26 (2.37, 11.65)</td>
<td>(6.3\times10^{-4})</td>
</tr>
</tbody>
</table>

\(P\)-values < 0.1 are in boldface.

\(^1\)HR: hazard ratio
\(^2\)Wald’s test \(p\)-values
\(^3\)Log rank test \(p\)-values
\(^4\)In Cox regressions, grade was treated as a continuous variable with natural ordering so as to compute only one regression coefficient for this variable. However, log-rank test \(p\)-values were computed by treating grade as a discrete variable.
\(^5\)Age was dichotomized to 60 years and older vs. younger than 60 years since patients 60 years of age and older are generally advised to receive adjuvant therapy.
\(^6\)Based on a multivariate Cox regression model that included all the variables in the table.
Figure Legends

Figure 1. Kaplan-Meier survival curves for patients with early-stage EEC (stages I & II) in the training and validation cohorts. For (A) and (B), patients are split by disease stage; for (C) and (D), patients are split by tumor grade.

Figure 2. A flow chart for construction and validation of the prognostic model for overall survival in early-stage endometrioid endometrial cancer.

Figure 3. Kaplan-Meier survival curves for patients with early-stage EEC (stages I & II) in the training and validation cohorts. Risk scores for patients were computed using the early-stage EEC prognostic model. Based on the risk scores, for (A) and (B), patients were classified into two risk groups using the 90th percentile of the risk scores of training samples as a cutoff. To check the robustness of our model, we used different cutoffs when classifying the patients into risk groups instead of the 90th percentile of risk scores. In (C) and (D), we used a cutoff value that makes the numbers of patients in low and high risk groups equal to the numbers of cases with Stage I and II, respectively. In (E) and (F), we used a cutoff value that makes the numbers of patients in three risk groups equal to the numbers of cases with Grade I, II and III, respectively.

Figure 4. Expression patterns of the protein markers and outcome prediction of the early-stage EEC prognostic model. Risk groups were defined by the 90th percentile of the risk scores of the training samples.
Figure 1

(A) Stage-wise comparison

(B) Stage-wise comparison

(C) Grade-wise comparison

(D) Grade-wise comparison
Figure 2

Modeling

Training Set (BERGEN)

Early Stage (Stages I & II)
# patients=183, # features=189

Initial Filtering
# features=31

Elastic net using leave-one-out CV

Integrative Model
# features=20

Testing

VALIDATION

Early Stage
# patients=333

Validation Set (MDACC & TCGA)
Figure 3

Based on Integrative Model

(A) Data-driven cutoff: 90th percentile

(B) Data-driven cutoff: 90th percentile

(C) Cutoff determined by # cases in different disease stage groups

(D) Cutoff determined by # cases in different disease stage groups

(E) Cutoff determined by # cases in different tumor grade groups

(F) Cutoff determined by # cases in different tumor grade groups
Figure 4

The figure shows a heat map of gene expression data for two groups: Training (n=183) and Validation (n=333). The heatmap is color-coded to represent different protein expression levels. The x-axis represents the different proteins, and the y-axis represents the sample number. The color intensity indicates the level of expression, with red representing high expression and green representing low expression.

The heatmap is divided into two categories based on prognosis: Good prognosis proteins and Poor prognosis proteins. The y-axis also includes a risk score, which is color-coded to distinguish between Low risk group and High risk group.

The proteins listed include EGFR, p-Myosin IIa, AR, c-Myc, p-STAT3, p-p38 MAPK, p-mTOR, Fibronectin, Hsp70, FASN, Acetyl-a-Tubulin, ACC1, Ku80, p-NF-kB p65, Bim, Chk2, c-Kit, and SMAD4.

The figure illustrates the correlation between gene expression and prognosis, with the risk score indicating the likelihood of a favorable or unfavorable outcome.
Integrative Protein-Based Prognostic Model for Early Stage Endometrioid Endometrial Cancer

Ji-Yeon Yang, Henrica M.J. Werner, Jane Li, et al.

Clin Cancer Res  Published OnlineFirst July 29, 2015.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-15-0104

Supplementary Material  Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2015/08/08/1078-0432.CCR-15-0104.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.