

## Predicting high-risk endometrioid carcinomas using proteins

Di Du<sup>1</sup>, Wencai Ma<sup>1</sup>, Melinda S. Yates<sup>2</sup>, Tao Chen<sup>3</sup>, Karen H. Lu<sup>2</sup>, Yiling Lu<sup>4</sup>, John N. Weinstein<sup>1</sup>, Russell R. Broaddus<sup>5</sup>, Gordon B. Mills<sup>4</sup> and Yuexin Liu<sup>1</sup>

<sup>1</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>2</sup>Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>3</sup>Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital of Fudan University, Shanghai, China

<sup>4</sup>Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>5</sup>Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

**Correspondence to:** Yuexin Liu, **email:** yliu8@mdanderson.org

**Keywords:** RPPA; stage; protein; biomarker; endometrioid carcinoma

**Received:** September 14, 2017

**Accepted:** February 24, 2018

**Published:** April 13, 2018

**Copyright:** Du et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**Background:** The lethality of endometrioid endometrial cancer (EEC) is primarily attributable to advanced-stage diseases. We sought to develop a biomarker model that predicts EEC surgical stage at the time of clinical diagnosis.

**Results:** PSES was significantly correlated with surgical stage in the TCGA cohort ( $P < 0.0001$ ) and in the validation cohort ( $P = 0.0003$ ). Even among grade 1 or 2 tumors, PSES was significantly higher in advanced than in early stage tumors in both the TCGA ( $P = 0.005$ ) and MD Anderson Cancer Center (MDACC) ( $P = 0.006$ ) cohorts. Patients with positive PSES score had significantly shorter progression-free survival than those with negative PSES in the TCGA (hazard ratio [HR], 2.033; 95% CI, 1.031 to 3.809;  $P = 0.04$ ) and validation (HR, 3.306; 95% CI, 1.836 to 9.436;  $P = 0.0007$ ) cohorts. The ErbB signaling pathway was most significantly enriched in the PSES proteins and downregulated in advanced stage tumors.

**Methods:** Using reverse-phase protein array expression profiles of 170 antibodies for 210 EEC cases from TCGA, we constructed a Protein Scoring of EEC Staging (PSES) scheme comprising 6 proteins (3 of them phosphorylated) for surgical stage prediction. We validated and evaluated its diagnostic potential in an independent cohort of 184 EEC cases obtained at MDACC using receiver operating characteristic curve analyses. Kaplan-Meier survival analysis was used to examine the association of PSES score with patient outcome, and Ingenuity pathway analysis was used to identify relevant signaling pathways. Two-sided statistical tests were used.

**Conclusions:** PSES may provide clinically useful prediction of high-risk tumors and offer new insights into tumor biology in EEC.

### INTRODUCTION

Endometrial carcinoma is the most common gynecologic malignancy; over 10,470 deaths from uterine corpus cancer were expected in the United States in 2016 [1], an approximately three-fold increase over the past 25 years. The lethality of endometrial cancer is primarily

correlated with stage III or IV (hereafter referred to as advanced stage) disease [2]. Typically, a five-year survival rate of 83%–97% is achieved for localized disease (stage I/II, hereafter referred to as early stage), in contrast to 43%–67% for stage III disease and only 13%–25% for stage IV disease [3]. Patients with advanced stage disease are at an increased risk of recurrence and are typically treated with

postoperative radiotherapy and/or chemotherapy; however, this is rarely curable [4]. Therefore, early stratification of endometrial carcinomas into surgical staging categories from tissue biopsies will help gynecologic oncologists more accurately choose which women should have extensive surgical staging procedures or receive systemic adjuvant therapy, and thus represents a key to reducing the morbidity and mortality associated with endometrial cancer.

Non-endometrioid carcinoma is known a priori to be clinically aggressive and can be readily segregated from its endometrioid counterpart at the time of diagnosis [5] because these two entities exhibit striking difference in histopathological (microscopic) appearance and molecular characteristics such as *TP53* mutation, [6] *CTNNB1* mutation [7], protein levels of estrogen receptor/progesterone receptor (ER/PR) [8], and gene expression profiling [9]. Endometrioid-type endometrial cancer (EEC), accounting for approximately 70 to 80% of endometrial cancer cases, is the focus of the current study. EEC is pathologically staged with the International Federation of Gynecology and Obstetrics (FIGO) system [2]. However, molecular characteristics have not been validated to help predict operative staging of EEC patients at the time of clinical diagnosis [10, 11]. Since molecular abnormalities predispose the manifestation of microscopic appearance and drive tumor progression, molecular biomarkers not only provide an opportunity for early detection but also offer the ability to direct therapeutic strategies; both are distinctive from the current microscopic analysis.

Cellular proteins are responsible for functional diversity because a vast array of regulatory or metabolic processes occur at the protein level; these cannot be adequately predicted from DNA or RNA data. The reverse-phase protein array (RPPA) platform allows identification of proteins for use as therapeutic intervention or as markers for the classification of cancer [12–14]. In this study, we analyzed the protein expression profiles of over 200 EEC cases with clinicopathologic characteristics obtained from The Cancer Genome Atlas (TCGA) and developed a Protein Scoring of EEC Staging (PSES) scheme that we further validated in an independent MD Anderson Cancer Center (MDACC) cohort comprising an additional 184 EEC cases. Prognostic and biological significance of PSES was further investigated. Our work demonstrates the potential of clinical stratification and novel therapeutic targets.

## RESULTS

### Patient characteristics

Clinicopathologic characteristics of patients included in the TCGA and MDACC cohorts are described and compared in Table 1. All patients were endometrial cancer with endometrioid histology. The two cohorts had a comparable percentage of advanced stage disease

(TCGA, 21.4%; MDACC, 26.1%), but there were significantly more grade 3 tumors in the TCGA cohort than in the MDACC cohort (39.0% versus 19.3%,  $P < 0.0001$ , Fisher's exact test). No statistical difference was observed between these two cohorts for patient age, death or recurrence events, or patient outcome, including overall survival (OS) and progression-free survival (PFS) (Supplementary Figure 1).

### Generation of PSES

Using the random partitioning method and an average  $P$  value cutoff of 0.05 (see Methods for details), we identified four proteins or phosphoproteins that were significantly associated with advanced stage disease: one upregulated protein (Dvl3) and three downregulated phosphoproteins, Shc-pY317, JNK-pT183-pT185, and HER3-pY1298 (hereafter referred to as pSHC, pJNK, and pHER3, respectively) (Figure 1A). The relative expression levels of these four proteins for each individual across the entire TCGA cohort were also depicted (Figure 1B). Consistent with the results rendered by the random approach, the four proteins were significantly differentially expressed between early and advanced stage tumors in the entire TCGA cohort. Moreover, the advanced stage tumors were significantly associated with grade 3 tumors and recurrent events, but not with MSI status or patient age (Figure 1B). Given that EEC tumors are typically characterized by the expression levels of estrogen receptor alpha (ER) and progesterone receptor (PR), we included these two proteins in the predictive model to account for baseline fluctuation on other clinicopathological variables such as age, grade, and recurrence status. A similar approach of inclusion of endogenous hormone levels in risk prediction models for postmenopausal breast cancer was applied previously in improving the ability to identify high-risk women [15]. Thus, total six proteins were used to compose the predictive model, and we termed this predictive protein set PSES (Protein Scoring of EEC Staging). The predictive model, expressed as PSES score, was defined as a linear combination of the six protein expression levels weighted by coefficients that were predefined on the basis of ROC analysis of the TCGA cohort (see Methods for details).

### Association of PSES with surgical stage in patients with EEC

We calculated PSES scores for each sample in the TCGA cohort. Values ranged from -1.9865 to 1.6682 across the 210 samples, of which 104 had positive PSES scores and 106 had negative scores. PSES scores were significantly higher in patients with advanced stage disease than in patients with early stage disease ( $P < 0.0001$ , Mann–Whitney test) (Figure 2A). To examine the impact of tumor purities on this result, we carried out two

**Table 1: Clinicopathologic characteristics of EEC patients in the training (TCGA) and validation (MDACC) cohorts\***

	TCGA (n = 210)	MDACC (n = 184)	P †
<b>Age</b>			
Mean, years [SD]	60.3 [11.3]	59.8 [11.6]	0.3934 <sup>‡</sup>
Range	30.5–87.5	27.2–84.7	
<b>FIGO Stage</b>			
I/II	165 (78.6)	136 (73.9)	
III/IV	45 (21.4)	48 (26.1)	0.2870 <sup>†</sup>
<b>Histological Grade</b>			
1/2	128 (61.0)	146 (80.7)	
3	82 (39.0)	35 (19.3)	< 0.0001 <sup>†</sup>
Unknown	0	3	
<b>Vital Status</b>			
Living	190 (90.5)	158 (87.8)	
Deceased	20 (9.5)	22 (12.2)	0.4162 <sup>†</sup>
Unknown	0	4	
<b>Recurrent Disease</b>			
No	162 (81.0)	123 (77.8)	
Yes	38 (19.0)	35 (22.2)	0.5079 <sup>†</sup>
Unknown	10	26	
<b>MSI Status</b>			
MSI <sup>‡</sup>	91 (43.3)	NA	
MSS	119 (56.7)	NA	

Abbreviations: TCGA, The Cancer Genome Atlas; MDACC, MD Anderson Cancer Center; SD, standard deviation; FIGO, International Federation of Gynecology and Obstetrics; MSI, microsatellite instability; MSS, microsatellite stable; NA, not applicable.

\*Values are reported as No. (%). Missing values are excluded from percentage calculation and statistical test.

†Statistical comparison of clinicopathological features between patients in the TCGA cohort and those in the MDACC cohort.

‡Mann–Whitney test.

†Fisher’s exact test.

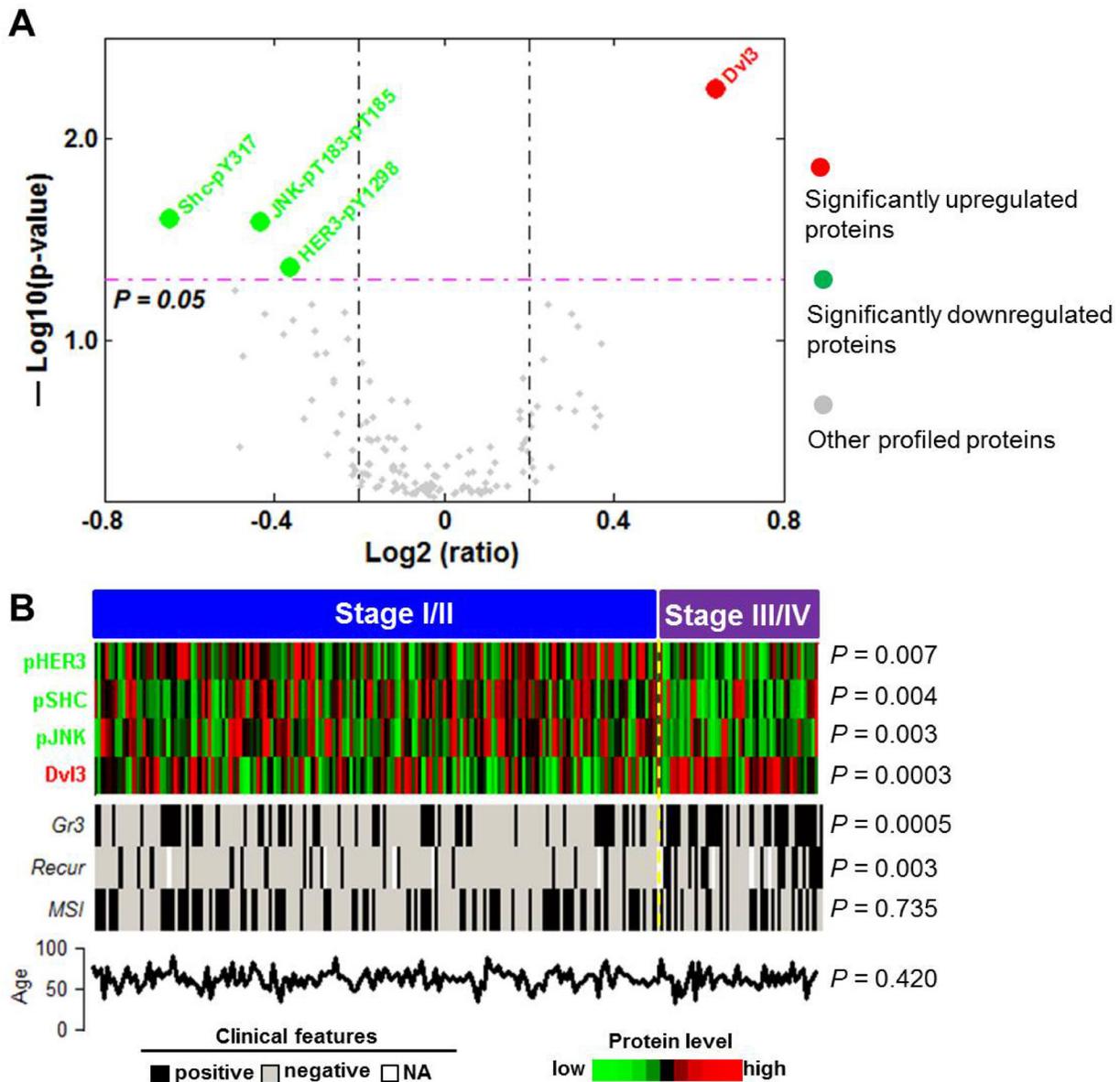
‡Including both MSI-H and MSI-L.

additional analyses. First, we compared tumor purities between early-stage and advanced-stage disease and found there was no significant difference in these two groups of patients ( $P = 0.4459$ , Supplementary Figure 2A). Secondly, we calculated the purity adjusted PSES score by dividing the original PSES score by tumor purity, and the purity-adjusted PSES score remained significant correlation with patient surgical stage ( $P = 0.0002$ , Supplementary Figure 2B). Note that the statistical significance was slightly compromised likely because the number of analyzed samples becomes smaller when tumor purity data were incorporated into this analysis. In addition, we obtained the percentage of stroma, lymphocyte, macrophage, and neutrophil cells in the EEC tissue specimen and found that PSES score was not

significantly correlated with these cellular compositions (Supplementary Figure 3). Collectively, these data indicate that PSES is significantly correlated with surgical stage, independent upon tumor purities. We next validated PSES in an independent patient cohort and performed RPPA profiling on an additional set of 184 EEC cases that were not included in the TCGA cohort. PSES score for each sample was computed in a similar manner in the validation set. Patients with advanced stage disease demonstrated significantly higher PSES scores than those with early stage disease ( $P = 0.0003$ , Mann–Whitney test) (Figure 2B), which was consistent with the results in the TCGA cohort. Because low-grade (Grade 1 or 2) tumors are less likely to have advanced stage disease, it is clinically of interest to be able to predict advanced stage tumors

among patients with low-grade EEC. Prominently, among patients with grade 1 or 2 tumors, those with advanced stage disease had statistically significantly higher PSES scores than did those with early stage disease in both the TCGA dataset ( $P = 0.0046$ , Mann–Whitney test) (Figure 2C) and the MDACC dataset ( $P = 0.0055$ ) (Figure 2D). Taken together, these results demonstrated that PSES is significantly correlated with tumor surgical stage even among patients with low-grade disease, as evidenced by an external validation.

To test whether PSES is an independent predictor of EEC patients with advanced stage disease, we further preformed multivariate logistic regression analyses. Even after adjusting for patients' age, grade, vital status, or recurrent disease status, the odds ratios for patients with high PSES scores were 4.64 (95% CI = 2.06 to 10.48,  $P = 2.2E-04$ ) in the TCGA cohort and 2.94 (95% CI = 1.23 to 7.00,  $P = 0.015$ ) in the MDACC cohort (Table 2). In addition, we calculated the sensitivity and specificity for each of the individual variables to predict EEC patients

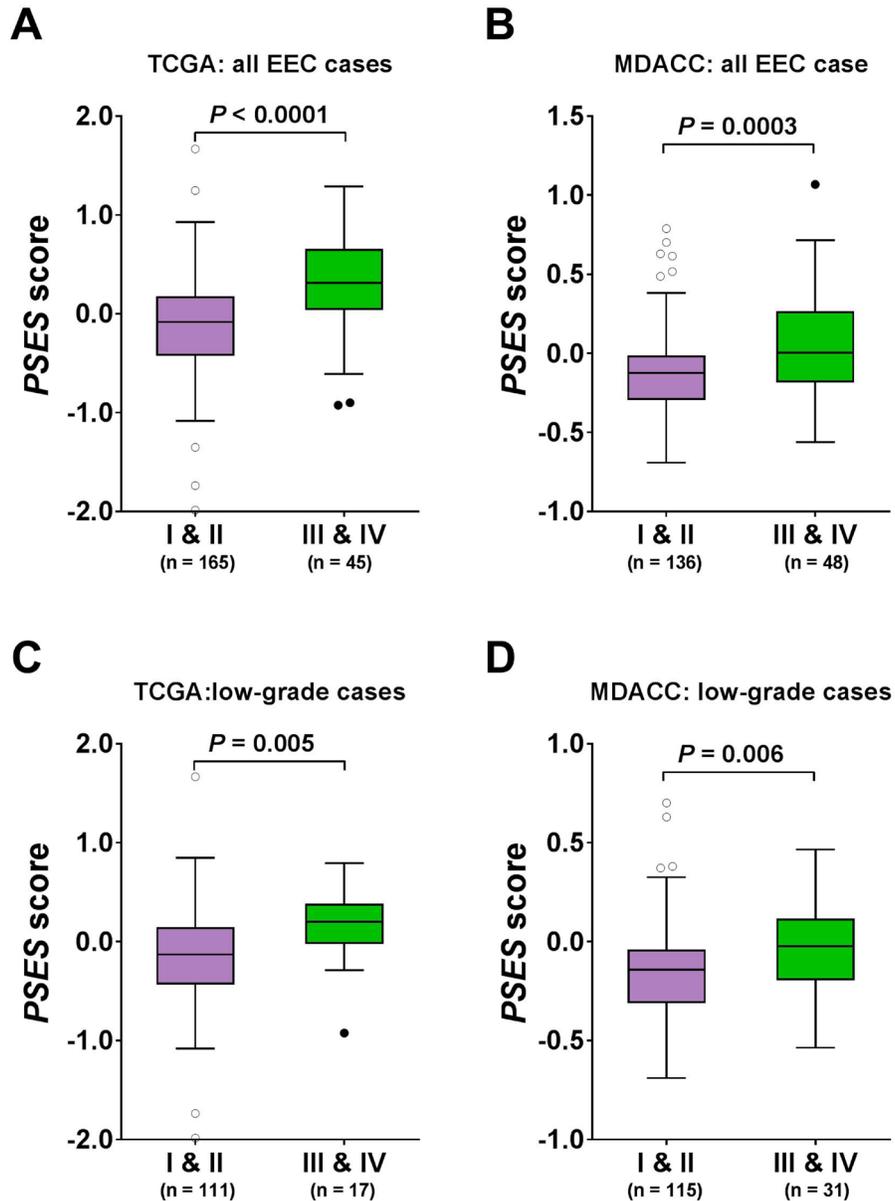


**Figure 1: Identification of protein markers used for construction of PSES.** (A) Volcano plot shows the protein expression difference (in terms of  $\log_2$  ratio) and the corresponding statistical tests (in terms of  $-\log_{10}(P \text{ value})$ ) between Stage III/IV tumors and Stage I/II tumors. Values were averaged from those calculated from the 500 different subsets randomly selected from the TCGA cohort. Four proteins exhibited a statistically significant difference. The dashed pink line indicates the  $P$  value cutoff of 0.05. (B) Relative expression levels of the four proteins and clinicopathologic characteristics for each individual patient in the entire TCGA cohort ( $n = 210$ ). The  $P$  values show the comparison between the early stage patients vs advanced stage patients and are calculated from the entire TCGA cohort.

with advanced stage disease (Supplementary Table 1). As compared to the other predictors, PSES improves the predictive sensitivity very much. We next correlated the PSES score with gene mutations frequently observed in endometrial cancer (Supplementary Figure 4). Patients with *TP53* mutation had significantly higher PSES scores than those with *TP53* wide-type. No significant correlation was observed between PSES scores and *PTEN* mutation or *CTNNB1* mutation.

## Association of PSES with patient outcomes

While PSES was generated on the basis of tumor stage, we additionally performed Kaplan-Meier survival analysis to assess the capacity of PSES to differentiate patients by PFS. Patients with positive PSES scores had statistically significantly worse PFS in both the TCGA cohort ( $P = 0.04$ , log-rank test) (Figure 3A) and the MDACC cohort ( $P = 0.0007$ , log-rank test) (Figure 3B).



**Figure 2: Association of PSES with tumor stage in patients with EEC.** (A) Box plots represent PSES scores in EEC samples from patients with stage I & II disease ( $n = 165$ ) and from patients with stage III & IV disease ( $n = 45$ ) in the TCGA cohort. The central line of each box is the median value, and the edges are the 25th and 75th percentiles. The whiskers extend to the 10th and 90th percentiles, and data points outside the whiskers are plotted individually as circles or dots.  $P$  values were calculated using two-sided Mann–Whitney test. (B) The predictive performance of PSES score was validated in an independent MDACC data set. PSES scores were significantly higher in patients with stage III & IV disease in the MDACC cohort ( $P = 0.0003$ , Mann–Whitney test). Among the low-grade (grade 1 & 2) EEC patients, tumors with advanced stage disease had statistically significantly higher PSES scores than did those with early stage disease in both (C) the TCGA cohort ( $n = 128$ ,  $P = 0.0046$ , Mann–Whitney test) and (D) the MDACC cohort ( $n = 146$ ,  $P = 0.0055$ , Mann–Whitney test).

**Table 2: Multivariate logistic analyses for PSES scores and various diagnostic factors in patients with EEC**

	Variables <sup>†</sup>	OR (95% CI)	P
<i>Advanced stage vs early stage tumors</i>			
TCGA	PSES score	3.96 (1.80 to 8.73)	.001
	Age, >60 years vs ≤60 years	0.53 (0.24 to 1.17)	.117
	Grade, Gr3 vs Gr1/2	2.33 (1.06 to 5.12)	.036
	Vital status, deceased vs living	1.87 (0.50 to 7.02)	.355
	Recurrence, yes vs no	4.72 (1.88 to 11.86)	.001
<i>Advanced stage vs early stage tumors</i>			
MDACC	PSES score	5.37 (1.27 to 22.65)	.022
	Age, >60 years vs ≤60 years	1.74 (0.70 to 4.32)	.228
	Grade, Gr3 vs Gr1/2	0.92 (0.31 to 2.76)	.882
	Vital status, deceased vs living	1.32 (0.30 to 5.92)	.712
	Recurrence, yes vs no	10.44 (3.55 to 30.73)	<.001

Abbreviations: CI, confidence interval; OR, odds ratio.

<sup>†</sup>PSES score was treated as a continuous variable and all other covariates were binary: age (0 for an age of 60 years or less and 1 for an age of greater than 60 years), grade (0 for a grade of 1 or 2 and 1 for a grade of 3); vital status (0 for living and 1 for deceased), and recurrence (0 for a tumor with no recurrence and 1 for a tumor with recurrence).

To test whether this result was independent of tumor stage, we performed two different types of analyses. First, we added both PSES score and tumor stage as covariates to a Cox proportional hazards model and estimated the hazard ratio statistic. Although stage remained a strong predictor in both cohorts, positive PSES score was significantly associated with shorter PFS in the MDACC cohort even after adjustment by tumor stage (HR, 2.33, 95% CI = 1.10 to 4.93,  $P = 0.026$ ) (Figure 3C). Secondly, we stratified patients into either early-stage or advanced-stage group and performed Kaplan-Meier survival analysis using the PSES scores separately in these two groups. PSES score was significantly associated with PFS in either group in the MDACC cohort but not in the TCGA cohort (Supplementary Figure 5). Taken together, these data suggested that PSES score is significantly correlated with patient PFS, independent of surgical stage. Although patients with higher PSES score appeared to have worse overall survival, this trend was not statistically significant (Supplementary Figure 6).

### Biological Interpretation of PSES

To characterize the biological properties of PSES, we used publicly available tools to associate the predictive proteins with gene ontology (using the GO database) [16] as well as pathway annotation (Ingenuity Pathway Analysis, <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>). The Ingenuity Knowledge Base including all proteins was used as a reference set and the statistical significance was determined by the Fisher's exact test. We focused the analyses on the four differentially

expressed proteins/phosphoproteins. Dvl3 is a binding protein that controls signal transduction activity in the Wnt pathway and that all three phosphorylated proteins (pHER3, pSHC, and pJNK) are involved in kinase signaling through cell surface receptors (Supplementary Table 2). Pathway analysis showed that ErbB signaling is most significantly enriched in the PSES proteins ( $P = 5.1E-05$ ) (Figure 4A). This pathway is associated with tumor angiogenesis and involves the three phosphorylated proteins that were significantly downregulated in advanced stage tumors (Supplementary Figure 7). Interestingly, a higher frequency of ErbB expression and activation is present in ductal carcinoma *in situ* than in invasive breast cancer. Canonically, Shc1, a docking partner of ErbB receptors [17], regulates the JNK signaling pathway [18]. Moreover, the expression levels of these three phosphoproteins were strongly and significantly correlated with one another, further suggesting significant inactivation of this signaling pathway in advanced stage EEC tumors (Figure 4B).

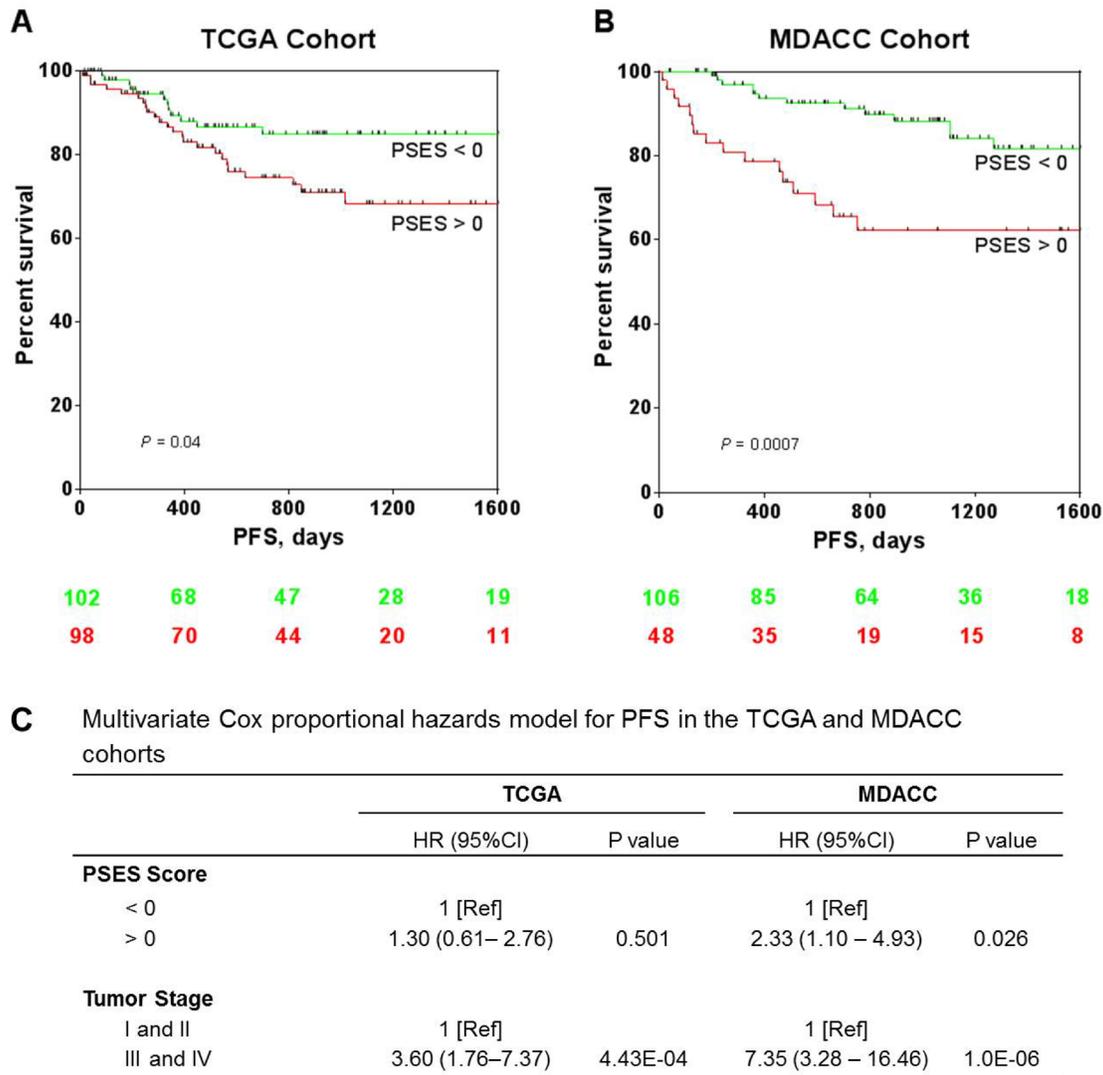
### DISCUSSION

Using RPPA profiling of nearly 400 EEC samples, the largest cohort with protein data yet reported, in this study we developed a predictive scheme that successfully distinguished patients with advanced stage disease from those with early stage disease, independent of patient's age, tumor grade, vital status, or recurrence status. We further demonstrated that PSES has both prognostic value and biological interest. The clinical stage determined at the initial diagnosis is typically based on the results of a physical exam, biopsy, and imaging tests before

surgery. The pathologic stage is determined much later by examining tissue removed during an operation, and thus gives the health care team more precise information that can be used to predict treatment response and outcome. One potential use of the PSES algorithm would be in analysis of preoperative EEC tumor biopsies obtained during initial diagnosis that could then be used to direct the appropriate surgical approach. Specifically, our PSES score could help select for patients who should undergo more extensive lymph node dissections at the time of hysterectomy [19]. Given its strong correlation with pathological stage, the clinical stage information should be incorporated into the PSES prediction scheme in order to improve the predictive sensitivity and

specificity. A protein-driven model that can be evaluated using immunohistochemistry may be more practical and accurate in clinical management [20], while also being less susceptible to expression variations of individual proteins. Different from other predictors such as histologic grade, patient age, or even clinical stage, the PSES predictor projects treatment strategies and likely facilitates identification of novel targets for therapeutic intervention.

Feature selection based on statistical comparison is dependent upon sample size and patient characteristics. For example, the gene sets discovered to be associated with platinum-based chemotherapy resistance exhibited a wide range of 14 to 1,727 genes from several ovarian cancer studies (sample sizes of 6 to 119), and only seven



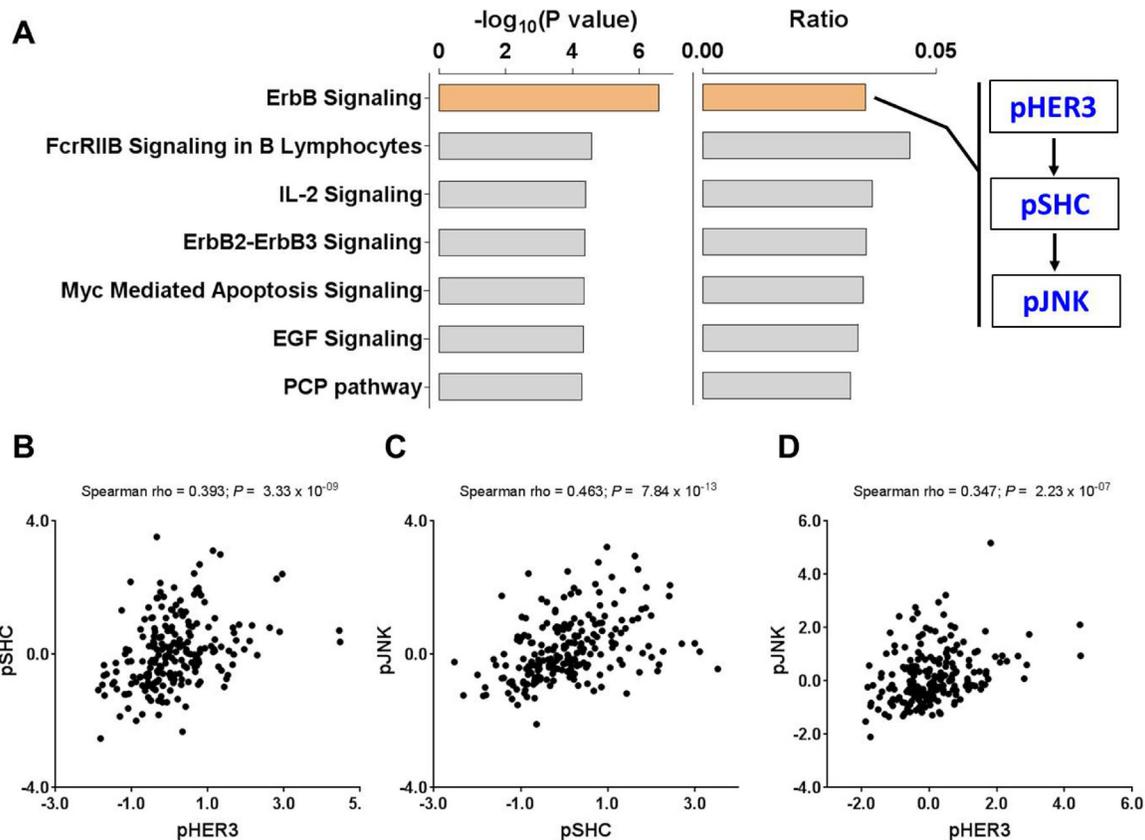
**Figure 3: Association of PSES with tumor prognosis in patients with EEC.** (A) The PFS rate in EEC patients with positive PSES scores ( $n = 98$ ) was statistically significantly lower than that for patients with negative PSES scores ( $n = 102$ ) ( $P = 0.04$ , log-rank test) in the TCGA cohort. (B) The PFS rate in EEC patients with positive PSES scores ( $n = 48$ ) was statistically significantly lower than that for those with negative PSES scores ( $n = 106$ ) ( $P = 0.0007$ , log-rank test) in the MDACC cohort. Patients with missing survival data were excluded from this analysis. The numbers of patients at risk at various time points are shown below each curve. (C) Multivariate Cox proportional hazards model analysis of PFS with PSES score and patient tumor stage as covariates in both the TCGA and MDACC cohorts.

genes were observed as overlapping [21]. To capture these factors, we sought to identify informative features that exhibited statistically significant differences on average in 500 different training sets that were randomly selected from the TCGA cohort and which comprised different numbers of samples. A predictive model consisting of the thus-selected features was robust to tumor heterogeneity and sample size, as demonstrated by validation in an independent cohort.

PSES also offers biological insight into endometrial tumor progression. Interestingly, the differentially expressed proteins are significantly enriched in the ErbB signaling pathway, which involves the three phosphoproteins (pHER3, pSHC, pJNK) that were significantly downregulated in advanced stage EEC tumors. Strong expression correlation among these phosphoproteins further supported an abnormality in this signaling pathway. Dysregulation of these proteins or related pathways in the role of tumor suppression or in disease with favorable prognosis has previously been demonstrated in several other cancers. Activation of the JNK signaling pathway was previously reported to contribute to apoptosis and growth inhibition in human

hepatoblastoma [22] and osteosarcoma [23] cells. In contrast, JNK deficiency significantly increases tumor formation in breast cancer [24]. Compared with other molecular subtypes, luminal A/B breast cancer, which has a relatively favorable prognosis, exhibits higher ERBB3 expression [25]. Likewise, low expression level of p66, an isoform of pSHC, was recently reported to be significantly correlated with worse survival rate in lung cancer [26]. Collectively, these findings are biologically consistent with the downregulation of the ErbB signaling pathway in advanced stage tumors in the setting of EEC. Consistently, higher expression levels of the EGFR protein (a member of the ErbB family) and p38 MAPK phosphoprotein (a member of the MAPK family) were previously reported to be associated with a good prognosis in early-stage EEC [27].

On the other hand, the disheveled segment polarity protein 3 (encoded by *DVL3*), a key mediator of Wnt/ $\beta$ -catenin signaling [28], which was significantly upregulated in the advanced stage EEC tumors, has been shown to be a driver of lung cancer metastases [29, 30]. Tumors with a relatively lower level of Dvl3 were correlated with improved sensitization to IGF1R inhibition and longer PFS



**Figure 4: Biological interpretation of PSES.** (A) Significantly enriched pathways in the four differentially expressed proteins used for construction of PSES where cartoon of ErbB signaling pathway involving the three phosphorylated proteins is also shown. The right plot on this panel shows the ratio of genes in the PSES proteins to all the genes included in the signaling pathway. (B–D) Expression correlation of the three proteins that are involved in the ErbB signaling pathway where (B) pHER3 vs pSHC, (C) pSHC vs pJNK and (D) pHER3 vs pJNK.

in patients treated with IGF1R antibodies [31]. The Wnt/ $\beta$ -catenin signaling pathway appears to be a predominant factor driving tumor progression in endometrial cancer, though mediators for the pathway activation are different from low-grade to high-grade EEC. The Wnt pathway is activated either by  $\beta$ -catenin exon 3 mutations in low-grade endometrioid carcinoma [7] or by overexpression of Dvl3 protein in high-grade endometrioid carcinoma.

It was previously reported that two randomized clinical trials demonstrated different effects of adjuvant chemotherapy on PFS [32]. In striking contrast to a lack of effect in the MaNGO trial, which recruited over 99% Stage IIB-IIIC patients, the NSGO/EORTC trial, which consisted of 98% Stage IA-IIA patients, showed a favorable effect of chemotherapy on PFS. These results suggest that benefit from adjuvant chemotherapy may reside largely within the cohort of patients with stage I-II tumors, which could be reliably predicted by PSES in this study. Nonetheless, the association of PSES with stage should be validated in a prospectively acquired EEC cohort before we could translate it into routine clinical practice.

In summary, PSES is a useful prognostic factor and offers potential targets for therapeutics. The finding has clinical implications for risk assessment and early intervention for patients with endometrial cancer.

## MATERIALS AND METHODS

### Patient samples

Protein expression profiling and clinicopathologic annotation for 210 EEC cases were obtained from the TCGA data portal [6] on March 1, 2013. Patients underwent surgical resection but received no prior treatment for their disease. Clinicopathologic diagnoses were made at local tissue source sites and further confirmed by TCGA. Cases were reviewed and pathologically staged according to the 2009 FIGO staging system into 4 major sub-stage divisions (stage I, II, III, and IV) [6]. As a validation cohort, an additional 184 EEC cases were diagnosed at MDACC from 1998 to 2009, and the tumor samples were reviewed for grade and stage by two independent pathologists. The study was approved by the National Cancer Institute (training set) and by the institutional review board at MDACC (validation set). Detailed patient demographic and clinicopathologic characteristics for both cohorts are described and compared in Table 1.

### RPPA Profiling

Reverse-phase protein array (RPPA) is a high-throughput antibody-based technique for simultaneously measuring protein expression levels in a large number of biological samples [33]. Quantitative protein expression profiles in the training set were obtained from TCGA [6]. The 210 TCGA EEC samples included in this study were those who had both RPPA data and clinical annotation.

RPPA profiling of samples in the validation set (MDACC cohort) was measured and pre-processed, including sample-wise median center and log 2 transformation, at the MD Anderson RPPA core facility using standard operating procedures and validated antibodies [33]. For the purpose of comparison between TCGA (170 antibodies) and MDACC (187 antibodies) data sets, a total of 138 antibodies that were in common across these two cohorts were selected for all downstream analyses.

### Identification of predictive proteins

The overall flow chart of the study design shown in Supplementary Figure 8 summarizes the procedure used to construct and validate the protein-based scoring scheme for predicting clinical stage of EEC patients. The predictive proteins (features) were first identified to be differentially expressed via supervised analysis of the advanced versus early stage tumors in the TCGA set. Firstly, we randomly selected a subset of patients from the TCGA cohort with a set size ranging from 30 to 210. Then, we performed supervised analysis to identify differentially expressed proteins/phosphoproteins between advanced stage versus early stage tumors within this subset and calculated expression (fold-change in term of log2 ratio) and statistical ( $P$  value) differences for all proteins and phosphoproteins. We repeated this process 500 times (Supplementary Figure 9). For each protein, the fold-changes and  $P$  values varied with subsets (Supplementary Figure 10), presumably because different subsets had different sample sizes and different patient characteristics. We then calculated the arithmetic mean of the fold-changes and geometric mean of the  $P$  values for each protein from all 500 randomly selected subsets. Using an average  $P$  value cutoff of 0.05, we identified the predictive protein to use in developing the prediction algorithm. EEC tumors are characterized by expression of estrogen receptor alpha (ER) and progesterone receptor (PR); hence, we included these two proteins in the predictive model to account for baseline fluctuation on other clinicopathological variables such as age, grade, and recurrence status.

### Construction of predictive model

The predictive model was constructed using a weighted voting algorithm [34, 35]. The protein scoring of EEC staging (PSES) of the  $j^{\text{th}}$  sample is defined as

$$\text{PSES}_j = \sum_{i=1}^N \omega_i \cdot (x_{ij} - \mu_i)$$

where  $N$  is the number of predictive proteins (6 in this case);  $j$  represents samples ( $j = 1, 2, \dots$ );  $i$  represents proteins with non-zero weights ( $i = 1, 2, \dots, 6$ );  $\omega_i$  is the weighting factor associated with the  $i^{\text{th}}$  protein;  $x_{ij}$  is the expression level of the  $i^{\text{th}}$  protein for the  $j^{\text{th}}$  patient; and

$\mu_i$  represents the average expression of the  $i^{\text{th}}$  protein across the entire cohort. The weighting factors reflect the contribution of proteins to the predictive model and are first initialized to their relative expressions (rx) [36] between the advanced stage and early stage groups:

$$\omega_i = \frac{rx_i}{\sum_{i=1}^N |rx_i|}$$

$$rx_i = \frac{\mu_{i,hi} - \mu_{i,lo}}{\sigma_{i,hi} + \sigma_{i,lo}}$$

where  $\mu_{i,hi}$ ,  $\sigma_{i,hi}$  ( $\mu_{i,lo}$ ,  $\sigma_{i,lo}$ ) represent the average expression and standard deviation of the  $i^{\text{th}}$  protein in the advanced stage (and early stage) groups, respectively.

To reduce the potential of overfitting [37], the initial weighting factors were next subjected to unconstrained nonlinear optimization using a derivative-free method [38]. The genetic algorithm was employed in Matlab, and refined using *fminsearch*. *fminsearch* starts at the initial weighting vector,  $x_0 = [\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6]$  and finds a local minimum of the custom-defined function described in *Fun(x)*.

$$\hat{x} = \text{fminsearch} [Fun(x), x_0]$$

$$Fun(x) = 1 - ROC(PSES_{hi}^{(x)}, PSES_{lo}^{(x)})$$

$$\hat{x} = [\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3, \hat{\omega}_4, \hat{\omega}_5, \hat{\omega}_6]$$

where  $PSES_{hi}^{(x)}$  and  $PSES_{lo}^{(x)}$  are the PSES scores for advanced and early stage tumors respectively and depend upon the weighting factor,  $x$ . ROC denotes the ROC curve analysis of these two scores; the corresponding value of the area under the curve (AUC) for each  $x$  was then calculated. Minimizing *Fun(x)* is equivalent to maximizing the predictive performance assessed by receiver operating characteristics (ROC) curve analysis. The  $\hat{x}$  represents the optimized weighting factor vector that gives rise to a maximum AUC value. The ROC curves evaluated via PSES scores were calculated as examples at the initialization and at the end of optimization (Supplementary Figure 11).

The PSES score was calculated for each of the 210 TCGA samples as the sum of the protein expression levels multiplied by the optimized weighting factors,

$$PSES_j = \sum_{i=1}^N \hat{\omega}_i \cdot (x_{ij} - \mu_i)$$

Specifically, PSES was defined as follows:  $0.3597 \times Dvl3 - 0.1874 \times JNK-pT183\_pT185 - 0.017 \times Shc\_pY317 - 0.1976 \times HER3\_pY1298 - 0.06 \times ER-alpha - 0.0678 \times PR$ , where protein expressions were mean centered. A plus or minus sign indicates that increased expression is associated with tumors with either advanced or early stage disease.

## Statistical analysis

Mann-Whitney analysis of variance was used to evaluate the statistical differences in protein levels and

PSES scores between early and advanced stage tumors. The Spearman correlation test was used to examine expression correlation among the three identified phosphorylated proteins. The Youden index [39] was used to determine the optimal PSES cutoff value in distinguishing patients with advanced and early stage disease. A multivariable logistic regression model was used to calculate the odds ratios (ORs) for clinicopathological variables associated with advanced stage tumors according to the PSES scores.

Survival analyses were performed using the Kaplan-Meier method, and the difference in survival was examined with the use of log-rank tests. Overall survival (OS) was defined as the interval from the date of initial surgical resection to the date of last known contact (censored) or death. Progression-free survival (PFS) was defined as the interval from the date of initial surgical resection to the date of progression, recurrence, or last known contact (censored). Statistical significance was defined as  $P < 0.05$ , and all tests were two-sided. Analyses were primarily performed using the scientific software Matlab, version 8.4 (MathWorks, Inc., Natick, MA, USA), SPSS version 18 (SPSS Inc., Chicago, IL, USA), and GraphPad Prism, version 6 (GraphPad Software, Inc., La Jolla, CA, USA).

## ACKNOWLEDGMENTS

We would like to acknowledge The Cancer Genome Atlas (TCGA) Research Network for archive of EEC datasets, and Jane Li for providing RPPA and clinical data for samples in the MD Anderson patient cohort. We thank Lee Ann Chastain in Department of Biostatistics and Sunita Patterson in Department of Scientific Publications at MD Anderson Cancer Center for editing this manuscript.

## CONFLICTS OF INTEREST

None.

## GRANT SUPPORT

This study was partially supported by grants from the National Institutes of Health/National Cancer Institute: a Career Development Award (to YL) from the MD Anderson Gynecologic SPORE in Uterine Cancers (2P50 CA098258-08, to KL and RB), the MD Anderson Institutional Research Grant (to YL), the MD Anderson TCGA Genome Data Analysis Center (U24 CA143883, to JW), and the MD Anderson Cancer Center Support Grant (P30 CA016672; used the Bioinformatics Shared Resource, Functional Proteomics Reverse Phase Protein Array [RPPA] Core, and Clinical Trials Support Resource). The study was also supported by the Cancer Prevention Research Institute of Texas (RP130397, to JW), the Mary K. Chapman Foundation, and the Michael & Susan Dell Foundation (honoring Lorraine Dell).

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016; 66:7–30. <https://doi.org/10.3322/caac.21332>.
2. Abu-Rustum NR, Zhou Q, Iasonos A, Alektiar KM, Leitao MM Jr, Chi DS, Sonoda Y, Soslow R, Hensley M, Barakat RR. The revised 2009 FIGO staging system for endometrial cancer: should the 1988 FIGO stages IA and IB be altered? *Int J Gynecol Cancer.* 2011; 21:511–16. <https://doi.org/10.1097/IGC.0b013e31820cc305>.
3. American Cancer Society. <http://www.cancer.org/cancer/endometrialcancer>. 2013.
4. Westin SN, Broaddus RR. Personalized therapy in endometrial cancer: challenges and opportunities. *Cancer Biol Ther.* 2012; 13:1–13. <https://doi.org/10.4161/cbt.13.1.18438>.
5. Salvesen HB, Carter SL, Mannelqvist M, Dutt A, Getz G, Stefansson IM, Raeder MB, Sos ML, Engelsen IB, Trovik J, Wik E, Greulich H, Bø TH, et al. Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. *Proc Natl Acad Sci USA.* 2009; 106:4834–39. <https://doi.org/10.1073/pnas.0806514106>.
6. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, et al, and Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. *Nature.* 2013; 497:67–73. <https://doi.org/10.1038/nature12113>.
7. Liu Y, Patel L, Mills GB, Lu KH, Sood AK, Ding L, Kucherlapati R, Mardis ER, Levine DA, Shmulevich I, Broaddus RR, Zhang W. Clinical significance of CTNNB1 mutation and Wnt pathway activation in endometrioid endometrial carcinoma. *J Natl Cancer Inst.* 2014; 106:dju245. <https://doi.org/10.1093/jnci/dju245>.
8. Gründker C, Günthert AR, Emons G. Hormonal heterogeneity of endometrial cancer. *Adv Exp Med Biol.* 2008; 630:166–88. [https://doi.org/10.1007/978-0-387-78818-0\\_11](https://doi.org/10.1007/978-0-387-78818-0_11).
9. Westin SN, Broaddus RR, Deng L, McCampbell A, Lu KH, Lacour RA, Milam MR, Urbauer DL, Mueller P, Pickar JH, Loose DS. Molecular clustering of endometrial carcinoma based on estrogen-induced gene expression. *Cancer Biol Ther.* 2009; 8:2126–35. <https://doi.org/10.4161/cbt.8.22.9740>.
10. Teng Y, Ai Z, Wang Y, Wang J, Luo L. Proteomic identification of PKM2 and HSPA5 as potential biomarkers for predicting high-risk endometrial carcinoma. *J Obstet Gynaecol Res.* 2013; 39:317–25. <https://doi.org/10.1111/j.1447-0756.2012.01970.x>.
11. Binder PS, Prat J, Mutch DG. Molecular staging of gynecological cancer: what is the future? *Best Pract Res Clin Obstet Gynaecol.* 2015; 29:776–89. <https://doi.org/10.1016/j.bpobgyn.2015.01.008>.
12. Nedelkov D, Kiernan UA, Niederkofler EE, Tubbs KA, Nelson RW. Population proteomics: the concept, attributes, and potential for cancer biomarker research. *Mol Cell Proteomics.* 2006; 5:1811–18. <https://doi.org/10.1074/mcp.R600006-MCP200>.
13. Tabchy A, Hennessy BT, Gonzalez-Angulo AM, Bernstam FM, Lu Y, Mills GB. Quantitative proteomic analysis in breast cancer. *Drugs Today (Barc).* 2011; 47:169–82. <https://doi.org/10.1358/dot.2011.47.2.1576695>.
14. Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, Kornblau SM. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Cancer Ther.* 2006; 5:2512–21. <https://doi.org/10.1158/1535-7163.MCT-06-0334>.
15. Tworoger SS, Zhang X, Eliassen AH, Qian J, Colditz GA, Willett WC, Rosner BA, Kraft P, Hankinson SE. Inclusion of endogenous hormone levels in risk prediction models of postmenopausal breast cancer. *J Clin Oncol.* 2014; 32:3111–16. <https://doi.org/10.1200/JCO.2014.56.1068>.
16. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, et al, and Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004; 32:D258–61. <https://doi.org/10.1093/nar/gkh036>.
17. Yang S, Raymond-Stintz MA, Ying W, Zhang J, Lidke DS, Steinberg SL, Williams L, Oliver JM, Wilson BS. Mapping ErbB receptors on breast cancer cell membranes during signal transduction. *J Cell Sci.* 2007; 120:2763–73. <https://doi.org/10.1242/jcs.007658>.
18. Hashimoto A, Kurosaki M, Gotoh N, Shibuya M, Kurosaki T. Shc regulates epidermal growth factor-induced activation of the JNK signaling pathway. *J Biol Chem.* 1999; 274:20139–43. <https://doi.org/10.1074/jbc.274.29.20139>.
19. Son JH, Kong TW, Kim SH, Paek J, Chang SJ, Lee EJ, Ryu HS. Prediction of lymph node metastasis in patients with apparent early endometrial cancer. *Obstet Gynecol Sci.* 2015; 58:385–90. <https://doi.org/10.5468/ogs.2015.58.5.385>.
20. Yang JY, Yoshihara K, Tanaka K, Hatae M, Masuzaki H, Itamochi H, Takano M, Ushijima K, Tanyi JL, Coukos G, Lu Y, Mills GB, Verhaak RG, and Cancer Genome Atlas (TCGA) Research Network. Predicting time to ovarian carcinoma recurrence using protein markers. *J Clin Invest.* 2013; 123:3740–50. <https://doi.org/10.1172/JCI74035>.
21. Helleman J, Smid M, Jansen MP, van der Burg ME, Berns EM. Pathway analysis of gene lists associated with platinum-based chemotherapy resistance in ovarian cancer: the big picture. *Gynecol Oncol.* 2010; 117:170–76. <https://doi.org/10.1016/j.ygyno.2010.01.010>.
22. Cai DT, Jin H, Xiong QX, Liu WG, Gao ZG, Gu GX, Qiu YH. ER stress and ASK1-JNK activation contribute to oridonin-induced apoptosis and growth inhibition in cultured human hepatoblastoma HuH-6 cells. *Mol Cell*

- Biochem. 2013; 379:161–69. <https://doi.org/10.1007/s11010-013-1638-2>.
23. Sau A, Filomeni G, Pezzola S, D'Aguanno S, Tregno FP, Urbani A, Serra M, Pasello M, Picci P, Federici G, Caccuri AM. Targeting GSTP1-1 induces JNK activation and leads to apoptosis in cisplatin-sensitive and -resistant human osteosarcoma cell lines. *Mol Biosyst.* 2012; 8:994–1006. <https://doi.org/10.1039/C1MB05295K>.
  24. Cellurale C, Girmius N, Jiang F, Cavanagh-Kyros J, Lu S, Garlick DS, Mercurio AM, Davis RJ. Role of JNK in mammary gland development and breast cancer. *Cancer Res.* 2012; 72:472–81. <https://doi.org/10.1158/0008-5472.CAN-11-1628>.
  25. Morrison MM, Hutchinson K, Williams MM, Stanford JC, Balko JM, Young C, Kuba MG, Sánchez V, Williams AJ, Hicks DJ, Arteaga CL, Prat A, Perou CM, et al. ErbB3 downregulation enhances luminal breast tumor response to antiestrogens. *J Clin Invest.* 2013; 123:4329–43. <https://doi.org/10.1172/JCI166764>.
  26. Li X, Xu Z, Du W, Zhang Z, Wei Y, Wang H, Zhu Z, Qin L, Wang L, Niu Q, Zhao X, Girard L, Gong Y, et al. Aiolos promotes anchorage independence by silencing p66Shc transcription in cancer cells. *Cancer Cell.* 2014; 25:575–89. <https://doi.org/10.1016/j.ccr.2014.03.020>.
  27. Yang JY, Werner HM, Li J, Westin SN, Lu Y, Halle MK, Trovik J, Salvesen HB, Mills GB, Liang H. Integrative protein-based prognostic model for early-stage endometrioid endometrial cancer. *Clin Cancer Res.* 2016; 22:513–23. <https://doi.org/10.1158/1078-0432.CCR-15-0104>.
  28. Kwan HT, Chan DW, Cai PC, Mak CS, Yung MM, Leung TH, Wong OG, Cheung AN, Ngan HY. AMPK activators suppress cervical cancer cell growth through inhibition of DVL3 mediated Wnt/ $\beta$ -catenin signaling activity. *PLoS One.* 2013; 8:e53597. <https://doi.org/10.1371/journal.pone.0053597>.
  29. Kafka A, Tomas D, Beroš V, Pečina HI, Zeljko M, Pečina-Šlaus N. Brain metastases from lung cancer show increased expression of DVL1, DVL3 and beta-catenin and down-regulation of E-cadherin. *Int J Mol Sci.* 2014; 15:10635–51. <https://doi.org/10.3390/ijms150610635>.
  30. Wang J, Qian J, Hoeksema MD, Zou Y, Espinosa AV, Rahman SM, Zhang B, Massion PP. Integrative genomics analysis identifies candidate drivers at 3q26-29 amplicon in squamous cell carcinoma of the lung. *Clin Cancer Res.* 2013; 19:5580–90. <https://doi.org/10.1158/1078-0432.CCR-13-0594>.
  31. Gao S, Bajrami I, Verrill C, Kigozi A, Ouaret D, Aleksic T, Asher R, Han C, Allen P, Bailey D, Feller S, Kashima T, Athanasou N, et al. Dsh homolog DVL3 mediates resistance to IGF1R inhibition by regulating IGF-RAS signaling. *Cancer Res.* 2014; 74:5866–77. <https://doi.org/10.1158/0008-5472.CAN-14-0806>.
  32. Hogberg T, Signorelli M, Oliveira CF, Fossati R, Lissoni AA, Sorbe B, Andersson H, Grenman S, Lundgren C, Rosenberg P, Boman K, Zola P, Kristensen G. Sequential adjuvant chemotherapy and radiotherapy in endometrial cancer - Results from two randomized studies. *Eur J Cancer.* 2010; 46:2422–31. <https://doi.org/10.1016/j.ejca.2010.06.002>.
  33. Carey MS, Agarwal R, Gilks B, Swenerton K, Kaloger S, Santos J, Ju Z, Lu Y, Zhang F, Coombes KR, Miller D, Huntsman D, Mills GB, Hennessy BT. Functional proteomic analysis of advanced serous ovarian cancer using reverse phase protein array: TGF-beta pathway signaling indicates response to primary chemotherapy. *Clin Cancer Res.* 2010; 16:2852–60. <https://doi.org/10.1158/1078-0432.CCR-09-2502>.
  34. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science.* 1999; 286:531–37. <https://doi.org/10.1126/science.286.5439.531>.
  35. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature.* 2002; 415:436–42. <https://doi.org/10.1038/415436a>.
  36. Liu Y, Sun Y, Broaddus R, Liu J, Sood AK, Shmulevich I, Zhang W. Integrated analysis of gene expression and tumor nuclear image profiles associated with chemotherapy response in serous ovarian carcinoma. *PLoS One.* 2012; 7:e36383. <https://doi.org/10.1371/journal.pone.0036383>.
  37. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004; 351:2817–26. <https://doi.org/10.1056/NEJMoa041588>.
  38. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J Optim.* 1998; 9:112–47. <https://doi.org/10.1137/S1052623496303470>.
  39. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J.* 2008; 50:419–30. <https://doi.org/10.1002/bimj.200710415>.