












6 Training, Validating, and Testing Machine Learning Prediction Models for Endometrial Cancer Recurrence

Jesus Gonzalez Bosquet, MD, PhD¹ ; Andrew Polio, MD¹ ; Erin George, MD² ; Ahmad A. Tarhini, MD, PhD³ ; Casey M. Cosgrove, MD⁴ ; Marilyn S. Huang, MD⁵ ; Bradley Corr, MD⁶ ; Aliza L. Leiser, MD⁷; Bodour Salhia, PhD⁸ ; Kathleen Darcy, MD^{9,10}; Christopher M. Tarney, MD⁹; Rob L. Dood, MD¹¹ ; Lauren E. Dockery, MD, MS¹²; Stephen B. Edge, MD¹³ ; Michael J. Cavnar, MD¹⁴; Lisa Landrum, MD, PhD¹⁵; Rob J. Rounbehler, PhD¹⁶; Michelle Churchman, PhD¹⁷; and Vincent M. Wagner, MD¹ 

DOI <https://doi.org/10.1200/PO-24-00859>

ABSTRACT




PURPOSE Endometrial cancer (EC) is the most common gynecologic cancer in the United States with rising incidence and mortality. Despite optimal treatment, 15%–20% of all patients will recur. To better select patients for adjuvant therapy, it is important to accurately predict patients at risk for recurrence. Our objective was to train, validate, and test models of EC recurrence using lasso regression and other machine learning (ML) and deep learning (DL) analytics in a large, comprehensive data set.

METHODS Data from patients with EC were downloaded from the Oncology Research Information Exchange Network database and stratified into low risk, The International Federation of Gynecology and Obstetrics (FIGO) grade 1 and 2, stage I (N = 329); high risk, or FIGO grade 3 or stages II, III, IV (N = 324); and nonendometrioid histology (N = 239) groups. Clinical, pathologic, genomic, and genetic data were used for the analysis. Genomic data included microRNA, long noncoding RNA, isoforms, and pseudogene expressions. Genetic variation included single-nucleotide variation (SNV) and copy-number variation (CNV). In the discovery phase, we selected variables informative for recurrence ($P < .05$), using univariate analyses of variance. Then, we trained, validated, and tested multivariate models using selected variables and lasso regression, MATLAB (ML), and TensorFlow (DL).

RESULTS Recurrence clinic models for low-risk, high-risk, and high-risk non-endometrioid histology had AUCs of 56%, 70%, and 65%, respectively. For training, we selected models with AUC >80%: five for the low-risk group, 20 models for the high-risk group, and 20 for the nonendometrioid group. The two best low-risk models included clinical data and CNVs. For the high-risk group, three of the five best-performing models included pseudogene expression. For the nonendometrioid group, pseudogene expression and SNV were overrepresented in the best models.

CONCLUSION Prediction models of EC recurrence built with ML and DL analytics had better performance than models with clinical and pathologic data alone. Prospective validation is required to determine clinical utility.

ACCOMPANYING CONTENT

-  [Data Sharing Statement](#)
-  [Data Supplement](#)
-  [Protocol](#)

Accepted March 26, 2025

Published May 5, 2025

JCO Precis Oncol 9:e2400859
© 2025 by American Society of
Clinical Oncology

Creative Commons Attribution
Non-Commercial No Derivatives
4.0 License

BACKGROUND

Endometrial cancer (EC) incidence and mortality continues to rise¹ despite advancements in adjuvant therapy over the past two decades, with a projected mortality increase of 55% by 2030.² In addition, important clinical trials have changed standards of treatment for low-risk and low-intermediate-risk EC (PORTEC-1 and GOG 99),^{3,4} high-intermediate-risk EC (PORTEC-2 and ASTEC),^{5,6} and high-risk EC (PORTEC-3, GOG 249, and GOG 258).^{7–10}

Furthermore, immunotherapy and targeted therapy have been introduced in advanced-stage and recurrent ECs with notable success (RUBY, GY-018, and DUO-E).^{11–13} Regardless of these advances, treatment failure occurs in approximately 10%–15% of patients with early-stage EC. Although nonendometrioid EC types account for a disproportionately high number of EC recurrences and cancer-related deaths,¹⁴ the majority of treatment failures and recurrences occur in endometrioid EC.^{14,15} Thus, identifying patients who might benefit from additional surveillance

CONTEXT

Key Objective

To build and test models of endometrial cancer recurrence integrated with clinical/pathologic risk.

Knowledge Generated

Models of recurrence integrating clinical and genomic data performed well in each risk group (low, high, and non-endometrioid) of the original database, Oncology Research Information Exchange Network, a network of US cancer centers. Testing in an independent database (The Cancer Genome Atlas) had some limitations and performed worse.

Relevance

Integrating genomic, clinical, and pathologic data improved performance of models for EC recurrence. Larger data sets with similar data are needed to externally validate these models.

and treatment to prevent recurrence and reduce mortality in EC would be of great value.

Historical studies included EC clinical and pathologic characteristics to stratify risk for recurrence and to inform adjuvant treatment.^{16,17} Since the publication of The Cancer Genome Atlas (TCGA) and description of specific molecular profiles for EC,¹⁸ efforts have been directed to stratify EC treatment on the basis of these four profiles¹⁹: (1) *POLE*-mut characterized by EC by mutation in DNA polymerase- ϵ ; (2) mismatch repair (MMR) deficiency with functional loss of the MMR proteins, resulting in microsatellite instability (MSI); (3) *TP53*-abnormal (*TP53abn*); and (4) no specific molecular profile.²⁰ The International Federation of Gynecology and Obstetrics (FIGO) new 2023 EC staging took consideration of this molecular classification and added *POLE*-mut as a characteristic of good prognosis in early-stage EC, and *TP53abn* as a sign of worse prognosis in early-stage EC.²¹ Models for EC recurrence integrating this new molecular classification and clinical data were superior to models with clinical data alone, with performances measured by the AUC over 70%²² versus below 70%,²³ respectively. Unfortunately, these models were not validated in independent data sets,²² nor their performances are ideal for being use in a clinical setting. Trials on the basis of those molecular-clinical models for treatment selection (RAINBO and PORTEC-4a) will be mature by the end of 2028.^{20,24} Therefore, there is room for improvement in EC recurrence prediction to better select patients for adjuvant treatment.

In a preliminary pilot study, we identified several prediction models for EC recurrence using integration of clinical and genomic data.²⁵ Genomic data included gene, exon, long noncoding RNA (lncRNA), and microRNA expression (MIR), single-nucleotide variation (SNV) and copy-number variation (CNV), and structural variation. The best performance model had an AUC of 90% (95% CI, 75% to 100%) and it was built with just five lncRNA.²⁵ However, validation of those models in an independent data set, the Oncology Research

Information Exchange Network (ORIEN), performed poorly, with an AUC of 57% (95% CI, 51% to 63%), likely because the initial model was constructed with a limited data set with only seven reported cases of EC recurrence. To improve the performance and accuracy of these models, we need a larger database of EC, with more diverse cases, and with larger representation of low-risk and high-risk EC endometrioid cases and nonendometrioid cases. The objective of this study is to train, validate, and test models of EC recurrence with lasso regression, other machine learning (ML), and deep learning (DL) analytics integrating clinical and genomic data from ORIEN, a large comprehensive EC database.^{26,27}

METHODS

Study Design

We performed a retrospective, multi-institution, case-control study with data originated from the ORIEN network EC data set. ORIEN is composed of multiple cancer centers that have agreed to use the same institutional review board-approved protocol and consent (Total Cancer Care Protocol) to follow patients throughout their lifetime.^{26,27} A copy of the protocol is included in the Data Supplement. Patients consent to donate medical records and tissue specimens for molecular profiling, as an approach to improve design and performance of personalized cancer care. RNA and DNA were extracted from tumor specimens and processed to obtain the necessary genomic data. The study analysis was carried out in two phases: (1) phase I: selection of variables and group of variables that were more informative for the outcome of interest, EC recurrence, using cross-validation and lasso regression; and (2) phase II: with the selected variables from phase I, we trained and validated models for EC recurrence using lasso regression, MATLAB classification learner app, and TensorFlow analytics. Finally, we tested these models, with lasso regression, MATLAB apps, and TensorFlow analytics in an independent EC data set, TCGA.

Patients' Inclusion and Clinical Data

All patients in the ORIEN database with EC, including all histologies that had information about recurrent disease. Patients with EC recurrence (or cases) were those that after completion of treatment with no evidence of disease (NED), EC reappeared, either locally (vaginal), regionally (pelvis), or distally. Index cases included women with a new event of EC cancer after treatment, those who had cancer at the last surveillance, or died from cancer. Controls were patients with NED during the whole follow-up. There were a total 892 women with EC included in this analysis with an average of follow-up of 31 months: 186 with EC recurrence (cases, average f/u of 28 months) and 706 without (controls, average of f/u of 31 months) who had RNA and DNA sequenced and had recurrence information. Clinical, pathologic, treatment, and molecular baseline characteristics of these patients are detailed in [Table 1](#) (14 clinical-pathologic) and the Data Supplement (Table S1; 28 laboratory values). Included patients were part of the ORIEN database since 2004 and up to 2021.

Patients with 2009 FIGO stage I and histologic grade 1 or 2 endometrioid EC had an overall recurrence rate of 11.6% (38/329) and were considered low risk for recurrence. Patients with a histologic grade 3 endometrioid EC or with FIGO stage II-IV had an overall recurrence rate of 21.3% (69/324) and were considered high risk for recurrence. Patients with nonendometrioid type EC (serous, carcinosarcoma, clear cell, undifferentiated, and mixed) had an overall recurrence rate of 33.1% (79/239) and were also considered high risk for recurrence. Given the different risks of recurrence for each group (different phenotype), we built a model of recurrence for each group.

Genomic Data

Data Preprocessing

Details about data preprocessing are found in the Data Supplement.

Genomic Variables

All genomic data were normalized and \log_2 -transformed before analysis, including number of SNVs, CNVs, and fusion transcripts. For analysis and modeling, we included gene, gene isoforms, MIR, lncRNA, pseudogene and fusion transcript expressions, as well as SNV and CNV data ([Table 2](#)).

Data Analysis and Modeling

Selection of Variables

Briefly, in the first phase of analysis, we selected those most informative variables for prediction of EC recurrence with analysis of variance ($P < .05$) and cross-validation with 10

replicates for each fold. Then, selected variables from the univariate analysis were incorporated into multivariate *lasso* regression prediction models of EC recurrence. Data types were progressively combined to create more complex prediction models. For details about selection of variables and training, validating, and testing of selection models, see the Data Supplement.

Training, Validating, and Testing EC Recurrence Models

Only variables included in models that were superior to an AUC >0.8 in phase I were brought forward to the second phase of analysis. Then, we trained, validated, and tested models including the selected variables from phase I, and used *lasso* regression, other ML included in *MATLAB* apps, and DL (*TensorFlow*) analytics.

Testing of Prediction Models in an External Data Set

Additionally, we used TCGA EC data set for external testing of the best prediction models for EC recurrence trained in the ORIEN set in phase II. The best prediction models of EC recurrence were tested with *lasso* regression, *MATLAB*, and *TensorFlow*, including TCGA data as the testing set.

Pathway Analysis

Pathway enrichment analysis for selected genes were performed in R environment with the package *clusterProfile*,^{28,29} which interrogates the Kyoto Encyclopedia of Genes and Genomes database to identify overrepresented pathways given a gene set.³⁰

RESULTS

Selection of Variables

Recurrence models with only clinical data for low-risk endometrioid, and high-risk endometrioid and non-endometrioid, had an AUC of 0.56, 0.81, and 0.65, respectively. For low-risk endometrioid EC, the only variable informative for recurrence was BMI, odds ratio (OR), 1.06 ([Table 2](#)). For high-risk endometrioid EC, there were six clinical variables that were included in the prediction model: FIGO stage (OR, 1.5), Hispanic ethnicity (OR, 1.5), radiation treatment after surgery (either brachytherapy or external-beam, OR, 0.69), and albumin (OR, 0.47), bilirubin (OR, 4.1) and RBC distribution width (RDW; OR, 0.96) values (Data Supplement, Table S1). For nonendometrioid type, FIGO stage (OR, 1.59), histologic type (serous, clear cell, carcinosarcoma, undifferentiated or dedifferentiated, or mixed, OR, 1.02), radiation treatment after surgery (either brachytherapy or external-beam, OR, 0.88), and albumin values (OR, 0.53) were in the clinical model.

Almost two thirds of MMR information was missing from the database ([Table 1](#)) and even more MSI testing, probably

TABLE 1. Patients' Baseline Characteristics

Characteristic	Low-Risk (FIGO stage I and grade 1 and 2) Endometrioid Type			High-Risk (FIGO stage II, III, IV or grade 3) Endometrioid Type			Nonendometrioid Type		
	Recurrent (n = 38)	Nonrecurrent (n = 291)	P	Recurrent (n = 69)	Nonrecurrent (n = 255)	P	Recurrent (n = 79)	Nonrecurrent (n = 160)	P
Age, years									
Average	61	61	.731	60	60	.973	66	64	.430
BMI									
Average	38.7	36.3	.224	33.6	35.5	.203	31.9	32.3	.800
Ethnicity			.412			.082 ^a			.746
Cuban	0	2		0	1		0	0	
Hispanic	1	9		6	10		4	9	
Mexican	0	2		3	5		2	2	
Non-Hispanic	36	260		60	230		72	145	
Puerto Rican	0	4		0	0		1	1	
Race			.885			.775			.986
American Native	0	0		3	0		0	0	
Asian	1	6		1	5		1	6	
Black	1	12		3	10		9	10	
Filipino	0	1		0	2		0	1	
Islander	1	0		0	0		0	0	
Other	0	6		0	6		0	3	
White	35	266		62	232		69	137	
Smoking			.132			.650			.036 ^a
Yes	7	89		41	156		50	97	
No	25	162		18	79		13	53	
Unknown	6	40		10	20		16	10	
Alcohol use			.217			.322			.945
Yes	12	119		29	103		33	77	
No	20	123		25	120		26	62	
Unknown	6	49		15	32		20	21	
Personal history									
Familial polyposis	0	1	.992	0	0	NA	0	0	NA
HPV	0	3	.991	2	4	.670	1	2	.949
Hyperplasia	3	31	.323	1	19	.053 ^a	2	8	.421
CIN	0	3	.991	0	0	NA	0	1	.987
Lynch syndrome	0	4	.992	2	3	.470	0	1	.987
Anemia	3	34	.509	16	37	.060 ^a	17	23	.063 ^a
COPD	3	11	.246	1	7	.540	2	8	.450

(continued on following page)

TABLE 1. Patients' Baseline Characteristics (continued)

Characteristic	Low-Risk (FIGO stage I and grade 1 and 2) Endometrioid Type			High-Risk (FIGO stage II, III, IV or grade 3) Endometrioid Type			Nonendometrioid Type		
	Recurrent (n = 38)	Nonrecurrent (n = 291)	P	Recurrent (n = 69)	Nonrecurrent (n = 255)	P	Recurrent (n = 79)	Nonrecurrent (n = 160)	P
DM	4	44	.928	12	40	.752	9	6	.018 ^a
Heart disease	0	1	.995	3	9	.701	6	10	.468
MI/CHF	4	36	.436	12	19	.016 ^a	8	20	.771
Stroke	3	29	.688	10	19	.078 ^a	8	20	.771
DVT	2	32	.296	16	27	.004 ^a	13	18	.105
PE	3	24	.370	13	20	.005 ^a	12	13	.035 ^a
Hyperlipidemia	4	61	.143	15	47	.433	11	36	.303
Hypertension	24	167	.480	42	139	.380	43	98	.852
Hypothyroidism	7	53	.928	6	39	.197	15	38	.827
Pain	11	67	.907	14	56	.888	28	37	.005 ^a
Grade			.705			.501			.989
1	17	188		5	18		0	2	
2	4	55		8	24		1	3	
3	NA	NA		7	16		24	41	
Undifferentiated	NA	NA		NA	NA		7	26	
FIGO stage									
I	38	291		4	70	Ref	19	71	Ref
II	NA	NA		11	35	.006 ^a	7	17	.406
II	NA	NA		27	78	.001 ^a	27	45	.023 ^a
IV	NA	NA		19	17	<.001 ^a	22	15	<.001 ^a
MMR			.581			1.000			.519
MMRp	11	72		10	47		11	30	
MMRd	9	45		10	47		14	28	
Unknown	18	174		49	161		54	102	
MI			.491			.157			.615
<50%	27	227		2	53		7	35	
≥50%	10	64		2	12		5	18	
Adjuvant radiation (any type)			.671			.199			<.001 ^a
No	27	217		42	133		54	71	
Yes	7	68		27	122		25	89	
Initial chemotherapy			.164			.161			.315
No	30	273		39	169		42	74	
Yes	4	16		29	85		37	86	

(continued on following page)

TABLE 1. Patients’ Baseline Characteristics (continued)

Characteristic	Low-Risk (FIGO stage I and grade 1 and 2) Endometrioid Type			High-Risk (FIGO stage II, III, IV or grade 3) Endometrioid Type			Nonendometrioid Type		
	Recurrent (n = 38)	Nonrecurrent (n = 291)	<i>P</i>	Recurrent (n = 69)	Nonrecurrent (n = 255)	<i>P</i>	Recurrent (n = 79)	Nonrecurrent (n = 160)	<i>P</i>
Histology									
Carcinoma							5	26	Ref
Carcinosarcoma							12	22	.085 ^a
Clear							7	9	.046 ^a
Mixed							11	52	.872
Mucinous							1	2	.469
Serosus							43	49	.004 ^a

NOTE. These are the baseline variables determined at treatment completion and included in the analysis.
Abbreviations: CIN, cervical intraepithelial neoplasia; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; DVT, deep vein thrombosis; FIGO, The International Federation of Gynecology and Obstetrics; HPV, human papillomavirus; MI/CHF, myocardial infarction/congestive heart failure; MI, myometrial invasion; MMRd, mismatch repair deficient; MMRp, mismatch repair proficient; NA, not available; PE, pulmonary embolism; Ref, reference variable.
^aStatistically significant with *P* value <.05.

TABLE 2. Variable Selection and Variables After Prediction Model Construction With Type of Data

Type of Data	Endometrioid			
	Baseline	Low Risk	High Risk	Nonendometrioid
Clinical ^a	42	1	6	4
SNV	19,239	800	1,257	1,461
CNV	23,445	5,311	7,052	4,234
Fusion	10,942	2,430	2,681	5,464
Gene expression	26,629	10,021	5,134	3,217
Isoforms expression	61,427	18,472	15,268	6,404
MIR ^a	1,881	—	—	—
LncRNA	16,849	4,915	4,033	1,285
Pseudogenes	15,250	5,247	3,053	1,171

NOTE. The baseline represents the initial number of variables for each type of data. After the selection with ANOVA (P value $<.05$), the most informative variables were kept for the multivariable lasso regression for all risks groups, low-risk and high-risk endometrioid EC, and nonendometrioid EC. Clinical baseline characteristics are detailed in Table 1 (14 clinical-pathologic) and the Data Supplement (Table S1; 28 laboratory values). The clinical selected variable for low-risk endometrioid EC was BMI; and the clinical selected variables for high-risk endometrioid EC were FIGO stage, Hispanic ethnicity, radiation treatment after surgery (either brachytherapy or external-beam), albumin, bilirubin and RDW; and the clinical selected variables for nonendometrioid EC were FIGO stage, histologic type (serous, clear cell, and carcinosarcoma), undifferentiated or dedifferentiated, or mixed, radiation treatment after surgery (either brachytherapy or external-beam, and albumin).

Abbreviations: ANOVA, analysis of variance; CNV, copy-number variation; EC, endometrial cancer; FIGO, The International Federation of Gynecology and Obstetrics; lncRNA, long noncoding RNA; MIR, microRNA expression; RDW, RBC distribution width.

^aLasso regression was performed directly with no preselection because of smaller number of variables in clinical data and MIR.

because more than a fourth of specimens were collected before the manuscript with TCGA data were published in 2013.¹⁸ Missing MMR and MSI information made it impractical to add them to the models. We assessed all SNVs of those genes included in TCGA classification: *POLE*, *TP53*, and MMR/MSI genes, *MLH1*, *MSH2*, *MSH6*, and *PMS2* (Data Supplement, Table S2), although none of the variants of these genes were present in any good performing model of EC recurrence for any risk level.

Initial models included only significant variables from each data type. Data types were later combined to create more complex models. For the training, validation, and testing phases, we only selected those variables included in models that had a performance ≥ 0.8 . Figure 1 details the composition of those models for the different risks' groups. These best models with combination of different data types were used in the second phase for training, validation, and testing of final prediction models for EC recurrence.

Training, Validation, and Testing Models for EC Recurrence

We built, validated, and tested models with the selected features from phase I. First, we trained and validated models (cross-validation) with selected variables using lasso regression. Additionally, we trained, validated, and tested models using selected variables in phase I from the ORIEN database in two analytical platforms: MATLAB (ML) and TensorFlow (DL; Table 3). In Table 3, we included only the best-performing model of the 35 possible results from MATLAB. For low-risk endometrioid EC, the only resulting clinical variable resulting from the analysis was BMI. For high-risk endometrioid EC, three of the five best-performing models included pseudogene expression. Some pseudogenes and SNVs were predominant in the best-performing models after validation and testing for non-endometrioid EC. Details of validation and training for some of these models are represented in the Data Supplement (Fig S3). Details about all variables included in best prediction models of EC for every risk group are detailed in the Data Supplement (Figs S4–S6 and Tables S7–S9). Pathway enrichment analysis results are also detailed in the Data Supplement (Fig S10).

External Testing of Models for EC Recurrence

We evaluated some of the best-performing models of endometrioid EC recurrence in TCGA data. TCGA endometrioid EC data set clinical characteristics that were found to be informative in the clinical model of recurrence for the ORIEN endometrioid data set are described in the Data Supplement (Table S11). There were some variables included in the high-risk endometrioid EC clinical model of ORIEN that were not available in TCGA, such as laboratory information, specifically albumin, bilirubin, and RDW values. The nonendometrioid data from ORIEN had more diverse histologic types than just serous cancers, so we considered that it was more problematic to evaluate those models in TCGA.

After downloading and preprocessing TCGA data set as we did with ORIEN, we selected those variables included in EC endometrioid best-performing models. Unfortunately, in half of the tested models, there were some missing variables in TCGA, therefore some modifications had to be made in the original ORIEN model to allow for testing. These modifications, or relearned models, were made on the Clin + PSE + ISO (clinical, pseudogene, and isoform expressions) model for the low-risk set, and the SNV, SNV + LNC + ISO (SNV, lncRNA, and isoforms expressions), MIR + PSE + mRNA (miRNA, pseudogene, and gene expressions), and Clin + PSE + FUS (clinical, pseudogene, and fusion transcript expressions) models for the high-risk set, and are marked on Table 4 with an asterisk. Testing of ORIEN models of endometrioid EC recurrence in TCGA data had good accuracy but poor AUCs (Table 4) for both analytical platforms (ML and DL). This is most likely due to the unbalanced data:

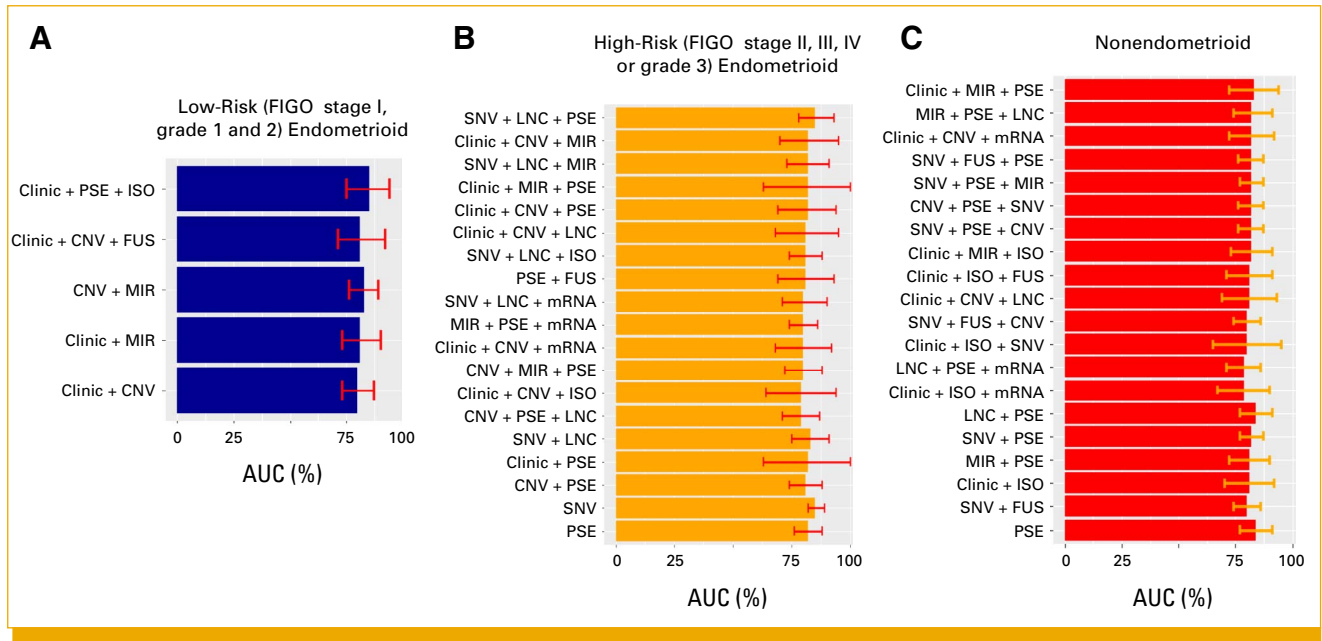


FIG 1. Selection of best models of EC recurrence after combination of data types. EC recurrence models for all risk groups with performances ≥ 0.8 measured by the AUC. The three panels represent risk-based groups: (A) Low-risk endometrioid EC best models (blue); (B) high-risk endometrioid EC best models (orange); and (C) nonendometrioid group best models (red). Different performances on all three panels are displayed in ascending order. The x-axis is AUC as a percentage (0%-100%). The red error mark displays the 95% CI. Overall, over 300 models with different combinations of datatypes were tested. We only displayed the best (A) five models for low-risk endometrioid EC, (B) 19 models for high-risk endometrioid EC, and (C) 20 for nonendometrioid EC. Genomic variation: CNV, copy-number variation; EC, endometrial cancer; SNV, single-nucleotide variation. Transcriptome: FUS, fusion transcript expression; ISO, gene isoform expression; LNC, long non-coding RNA expression; MIR, microRNA expression; mRNA, gene expression; PSE, pseudogene expression.

recurrences account for <10% of all samples in the low-risk group and <19% for the high-risk group. Details about some of these results are depicted in the Data Supplement (Fig S12). Additionally, other factors may have influenced the testing performance in TCGA EC data: (1) missing variables with respect to the original ORIEN model; (2) more reported recurrences in ORIEN versus TCGA in each risk group: low-risk: 12% versus 10%, high-risk: 21% versus 18%, respectively, although these differences were not statistically significant (chi-square P value $> .05$), underreported outcomes (recurrences) may have detrimental effects on prediction performances; and (3) 61% (252 of 411) of all EC TCGA samples were collected and processed before 2010, in comparison with only 7% (52 of 708) of all ORIEN EC samples.

DISCUSSION

In this study, we trained, validated, and tested models for EC recurrence stratified by risk factors. Risk factors were based on historical clinical-pathologic characteristics that were used in the past 40 years to determine adjuvant treatment for EC.^{3-5,8} The resulting prediction models were tested in a subset of the ORIEN database and were found to have excellent performances, on the basis of their AUC. ORIEN is one of the largest databases of EC clinical and genomic information maintained prospectively by a network of academic

institutions.^{26,27,31} Accordingly, clinical information and surveillance is optimized. Additionally, we tried to evaluate these EC recurrence models in TCGA clinical-genomic data. Unfortunately, not all data were available for testing, so some compromises had to be made. Also, there was a concern about recurrence reporting: recurrence in early-stage, low-risk EC is a rare event and missed reporting may result in misclassification. In higher-risk EC, recurrence is more frequent, but still there was less reporting of disease relapse in the TCGA data set than in ORIEN's network data. Furthermore, almost two thirds of specimen collection, processing, and analysis for TCGA was performed before 2010 with older technology and shorter reads (50mers v 100-150mers), in comparison with only 7% for ORIEN's, which may have affected overlapping sequence reading and counts for fusion transcripts, CNV, and other somatic structural variations.³² All these factors lead us to conclude that ORIEN-trained models tested in TCGA data may have had conflicting performances.

Models of EC prediction with only clinical-pathologic data performed as previous historical models.^{16,17,22,23} What was unique of our study is that we separated model building by EC risk group, so the prediction potential of some of the variables that could have been diluted in the whole data set showed prediction capabilities for individual risk groups. For example, for the low-risk endometrioid EC group, which is

TABLE 3. Validation and Testing of Best Prediction Models

Risk	Groups	Variables	Lasso		MATLAB		TensorFlow	
			Validation		Validation	Testing	Validation	Testing
			AUC	95% CI	AUC	AUC	AUC	AUC
Low-risk endometrioid	2	Clinic + CNV	0.97	0.95 to 0.99	0.87	0.98	0.99	0.91
	2	CNV + MIR	0.90	0.82 to 0.97	0.88	0.98	0.99	0.95
	3	Clinic + PSE + ISO	0.95	0.90 to 1.00	0.96	0.99	0.99	0.95
High-risk endometrioid	1	PSE	0.92	0.87 to 0.98	0.85	0.94	0.99	0.97
	1	SNV	0.97	0.94 to 0.99	0.88	0.95	0.99	0.96
	3	MIR + PSE + mRNA	0.96	0.94 to 0.99	0.90	0.91	1.00	0.93
	3	Clinic + PSE + fusion	0.95	0.88 to 1.02	0.91	1.00	0.94	0.92
	3	SNV + LNC + ISO	0.98	0.97 to 0.99	0.94	0.97	1.00	0.94
High-risk nonendometrioid	2	SNV + fusion	0.92	0.88 to 0.96	0.97	0.88	0.99	0.91
	2	MIR + PSE	0.95	0.93 to 0.98	0.92	0.95	1.00	0.98
	2	SNV + PSE	0.96	0.94 to 0.99	0.95	0.98	0.99	0.94
	3	Clinic + ISO + mRNA	0.98	0.97 to 1.00	0.91	1.00	0.99	1.00
	3	Clinic + ISO + SNV	0.93	0.87 to 0.99	0.95	0.96	1.00	1.00
	3	CNV + PSE + SNV	0.96	0.94 to 0.99	0.93	0.97	0.99	0.94
	3	SNV + PSE + MIR	0.96	0.94 to 0.99	0.93	1.00	1.00	0.97

NOTE. Validation of best models of EC recurrence on the basis of risk classification. The initial model was built and validated with cross-validation with a lasso regression in an R environment (left side of the table). Validation and testing were performed in two analytical platforms (right side of the table): MATLAB (ML) and TensorFlow (DL). The upper part of the table has patients with low-risk endometrioid EC: two of the best models include clinical data and CNVs. The only resulting variable for clinical data is BMI, other variables are not informative for recurrence in this risk group. The middle part of the table has patients with high-risk endometrioid EC: three of the five best-performing models include PSE. The lower part of the table has patients with nonendometrioid EC: PSE and SNV were overrepresented in best performance models.

Abbreviations: CNV, copy-number variation; DL, deep learning; EC, endometrial cancer; FUS, fusion transcript expression; ISO, gene isoform expression; LNC, long noncoding RNA expression; MIR, microRNA expression; ML, machine learning; mRNA, gene expression; PSE, pseudogene expression; SNV, single-nucleotide variation.

the most common group of all EC groups (58% in our database), grade and myometrial invasion did not show any prediction potential and only BMI performed fairly in predicting EC recurrence. The American Cancer Society just projected that EC will surpass ovarian cancer in mortality this year,² and that it seems to be driven by the increasing incidence of high-risk histologic subtypes accounting for a disproportionate number of EC deaths.³³ This has to be coupled with an obesity epidemic with links to cancer incidence and mortality.^{34,35} In clinical-pathologic models for higher-risk groups, both endometrioid and non-endometrioid, FIGO stage and radiation after surgery were predictors of recurrence: the higher the stage, the higher the risk for recurrence and radiation protected from recurrence. Notably, in high-risk endometrioid type, Hispanic ethnicity conferred higher risk for recurrence, but race was not a factor. Black women tend to have more nonendometrioid EC types and more mortality rate by EC than White women.³⁶ However, in our study, Black race did not confer more risk for recurrence in these more aggressive EC types, either endometrioid or nonendometrioid. This could be due to the relative low numbers of Black women included, only 5% of the total. Administration of initial chemotherapy was a predictor of recurrence, probably because the distribution of women who received chemotherapy was similar for each risk

group (Table 1). Other laboratory values were associated with high-risk endometrioid EC recurrence: increasing levels of albumin and RDW conferred less risk for recurrence, while elevated levels of bilirubin increased it. Likewise, increasing levels of albumin was protective of recurrence for non-endometrioid EC. Lower serum albumin levels (or hypoalbuminemia) has been considered a marker for illness severity and has been incorporated in several prognostic scores such as the Acute Physiology and Chronic Health Evaluation score, Child's classification in patients with liver cirrhosis, and the Glasgow Prognostic Score.³⁷ Additionally, hypoalbuminemia has been associated with poor prognosis in ovarian cancer³⁸ and EC.³⁷ Therefore, it is plausible that decreasing levels of albumin confer higher risk of disease recurrence in more aggressive types of EC, both endometrioid and nonendometrioid.

The integrated genomic characterization of EC by TCGA represented a shift in EC tumor classification.¹⁸ TCGA initial classification resulted from clustering results from gene copy number, whole exome sequencing (WES) of 24,8 tumor-normal pairs, MSI status, RNA expression, protein expression, and DNA methylation analyses. Clustering, a unsupervised learning algorithm, is a great method to identify underlying groups on the basis of the available

TABLE 4. External Testing of Best Prediction Models in TCGA Endometrioid EC Dataset

Risk	Groups	Variables	Lasso		MATLAB		TensorFlow	
			AUC (95% CI)		Testing		Testing	
			Validation	Testing	Accuracy	AUC	Accuracy	AUC
Low-risk endometrioid	2	Clinic + CNV	0.97 (0.95 to 0.99)	0.58 (0.43 to 0.72)	0.90	0.54	0.91	0.49
	2	CNV + MIR	0.90 (0.82 to 0.97)	0.52 (0.36 to 0.67)	0.90	0.54	0.86	0.42
	3	Clinic + PSE + ISO ^a	0.95 (0.90 to 1.00)	0.54 (0.37 to 0.71)	0.90	0.54	0.90	0.50
High-risk endometrioid	1	PSE	0.92 (0.87 to 0.98)	0.52 (0.42 to 0.62)	0.83	0.53	0.69	0.52
	1	SNV ^a	0.97 (0.94 to 0.99)	0.50 (0.50 to 0.50)	0.82	0.50	0.82	0.50
	3	MIR + PSE + mRNA ^a	0.93 (0.87 to 0.99)	0.53 (0.44 to 0.63)	0.73	0.54	0.18	0.50
	3	Clinic + PSE + FUS ^a	0.95 (0.88 to 1.02)	0.54 (0.45 to 0.64)	0.82	0.54	0.35	0.55
	3	SNV + LNC + ISO ^a	0.98 (0.97 to 0.99)	0.50 (0.50 to 0.50)	0.82	0.52	0.18	0.50

NOTE. Evaluation (testing) of best models of endometrioid EC recurrence in TCGA endometrioid EC data set by risk classification. The initial model was built and validated (cross-validation) with a lasso regression in an R environment on the ORIEN EC data set (left side of the table). This ORIEN model was validated in TCGA data (lasso testing). Additional testing was performed on two analytical platforms (right side of the table): MATLAB (ML) and TensorFlow (DL). Performance was measured in terms of AUC and accuracy. The upper part of the table has patients with low-risk endometrioid EC: two of the best models include clinical data and CNVs. The lower part of the table has patients with high-risk endometrioid EC: three of the five best-performing models include PSE.

Abbreviations: CNV, copy-number variation; DL, deep learning; EC, endometrial cancer; FUS, fusion transcript expression; ISO, gene isoform expression; LNC, long noncoding RNA expression; MIR, microRNA expression; ML, machine learning; mRNA, gene expression; PSE, pseudogene expression; SNV, single-nucleotide variation; TCGA, the cancer genome atlas.

^aWhen TCGA had not all the original clinical/genomic data, the ORIEN prediction models had to be relearned first with the data available in TCGA, and then were tested in TCGA data with different analytics platforms (lasso, MATLAB, and TensorFlow).

data, which is very useful when there is no previous knowledge about grouping.³⁹ One limitation of clustering algorithms is overlap between groups with similar data points even when they are of a different class.⁴⁰ Using methods available in clinical practice, investigators were able to refine TCGA four molecular subgroups with surrogate markers, p53 abnormalities, MSI, and *POLE* mutations, resulting in a classification tool.^{22,41} Models for EC recurrence created with these integrating clinical and molecular markers had an AUC around 0.7, without external validation.²² In our study, we trained, validated, and tested integrated models of EC recurrence with superior performance than those using TCGA molecular surrogates.

Previously described *POLE* pathologic somatic mutations (exons 9, 13, and 14)⁴² have low incidence in both recurrent and nonrecurrent ECs (Data Supplement, Table S2), with no statistical differences between both by EC risk groups. When we take all risk groups together, there were 34 somatic variants in nonrecurrent EC (of 672) and only two in recurrent cases (of 184), with a chi-square *P* value = .022. There were cases with recurrent EC (even in the low risk) that had *POLE* mutations. As larger data sets of SNV from WES or whole genome sequencing are available, we would be able to assess the real frequency of that genomic alteration in all EC groups and their association with recurrence. In our analysis, we did not find *POLE* somatic variation as a predictor for recurrence for any of the risk groups, including the low risk. Similarly, *TP53* variation was not predictor of EC recurrence on any of the groups, including the nonendometrioid, despite having significantly more recurrent cases with

mutations, other than p.P72R (Data Supplement, Table S2). Our interpretation is that *TP53* SNVs are so prevalent in nonendometrioid types (including serous) that they do not discriminate well which samples are at risk for recurrence. Neither variation of the genes involved in MMR was a predictor of EC recurrence in any group. All these results point to the fact that not all molecular characteristics that are associated with prognosis in EC are necessarily good classifiers or predictors of EC recurrence.

The best-performing models for low-risk endometrioid EC recurrence included altered CNV in some lncRNAs. Most of them were protective for disease relapse. In previous analyses, with less precise definition of outcomes' phenotypes, and smaller sample size, we also detected lncRNA as important variables in predicting EC recurrence.²⁵ In this study, we identified some ncRNAs with altered copy number that were part of the DNA repair mechanism and conferred protection for EC recurrence in low-risk EC. The association between DNA damage, DNA repair, and cancer is well known and is the basis for novel therapies, such as poly(ADP-ribose) polymerase inhibitors, checkpoints inhibitors, and even immunotherapies.⁴³ DNA damage response (DDR) coordinates DNA repair through a complex network of cellular pathways. Genes encoding DDR factors are frequently mutated in cancer, causing genomic instability.⁴⁴ In some cancers, such as colorectal cancer, ncRNAs have been associated with prognosis, cancer progression, or suppression.⁴⁵ For example, *LINC00905* has been associated with worse recurrence in cervical cancer,⁴⁶ *LINC00847* was associated with worse prognosis in pancreatic cancer,⁴⁶

ZNF674-AS1 may inhibit migration and invasion in lung cancer,⁴⁷ and TPRG1-AS1 inhibits liver cancer progression.⁴⁸ It seems that the last two effects on cancer progression were mediated by interactions with MIRs. Interactions between ncRNAs, DDR mechanisms, and disease relapse in low-risk EC must be elucidated before we can tap into their potential for treatment targeting.

High-risk endometrioid and nonendometrioid ECs had several common variables that were included in the best-performing prediction models of recurrence. The majority of those were pseudogenes but also there were two SNVs, SAMM50 and SELENOH, and the pathway analysis point out to an overrepresentation of the mitophagy machinery. Mitophagy is a specialized form of autophagy that plays a significant role in the occurrence and development of cancers.⁴⁹ In EC, mitophagy activity is closely associated with tumor cell metabolism, proliferation, survival, and resistance to treatment.⁵⁰ Additionally, pseudogene expression alone can accurately classify the major histologic subtypes of EC.⁵¹ Pseudogenes are evolutionary relics present in the genomes of a wide variety of species, and recent multiomics studies have determined that dysregulation of many pseudogenes is associated with relapse of disease in diverse cancer types.⁵²

One of the strengths of this study is that was performed on the ORIEN network EC data set, a prospectively collected database from US academic institutions with comprehensive clinical, pathologic, and genomic data. Despite the database prospective collection, this was an observational, case-control study, with limitations inherent to this type of design. For example, MMR and MSI status was partially available, thus not useful for modeling. However, we took advantage of the prospective nature of the data collection and outcome surveillance, including disease relapse. Additionally, all genomics analyses were performed uniformly, following better practices analytics and National Cancer

Institute analysis recommendations.⁵³ We grouped patients on the basis of classic characteristics of risk so, for the modeling training, we had homogeneous phenotypes of EC recurrence: low-risk, high-risk, and nonendometrioid groups. All models were trained with cross-validation and then were tested on samples that were left out of the initial training. Additionally, we did external validation of the best-performing models in an independent data set, TCGA. TCGA validation had some limitations because of potential disease relapse underreporting and surveillance shortcomings, the historical nature of the database collection that limited some analytics, and potential differences in genetic background between both populations, ORIEN and TCGA.⁵⁴

To avoid overfitting, we performed cross-validation in the discovery phase as well in the training of models and left out samples for further testing. However, the discovery phase and model training were performed in the same data, and that could lead to overfitting. To better evaluate the clinical value of these prediction models, we will need to perform prospective evaluation with independent EC data collected from collaborative institutions, like the ORIEN network. Other data could be included in these models in the future to improve their performance if external testing is disappointing. Artificial intelligence analysis of histopathology slides and their association with outcome prediction is evolving rapidly,⁵⁵ and there are some DL trained models with slides predicting outcomes that are promising.⁵⁶ Additionally, we could create multimodal models, integrating DL models from tabular and image data, to create more robust and better performing models for EC recurrence.

In summary, training, validating, and testing models of EC recurrence in a comprehensive database from the ORIEN network resulted in excellent performing models that, after prospectively evaluated, could help to assess which patients are at risk of relapse and are potential candidates for clinical trials.

AFFILIATIONS

¹Department of Obstetrics and Gynecology, Gynecologic Oncology, University of Iowa, Iowa City, IA

²Gynecologic Oncology, Moffit Cancer Center, Tampa, FL

³Departments of Cutaneous Oncology and Immunology, H. Lee Moffit Cancer Center & Research Institute, Tampa, FL

⁴Gynecologic Oncology, The Ohio State University, Columbus, OH

⁵Gynecologic Oncology, University of Virginia, Charlottesville, VA

⁶Gynecologic Oncology, University of Colorado, Aurora, CO

⁷Gynecologic Oncology, Rutgers, New Brunswick, NJ

⁸Department of Translational Genomics, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA

⁹Gynecologic Cancer Center of Excellence, Department of Gynecologic Surgery and Obstetrics, Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Bethesda, MD

¹⁰Murtha Cancer Center Research Program, Department of Surgery, Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Bethesda, MD

¹¹Gynecologic Oncology, Huntsman Cancer Institute, Salt Lake City, UT

¹²Gynecologic Oncology, University of Oklahoma, Oklahoma City, OK

¹³Surgical Oncology, Roswell Park Comprehensive Cancer Center, Elm & Carlton Streets, Buffalo, NY

¹⁴Surgical Oncology, University of Kentucky, UKMC—C224, Lexington KY

¹⁵Gynecologic Oncology, Indiana University, Indianapolis, IN

¹⁶Department of Clinical and Life Sciences, Aster Insights, Hudson, FL

¹⁷Aster Insights, Hudson, FL

CORRESPONDING AUTHOR

Jesus Gonzalez Bosquet, MD, PhD; e-mail: jesus-gonzalezbosquet@uiowa.edu

DISCLAIMER

The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views, assertions, opinions, or policies of the Uniformed Services University of the Health Sciences,

the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc, the Department of Defense, or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the US government.

SUPPORT

Supported in part by the NIH 5R01CA99908-18 (K. Leslie PI), and by the Research Fund of the Gynecologic Oncology Division of the University of Iowa Hospitals and Clinics, and supported in part by the American Association of Obstetricians and Gynecologists Foundation (AAOGF) Bridge Funding Award.

DATA SHARING STATEMENT

A data sharing statement provided by the authors is available with this article at DOI <https://doi.org/10.1200/PO-24-00859>.

AUTHOR CONTRIBUTIONS

Conception and design: Jesus Gonzalez Bosquet, Rob L. Dood, Vincent M. Wagner

Financial support: Jesus Gonzalez Bosquet

Administrative support: Jesus Gonzalez Bosquet, Michelle Churchman

Provision of study materials or patients: Erin George, Casey M.

Cosgrove, Kathleen Darcy, Lisa Landrum, Rob J. Rounbehler, Michelle Churchman

Collection and assembly of data: Erin George, Ahmad A. Tarhini, Casey M. Cosgrove, Bodour Salhia, Lauren E. Dockery, Stephen B. Edge, Lisa Landrum, Rob J. Rounbehler, Michelle Churchman

Data analysis and interpretation: Jesus Gonzalez Bosquet, Andrew Polio, Erin George, Ahmad A. Tarhini, Marilyn S. Huang, Bradley Corr, Aliza L. Leiser, Kathleen Darcy, Christopher M. Tarney, Rob L. Dood, Michael J. Cavnar

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information

about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](#)).

Erin George

Consulting or Advisory Role: Incyclix Bio

Research Funding: Merck Serono

Ahmad A. Tarhini

Consulting or Advisory Role: Bristol Myers Squibb, Merck, Genentech/Roche, Novartis, Sanofi/Regeneron, Partner Therapeutics, Clinigen Group, Eisai, Bayer, Instil Bio, ConcertAI, BioNTech, AstraZeneca, Nested

Research Funding: Bristol Myers Squibb (Inst), Merck (Inst), Genentech/Roche (Inst), OncoSec (Inst), Sanofi/Regeneron (Inst), Clinigen Group (Inst), InflaRx (Inst), Acrotech Biopharma (Inst), Pfizer (Inst), Agenus (Inst), Scholar Rock (Inst), Agenus (Inst)

Casey M. Cosgrove

Honoraria: UpToDate, GOG Foundation, Immunogen

Consulting or Advisory Role: GlaxoSmithKline, AstraZeneca, Imvax, Intuitive Surgical

Research Funding: GlaxoSmithKline, Regeneron

Marilyn S. Huang

Honoraria: Intersphere MJH

Consulting or Advisory Role: Tesaro, Seagen, Aptitude Health, Agenus, Cooper Surgical, touchIME, FLASCO, Eisai, Immunogen, Aspira Women's Health, Clovis Oncology, Curio Science, VBL Therapeutics, Swedish Cancer Center, Voluntas, Seagen, IntegrityCE, Elsevier, MJH Healthcare Holdings, LLC, Natera, Immunogen, Merck, AstraZeneca, Pfizer, AbbVie

Research Funding: Merck (Inst)

Bradley Corr

Consulting or Advisory Role: GlaxoSmithKline (Inst), Merck (Inst),

AstraZeneca/Merck (Inst), Immunogen, Imvax, Gilead Sciences (Inst), Corcept Therapeutics (Inst), Zentalis (Inst)

Research Funding: Clovis Oncology (Inst), Immunogen (Inst)

Bodour Salhia

Leadership: CpG Diagnostics

Stock and Other Ownership Interests: CpG Diagnostics Inc

Consulting or Advisory Role: AstraZeneca

Patents, Royalties, Other Intellectual Property: Patents filed and pending at University of Southern California

Travel, Accommodations, Expenses: CpG Diagnostics Inc

Stephen B. Edge

Honoraria: North American Center for Continuing Medical Education

Lisa Landrum

Consulting or Advisory Role: GlaxoSmithKline

No other potential conflicts of interest were reported.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, et al: Cancer statistics, 2022. *CA Cancer J Clin* 72:7-33, 2022
2. Sheikh MA, Althouse AD, Freese KE, et al: USA endometrial cancer projections to 2030: Should we be concerned? *Future Oncol* 10:2561-2568, 2014
3. Creutzberg CL, van Putten WL, Koper PC, et al: Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: Multicentre randomised trial. PORTEC Study Group. *Post Operative Radiation Therapy in Endometrial Carcinoma*. *Lancet* 355:1404-1411, 2000
4. Keys HM, Roberts JA, Brunetto VL, et al: A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: A Gynecologic Oncology Group study. *Gynecol Oncol* 92:744-751, 2004
5. Nout RA, Smit VT, Putter H, et al: Vaginal brachytherapy versus pelvic external beam radiotherapy for patients with endometrial cancer of high-intermediate risk (PORTEC-2): An open-label, non-inferiority, randomised trial. *Lancet* 375:816-823, 2010
6. Barton DP, Naik R, Herod J: Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): A randomized study. *Int J Gynecol Cancer* 19:1465, 2009
7. de Boer SM, Powell ME, Mileskin L, et al: Adjuvant chemoradiotherapy versus radiotherapy alone in women with high-risk endometrial cancer (PORTEC-3): Patterns of recurrence and post-hoc survival analysis of a randomised phase 3 trial. *Lancet Oncol* 20:1273-1285, 2019
8. Randall ME, Filiaci V, McMeekin DS, et al: Phase III trial: Adjuvant pelvic radiation therapy versus vaginal brachytherapy plus paclitaxel/carboplatin in high-intermediate and high-risk early stage endometrial cancer. *J Clin Oncol* 37:1810-1818, 2019
9. Matei D, Filiaci V, Randall ME, et al: Adjuvant chemotherapy plus radiation for locally advanced endometrial cancer. *N Engl J Med* 380:2317-2326, 2019
10. Simon R, Lam A, Li MC, et al: Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform* 3:11-17, 2007

11. Westin SN, Moore K, Chon HS, et al: Durvalumab plus carboplatin/paclitaxel followed by maintenance durvalumab with or without olaparib as first-line treatment for advanced endometrial cancer: The phase III DUO-E trial. *J Clin Oncol* 42:283-299, 2024
12. Eskander RN, Sill MW, Beffa L, et al: Pembrolizumab plus chemotherapy in advanced endometrial cancer. *N Engl J Med* 388:2159-2170, 2023
13. Mirza MR, Chase DM, Slomovitz BM, et al: Dostarlimab for primary advanced or recurrent endometrial cancer. *N Engl J Med* 388:2145-2158, 2023
14. Del Carmen MG, Boruta DM II, Schorge JO: Recurrent endometrial cancer. *Clin Obstet Gynecol* 54:266-277, 2011
15. Restaino S, Dinoi G, La Fera E, et al: Recurrent endometrial cancer: Which is the best treatment? Systematic Review of the literature. *Cancers (Basel)* 14:4176, 2022
16. Versluis MA, de Jong RA, Plat A, et al: Prediction model for regional or distant recurrence in endometrial cancer based on classical pathological and immunological parameters. *Br J Cancer* 113:786-793, 2015
17. Devor EJ, Miecznikowski J, Schickling BM, et al: Dysregulation of miR-181c expression influences recurrence of endometrial endometrioid adenocarcinoma by modulating NOTCH2 expression: An NRG Oncology/Gynecologic Oncology Group study. *Gynecol Oncol* 147:648-653, 2017
18. Cancer Genome Atlas Research N, Kandoth C, Schultz N, et al: Integrated genomic characterization of endometrial carcinoma. *Nature* 497:67-73, 2013
19. Walsh CS, Hacker KE, Secord AA, et al: Molecular testing for endometrial cancer: An SGO clinical practice statement. *Gynecol Oncol* 168:48-55, 2023
20. RAINBO Research Consortium: Refining adjuvant treatment in endometrial cancer based on molecular features: The RAINBO clinical trial program. *Int J Gynecol Cancer* 33:109-117, 2022
21. Berek JS, Matias-Guiu X, Creutzberg C, et al: FIGO staging of endometrial cancer: 2023. *Int J Gynaecol Obstet* 162:383-394, 2023
22. Stelloo E, Nout RA, Osse EM, et al: Improved risk assessment by integrating molecular and clinicopathological factors in early-stage endometrial cancer-combined analysis of the PORTEC cohorts. *Clin Cancer Res* 22:4215-4224, 2016
23. Creutzberg CL, van Stiphout RG, Nout RA, et al: Nomograms for prediction of outcome with or without adjuvant radiation therapy for patients with endometrial cancer: A pooled analysis of PORTEC-1 and PORTEC-2 trials. *Int J Radiat Oncol Biol Phys* 91:530-539, 2015
24. van den Heerik A, Horeweg N, Nout RA, et al: PORTEC-4a: International randomized trial of molecular profile-based adjuvant treatment for women with high-intermediate risk endometrial cancer. *Int J Gynecol Cancer* 30:2002-2007, 2020
25. Gonzalez-Bosquet J, Gabrilovich S, McDonald ME, et al: Integration of genomic and clinical retrospective data to predict endometrioid endometrial cancer recurrence. *Int J Mol Sci* 23:16014, 2022
26. ORIEN. <http://www.oriencancer.org>
27. Dalton WS, Sullivan D, Ecsedy J, et al: Patient enrichment for precision-based cancer clinical trials: Using prospective cohort surveillance as an approach to improve clinical trials. *Clin Pharmacol Ther* 104:23-26, 2018
28. Yu G, Wang LG, Han Y, et al: clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16:284-287, 2012
29. Xu S, Hu E, Cai Y, et al: Using clusterProfiler to characterize multiomics data. *Nat Protoc* 19:3292-3320, 2024
30. Kanehisa M, Furumichi M, Sato Y, et al: KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 51:D587-D592, 2023
31. Davis HA, Hoberg AA, Jacobus LS, et al: Leveraging oncology collaborative networks and biomedical informatics data resources to rapidly recruit and enroll rural residents into oncology quality of life clinical trials. *J Clin Transl Sci* 8:e135, 2024
32. Roberts HE, Lopopolo M, Pagnamenta AT, et al: Short and long-read genome sequencing methodologies for somatic variant detection; genomic analysis of a patient with diffuse large B-cell lymphoma. *Sci Rep* 11:6408, 2021
33. Eakin CM, Lai T, Cohen JG: Alarming trends and disparities in high-risk endometrial cancer. *Curr Opin Obstet Gynecol* 35:15-20, 2023
34. Steele CB, Thomas CC, Henley SJ, et al: Vital signs: Trends in incidence of cancers associated with overweight and obesity—United States, 2005-2014. *MMWR Morb Mortal Wkly Rep* 66:1052-1058, 2017
35. The Lancet Diabetes Endocrinology: The obesity-cancer link: Of increasing concern. *Lancet Diabetes Endocrinol* 8:175, 2020
36. Clarke MA, Devesa SS, Hammer A, et al: Racial and ethnic differences in hysterectomy-corrected uterine corpus cancer mortality by stage and histologic subtype. *JAMA Oncol* 8:895-903, 2022
37. Seebacher V, Grimm C, Reinthaller A, et al: The value of serum albumin as a novel independent marker for prognosis in patients with endometrial cancer. *Eur J Obstet Gynecol Reprod Biol* 171:101-106, 2013
38. Parker D, Bradley C, Bogle SM, et al: Serum albumin and CA125 are powerful predictors of survival in epithelial ovarian cancer. *Br J Obstet Gynaecol* 101:888-893, 1994
39. Hastie T, Tibshirani R, Friedman JH: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (ed 2). New York, NY, Springer, 2009
40. Nerurkar P, Shirke A, Chandane M, et al: Empirical analysis of data clustering algorithms. *Proced Comput Sci* 125:770-779, 2018
41. Talhouk A, McConechy MK, Leung S, et al: A clinically applicable molecular-based classification for endometrial cancers. *Br J Cancer* 113:299-310, 2015
42. Cancer Genome Atlas Research Network; Kandoth C, Schultz N, et al: Integrated genomic characterization of endometrial carcinoma. *Nature* 497:67-73, 2013
43. Pillie PG, Tang C, Mills GB, et al: State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat Rev Clin Oncol* 16:81-104, 2019
44. Groelly FJ, Fawkes M, Dagg RA, et al: Targeting DNA damage response pathways in cancer. *Nat Rev Cancer* 23:78-94, 2023
45. Marmol I, Sanchez-de-Diego C, Pradilla Dieste A, et al: Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. *Int J Mol Sci* 18:197, 2017
46. Zhang Y, Zhang X, Zhu H, et al: Identification of potential prognostic long non-coding RNA biomarkers for predicting recurrence in patients with cervical cancer. *Cancer Manag Res* 12:719-730, 2020
47. Wang J, Liu S, Pan T, et al: Long non-coding RNA ZNF674-AS1 regulates miR-23a/E-cadherin axis to suppress the migration and invasion of non-small cell lung cancer cells. *Transl Cancer Res* 10:4116-4124, 2021
48. Choi JH, Kwon SM, Moon SU, et al: TPRG1-AS1 induces RBM24 expression and inhibits liver cancer progression by sponging miR-4691-5p and miR-3659. *Liver Int* 41:2788-2800, 2021
49. Wei X, Xiong X, Wang P, et al: SIRT1-mediated deacetylation of FOXO3 enhances mitophagy and drives hormone resistance in endometrial cancer. *Mol Med* 30:147, 2024
50. Song C, Pan S, Zhang J, et al: Mitophagy: A novel perspective for insighting into cancer and cancer treatment. *Cell Prolif* 55:e13327, 2022
51. Han L, Yuan Y, Zheng S, et al: The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. *Nat Commun* 5:3963, 2014
52. Sun M, Wang Y, Zheng C, et al: Systematic functional interrogation of human pseudogenes using CRISPRi. *Genome Biol* 22:240, 2021
53. Zhang Z, Hernandez K, Savage J, et al: Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat Commun* 12:1226, 2021
54. Miller MD, Devor EJ, Salinas EA, et al: Population substructure has implications in validating next-generation cancer genomics studies with TCGA. *Int J Mol Sci* 20:1192, 2019
55. Mobadersany P, Yousefi S, Amgad M, et al: Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 115:E2970-E2979, 2018
56. Volinsky-Fremont S, Horeweg N, Andani S, et al: Prediction of recurrence risk in endometrial cancer with multimodal deep learning. *Nat Med* 30:1962-1973, 2024