Gynecologic Oncology 142 (2016) 150–157

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Stratification of endometrioid endometrial cancer patients into risk levels using somatic mutations

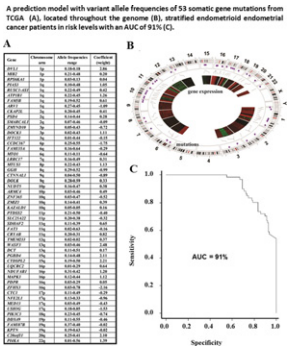
Donghai Dai [a], Kristina W. Thiel [a], Erin A. Salinas [a], Michael J. Goodheart [a,b], Kimberly K. Leslie [a,b], Jesus Gonzalez Bosquet [a,*]

[a] Department of Obstetrics and Gynecology, University of Iowa, Iowa City, IA, United States
[b] Holden Comprehensive Cancer Center, University of Iowa, Iowa City, IA, United States

HIGHLIGHTS

- We designed a prediction model to stratify endometrial cancer patients by risk levels using somatic mutations from TCGA.
- The prediction model including variant allele frequencies for each somatic gene mutation was superior to any other strategy.
- Stratifying patients accordingly to risk could individualize cancer treatment before and after surgery.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 2 March 2016
Received in revised form 9 May 2016
Accepted 11 May 2016
Available online 16 May 2016

Keywords:
Endometrial cancer
Somatic mutations
Prediction model
Risk levels
Individualized treatment

ABSTRACT

Objective. Patients with endometrioid endometrial cancer are stratified as high risk and low risk for extrauterine disease by surgical staging. Since patients with low-grade, minimally invasive disease do not benefit from comprehensive staging, pre-surgery stratification into a risk category may prevent unnecessary surgical staging in low risk patients. Our objective was to develop a predictive model to identify risk levels using somatic mutations that could be used preoperatively.

Methods. We classified endometrioid endometrial cancer patients in The Cancer Genome Atlas (TCGA) dataset into high risk and low risk categories: high risk patients presented with stage II, III or IV disease or stage I with high-intermediate risk features, whereas low risk patients consisted of the remaining stage I patients with either no myometrial invasion or low-intermediate risk features. Three strategies were used to build the prediction model: 1) mutational status for each gene; 2) number of somatic mutations for each gene; and 3) variant allele frequencies for each somatic mutation for each gene.

Results. Each prediction strategy had a good performance, with an area under the curve (or AUC) between 61% and 80%. Analysis of variant allele frequency produced a superior prediction model for risk levels of endometrial cancer as compared to the other two strategies, with an AUC = 91%. Lasso and Ridge methods identified 53 mutations that together had the highest predictability for high risk endometrioid endometrial cancer.

Conclusions. This prediction model will assist future retrospective and prospective studies to categorize endometrial cancer patients into high risk and low risk in the preoperative setting.

* Corresponding author at: Department of Obstetrics & Gynecology, University of Iowa Hospitals and Clinics, 200 Hawkins Dr., Iowa City, IA 52242, United States.
E-mail address: jesus-gonzalezbosquet@uiowa.edu (J. Gonzalez Bosquet).

http://dx.doi.org/10.1016/j.ygyno.2016.05.012
0090-8258/

1. Introduction

Endometrial cancer is the most common gynecological cancer, with over 54,000 new cases predicted for 2015 [1]. >70% of all endometrial tumors are at early stage (stage I) at the time of diagnosis [2]. Type I, or endometrioid endometrial (EEC), is the most common histologic type of endometrial cancer, and most are low grade and confined to the uterus. Surgical staging, including removing the uterus, cervix, adnexa, and pelvic and para-aortic lymph node tissues as well as obtaining pelvic washings, is the standard of care to accurately stage and assign patients into high risk and low risk categories [2–5], which can inform subsequent treatment [6]. Low risk patients have been defined as those with very low risk of having extrauterine disease and thus not requiring further treatment beyond a simple hysterectomy and removal of adnexa, preferably through a minimally invasive approach [7–9]. In comparison, high risk patients are at risk of having extrauterine disease, present worse prognosis and will likely require adjuvant treatment after surgery, including brachytherapy, external beam radiation or systemic chemotherapy.

Recent clinical trials have demonstrated that low risk patients with low-grade, minimally invasive disease do not clearly benefit from comprehensive staging. In fact, such patients will have excellent outcomes even with limited surgical intervention, hysterectomy and removal of the adnexa, with no further adjuvant therapy [4,10]. Moreover, the complex surgical procedures used for staging also increase complication rates and overall cost of care [11]. Thus, several groups have tried to identify predictors of disseminated disease to limit comprehensive surgical staging to those patients that will benefit (i.e., high risk patients) [12–15]. The Gynecologic Oncology Group (GOG) and other groups determined that histological grade and depth of myometrial invasion are associated with extension of disease outside the uterus [5,13,15]. However, the assessment of these variables may be limited by inaccurate and unreliable analysis performed on frozen specimens during the surgical procedure [16], highlighting the need for alternative approaches to stratify patients into risk categories.

Prediction models designed to assess high risk endometrial cancer have been constructed with information from surgical specimens [5, 15,17,18]. These models usually are excellent at detecting patients at low risk, with negative predictive values (NPV) close to 100%. However, they are not able to accurately detect patients as high risk, and their positive predictive value (PPV) is at most fair, around 20% in EEC [17,18]. Prediction models of risk using preoperative data have the same limitations [16,19]. In general, with these algorithms, it is necessary to perform 4 to 8 lymphadenectomies to find one patient with positive lymph nodes [19], and this surgery is not without extra costs and complications. Thus, to date, there is no preoperative predictive model that accurately identifies women with high risk EEC [16].

Our objective in this study was to create a test for risk using somatic mutations from whole exome next generation sequencing (NGS) that could be used preoperatively. Prediction models based on variant allele frequencies for somatic mutations were able to discriminate low versus high risk EEC with a mean area under the ROC (receiver operator characteristic) curve, or AUC, of 91%.

2. Materials and methods

2.1. Patients and data collection

Patients were selected from the Cancer Genome Atlas (TCGA) database of endometrial cancer. Patients with serous histology and other Type II endometrial cancer were excluded. Of those patients with Type I endometrial cancer or EEC, we only included patients who had undergone whole exome next generation sequencing with a somatic mutation report from TCGA (n = 190) [20]. Assuming that patients with myometrial invasion <50% (2009 FIGO stage IA) and histological grade 1 or 2 rarely would need lymph node assessment, 182 out of 190 patients that considered fully staged, or 96%. Mutations from these TCGA patients were downloaded from two different sources:

1. Level 2 and 3 of mutation analysis from the exomes of 190 ECC tumors were sequenced on Illumina GAIIx or HiSeq 2000 platforms (Illumina Inc., San Diego, CA). Somatic single variants and indels were called using software described elsewhere and were filtered for potential false positives [20]. The final list of somatic mutations in EEC is available online (https://tcga-data.nci.nih.gov/tcga/). A total of 177,057 somatic mutations were identified in the targeted exons of the 19,552 genes analyzed in EEC samples.
2. Original BAM files for each endometrial cancer patient were downloaded with permission from NCI from Cancer Genomics Hub (https://cghub.ucsc.edu). SAMtools [21] was then used to pile reads against Human Reference Genome 19. All reads for every synonymous (silent) and nonsynonymous mutation site were obtained. The Bcftools package in SAMtools was used to call variant sites from this subset of loci. Only sites for which one allele matched the reference genome were used in this analysis. Next, the number of reads for the reference and the strongest non-reference allele were calculated by summing the number of forward and reverse paired-end reads for each site. The strongest non-reference allele helped alleviate concerns of sequencing errors because each site required multiple reads per allele (>3) before being considered as a variant site. As an additional layer of stringency, only reads with Phred quality score of >30 (i.e. 99.9% accurate) were used in this study. The variant allele frequency (VAF) was calculated by the number of variant allele reads divided by the total number of reads over a genome locus. Thus, every patient was presented with a variant allele frequency for each somatic mutation.

2.2. Classification of EEC risk

Classification of EEC risk was based on the results and criteria from GOG 33 study, GOG 99 clinical trial and modified in the PORTEC trials [3,5,15,22]. "High risk" patients were defined as those at risk of having extrauterine disease and most likely needing any type of adjuvant treatment after surgery. Specifically, all patients presenting with stage II, III and IV as defined by 2009 FIGO classification (and sanctioned in 2014) [23], and patients with initial stage I and high-intermediate risk features by GOG 99 criteria [5] were classified as high risk. High-intermediate features of stage I included three risks factors: 2 or 3 tumor grade, presence of lymphovascular invasion, and outer-third myometrial invasion, with the following criteria: 1) at least 70 years of age with only one of the risk factors, 2) at least 50 years of age with any two of the other risk factors, or 3) any age with all three of the other risk factors. High risk patients are also at higher risk for disease recurrence and needing adjuvant treatment [6,24]. "Low risk" patients were the remaining stage I patients, either with no myometrial invasion or low-intermediate risk features by GOG 99 criteria [5]. There were 62 high risk and 128 low risk patients available for the study. A summary of clinical and pathological characteristics of both risks groups are shown in Supplemental Table S1. In the multivariable analysis of these features, there were significant differences between those variables that defined low and high risks groups: age, stage, myometrial invasion and histological grade. The stratification by risk was also independently associated with survival in TCGA patients with EEC (Supplementary Fig. S2). Five-year overall survival for low risk patients was 93% versus 82% for high risk patients.

2.3. Prediction model construction

Three strategies were used to build the prediction model: 1) mutational status; 2) the number of somatic mutations for each gene abstracted from the Level 2 dataset from TCGA; 3) variant allele