OXFORD

JNCI J Natl Cancer Inst (2015) 107(11): djv249

doi:10.1093/jnci/djv249
First published online September 1, 2015
Article

ARTICLE

Patterns of *CTCF* and *ZFHX3* Mutation and Associated Outcomes in Endometrial Cancer

Christopher J. Walker, Mario A. Miranda, Matthew J. O'Hern, Joseph P. McElroy, Kevin R. Coombes, Ralf Bundschuh, David E. Cohn, David G. Mutch, Paul J. Goodfellow

Affiliations of authors: Department of Obstetrics and Gynecology, Division of Gynecology Oncology (CJW, MAM, MJO, DEC, PJG), Department of Biomedical Informatics, Center for Biostatistics, College of Medicine (JPM, KRC), The Ohio State University Comprehensive Cancer Center - James Cancer Hospital and Solove Research Institute (CJW, MAM, MJO, JPM, KRC, DEC, PJG), Department of Physics, Department of Chemistry & Biochemistry, Department of Internal Medicine, Division of Hematology, Center for RNA Biology (RB), The Ohio State University, Columbus, OH; Department of Obstetrics and Gynecology, Division of Gynecology Oncology, Washington University, St. Louis, MO (DGM).

Correspondence to: Paul Goodfellow, PhD, Professor Obstetrics and Gynecology, 460W. 12th Avenue, BRT 808, Columbus, OH 43210 (e-mail: paul.goodfellow@osumc.edu).

Abstract

Background: The genetic events responsible for tumor aggressiveness in endometrioid endometrial cancer (EEC) remain poorly understood. The chromosome 16q22 tumor suppressor genes *CTCF* and *ZFHX3* are both frequently mutated in EEC, but their respective roles in outcome have not been determined.

Methods: Targeted deep sequencing of *CTCF* and *ZFHX3* was performed for 542 EEC samples. Copy number loss (CNL) was determined using microsatellite typing of paired tumor and normal DNA and a novel Bayesian method based on variant allele frequencies of germline polymorphisms. All statistical tests were two-sided.

Results: Mutation rates for *CTCF* and *ZFHX3* were 25.3% and 20.4%, respectively, and there was a statistically significant excess of tumors with mutation in both genes ($P = .003$). CNL rates were 17.4% for *CTCF* and 17.2% for *ZFHX3*, and the majority of CNLs included both *CTCF* and *ZFHX3*. Mutations were more frequent in tumors with microsatellite instability, and CNLs were more common in microsatellite-stable tumors ($P < .001$). Patients with *ZFHX3* mutation and/or CNL had higher-grade tumors ($P = .001$), were older ($P < .001$), and tended to have more frequent lymphovascular space invasion ($P = .07$). These patients had reduced recurrence-free and overall survival (RFS: hazard ratio [HR] = 2.35, 95% confidence interval [CI] = 1.38 to 3.99, $P = .007$; OS: HR = 1.51, 95% CI = 1.11 to 2.07, $P = .04$).

Conclusions: Our data demonstrate there is strong selection for inactivation of both *CTCF* and ZFHX3 in EEC. Mutation occurs at high frequency in microsatellite-unstable tumors, whereas CNLs are common in microsatellite-stable cancers. Loss of these two tumor suppressors is a frequent event in endometrial tumorigenesis, and ZFHX3 defects are associated with poor outcome.

It is estimated that more than 50 000 cases of endometrial cancer (EC) will be diagnosed in 2015 and over half of all EC deaths can be attributed to patients with endometrioid subtype tumors (1–3). Treating cases of recurrent or advanced-stage (not confined to the uterus) disease is clinically challenging. To improve outcomes in this population, three major questions must be addressed: How do we identify early-stage patients who will recur, can we alter treatment to prevent recurrence, and how do we treat recurrent/metastatic disease? Many studies, including The Cancer Genome Atlas (TCGA) for EC, have identified

Received: March 11, 2015; **Revised:** June 14, 2015; **Accepted:** August 5, 2015

commonly mutated genes in endometrioid endometrial carcinoma (EEC) (4–9). However, focused investigation is necessary to better understand the dysregulated molecular mechanisms responsible for tumorigenesis and disease maintenance, especially those associated with aggressive tumors.

CTCF is the major insulator protein in vertebrates and mediates long-range chromatin regulation (10–14). TCGA reports have shown that *CTCF* is specifically mutated at a high frequency in EEC tumors but not other malignancies (5,15). A recent report by Kemp et al. demonstrated that hemizygous *Ctcf* knockout is sufficient to globally dysregulate the methylome and predispose animals to tumor formation (16). This confirmed our earlier work implicating *CTCF* as a haploin-sufficient tumor suppressor gene (TSG) (17). *CTCF* maps to 16q22, a region of frequent deletion in a variety of malignancies (18–26). However, most 16q22 deletions are large and involve additional tumor suppressors including *ZFHX3* (also called *ATBF1*), *NQO1*, *CDH1*, and *E2F4* (reviewed in [27]). It remains uncertain which of these genes or combinations of genes confer selective advantage and which are coincidentally deleted.

ZFHX3 is frequently mutated in EEC (5) and studies in other tumor types suggest ZFHX3 acts as a tumor suppressor through multiple mechanisms (22,28–32). ZFHX3 is involved in the regulation of the estrogen receptor (ER) pathway through a complex feedback loop (33,34), and dysregulated ER signaling is the hallmark of EEC. Mouse models have shown that, like *Ctcf*, *Zfhx3* is a haploin-sufficient TSG and loss of *Zfhx3* causes neoplastic growth and differential expression of genes in the progesterone/estrogen signaling pathway (35,36).

Thus we sought to more closely investigate genetic alterations of *CTCF* and *ZFHX3* in EEC. We used targeted deep sequencing to detect mutation and copy number alterations in these two genes in a large cohort of patients with long-term follow-up. We also determined the clinicopathologic associations between *CTCF* and *ZFHX3* defects (alone and co-occurring) with features that portend poor outcome.

Methods

Study Population

The patient population and tumor specimens investigated have been previously described (17,37,38). Our cohort is representative of a typical EEC patient population, as presented in the Supplementary Methods and Supplementary Table 1 (available online). All 542 subjects studied were treated at Washington University School of Medicine, St. Louis, MO, and provided written, informed consent to Washington University Human Research Protection Office. All protocols were by the Washington University Human Research Protection Office (91–507 and 93–0828). All tumor DNAs were prepared from high neoplastic cellularity (>70%) flash-frozen tissues. Results for tumor microsatellite instability (MSI) testing and DNA Polymerase ε (*POLE*) proofreading domain mutation status have been previously reported (37,38). Seventeen MSI-low tumors were included with the microsatellite-stable (MSS) group. The mean and median follow-up for the cohort are 5.6 and 6.2 years, respectively.

Targeted Deep Sequencing

CTCF and *ZFHX3* sequencing was performed using TruSeq Custom Amplicon Assay (Illumina, San Diego, CA). Approximately 32 kb of *CTCF* and *ZFHX3* sequences was targeted using 79 425 bp amplicons (Illumina DesignStudio). Bar-coded patient samples were multiplexed and sequencing reactions performed with 250 base-paired end reads on an Illumina MiSeq. Variants were identified using Miseq Reporter software version 2.5.1 (Illumina, San Diego, CA) with the GATK variant caller (39) and filtered based on read depth and "qual," a phred-scaled quality score. Technical validation by Sanger sequencing of matched tumor and normal DNA established the accuracy of filtering and somatic origin of mutations. All calls were queried against dbSNP (build 137, National Center for Biotechnology Information) and COSMIC (40) databases. *CTCF* and *ZFHX3* variants are reported corresponding to the NM_006565.3 (Genbank: U25435.1) and NM_006885.3 (Genbank: NM_006885.3) transcripts, respectively.

Assessing Loss of Heterozygosity and Determining Copy Number Deletion

Fluorescence-based microsatellite typing was used to test for loss of heterozygosity (LOH) (41). Because of frequent strand slippage mutations in MSI tumors, microsatellite LOH analyses were limited to MSS tumors (n = 306). Seven highly informative dinucleotide repeats in the 16q22–23 region were evaluated (Supplementary Table 2, available online). Fragment sizing was performed using a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). Three hundred and one samples were informative for one or more marker tested.

We also applied a Bayesian method to determine allelic imbalances based on the variant allele frequencies (VAFs) of heterozygous polymorphisms detected by deep sequencing (personal communication Kevin R. Coombes). Sixty-three germline single nucleotide variants (SNVs) were detected with 396 cases heterozygous for one or more SNV. In brief, the model considers the observed VAFs relative to the percent tumor cellularity and predicts the probability of a deletion/amplification event based on deviation from the expected VAF of .50. For our cohort, the mean and median neoplastic cellularity were 78.5% and 80.0%, respectively. The observed number k of variant reads out of a total of n reads is modeled using a binomial distribution, Binom (k, n; φ), where the 'success' parameter φ depends on the copy number state (Deleted, Normal, or Gained) and the fraction, ν, of normal cells. This binomial model defines the likelihood function for the copy number state and normal fraction. We then place a discrete prior on the copy number state and a beta prior on the normal fraction. In this context, the posterior distribution can be computed by brute force by evaluating the likelihood at a grid of points over the interval (0,1) and applying Bayes' formula. Microsatellite typing and the Bayesian method showed high concordance (98%) and were performed by researchers blinded to clinical data.

Normalized read depth was calculated by jointly minimizing the logarithmic normalized coverage difference for the middle two quartiles of all amplicons for each sample pair, fitting a Gaussian distribution to this data for the middle two quartiles of samples for each amplicon, and reporting z-scores relative to this Gaussian distribution for each amplicon of each sample.

Statistical Analysis

Clinicopathologic variables were compared between indicated mutational subgroups using chi-squared or Fisher's exact tests for dichotomized variables. The continuous variables age and body mass index (BMI) were dichotomized at 60 years (42) and 30 kg/m^2 (43), respectively. Stage and grade were determined by FIGO 2009 classification (44), and stage was treated as a

ARTICLE

dichotomous variable (stage I/II vs stage III/IV). Overall survival (OS) and recurrence-free survival (RFS) were calculated using the log-rank test for univariate analyses. Cox proportional hazards regression was used for multivariate models. The proportional hazards assumption was examined by testing the association between the scaled Schoenfeld residuals and the Kaplan-Meier transformed survival times. For both multivariate models, none of the variables was statistically significant ($P > .05$), indicating a lack of evidence for departure from the proportional hazards assumption. OS was defined as the time from surgery to death. RFS was defined as the time from surgery to first recurrence. Patients with disease remaining after surgery were excluded from RFS analyses, and patients who died perioperatively were excluded from RFS and OS analyses. *P* values for associations with clinical features were adjusted using the False Discovery Rate (FDR) method (45) as noted, and FDR *P* values under .1 were considered statistically significant. *P* values for other comparisons are comparison-wise. All statistical tests were two-sided.

Results

Mutation and Allelic Deletion of *CTCF* and *ZFHX3* in EEC

Targeted deep sequencing of *CTCF* and *ZFHX3* was successfully completed for 538 of 542 EEC samples with an average read depth of 345X. The entire *CTCF* coding region was sequenced. Three regions of *ZHFX3* were not analyzed because of poor coverage, which code for amino acids 765–833, 2026–2171, and 3187–3703. Overall 79.0% of the *ZFHX3* coding sequence was covered. We undertook extensive technical validation of variant calls. Deep sequencing of 109 tumors that were previously Sanger sequenced for mutations in *CTCF* (17) correctly identified 20 different mutations in 34 tumors, and four additional nonclonal mutations were discovered (minor sequencing peaks were evident upon rereview of the original chromatograms). An additional 83 potential mutations in *CTCF* and *ZFHX3* within the rest of the cohort were analyzed using Sanger sequencing of matched tumor and normal DNA. Sixty-five calls were confirmed as somatic mutations, 16 were novel germline polymorphisms, and two were false-positives.

The overall rates of somatic mutation in *CTCF* and *ZFHX3* were 25.3% (87 different mutations in 136 tumors) and 20.4% (148 different mutations in 110 tumors), respectively. The vast majority of *CTCF*-mutated and most *ZFHX3*-mutated tumors had loss-of-function (LOF) mutations (Figure 1; details provided in Supplementary Table 3, available online). Twenty-nine tumors in our series have *POLE* mutations that have been associated with an ultramutated phenotype, and 24 (82.8%) had mutations in *CTFC* and/or *ZFHX3*. Because the *POLE*-mutated samples likely carry many passenger mutations, we considered this subset separately (Supplementary Table 3, available online).

CTCF and *ZFHX3* mutations were more frequent in MSI tumors than MSS tumors ($P < .001$ for either gene, two-sided Fisher's exact test). Among the 214 MSI tumors, 83 had *CTCF* mutations (38.8%) and 77 had *ZFHX3* mutations (36.0%), compared with 12.8% and 4.7% for MSS tumors, respectively. The *CTCF* p.T204fs hotspot mutation in an A_7 coding repeat (17) was seen in 49 MSI tumors, and two hotspot frameshift mutations in *ZFHX3*, p.E763fs and p.R1893fs*35/p.D1894fs*21, were observed exclusively in MSI tumors (Figure 1). Both ZFHX3 hotspots involved repeat regions: a G_7 coding repeat in exon 2 was mutated in eight patients, and an AG_5 repeat in exon 9 (not previously reported in EC) was mutated in eight patients (Supplementary Table 3, available online). Comutation of both *CTCF* and *ZFHX3* was seen in 33 specimens, which is statistically significantly greater than expected based on the individual mutation rate of each gene (22 expected, $P = .003$ two-sided Fisher's exact, excluding samples with allelic deletion outlined below).

In addition to mutations, we detected frequent LOH/allelic imbalance in the 16q22 region. Combined microsatellite typing and determining LOH using deep sequencing data (see Methods) gave an overall allelic loss rate of 18.1%. Representative findings for LOH detected by each method are presented in Figure 2. Analysis of normalized amplicon read depth revealed that allelic imbalance was typically associated with copy number loss (CNL) (Supplementary Figure 1, available online).

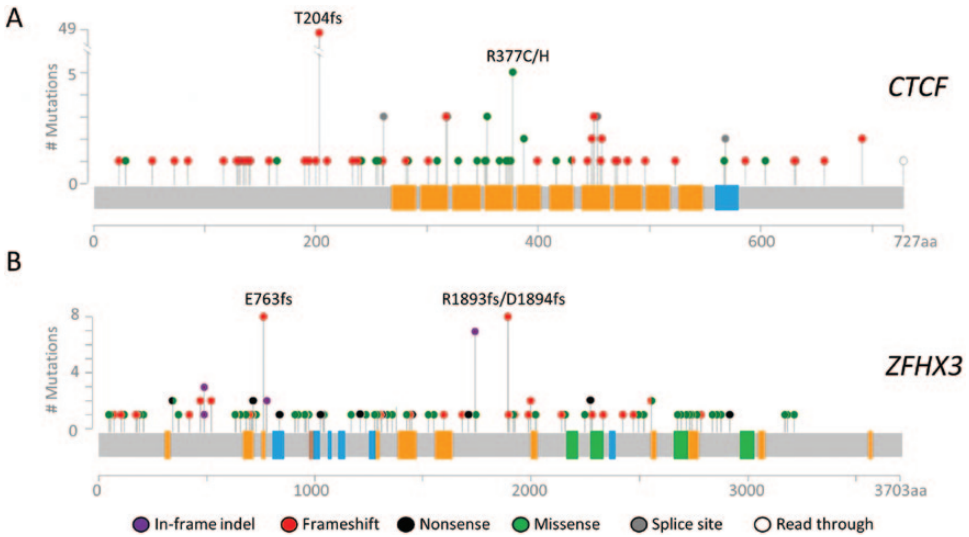

Figure 1. Mutations identified in 509 endometrioid endometrial carcinomas. **A) Lollipop plot** for *CTCF*. The hotspot mutations, p.T204fs and p.R377C/H, were seen in 49 and five tumors, respectively. **B) Lollipop plot** for *ZFHX3* shows that p.E763fs and p.R1893fs/p.D1894fs were each seen in eight tumors. Protein domains are indicated as follows: **orange**, typical C2H2 finger; **blue**, atypical C2H2 zinc fingers; **brown**, degenerate C2H2 zinc fingers; **green**, homeoboxes. Mutations in *POLE*-mutated tumors are not shown. A complete description of mutations can be found in Supplementary Table 3 (available online).

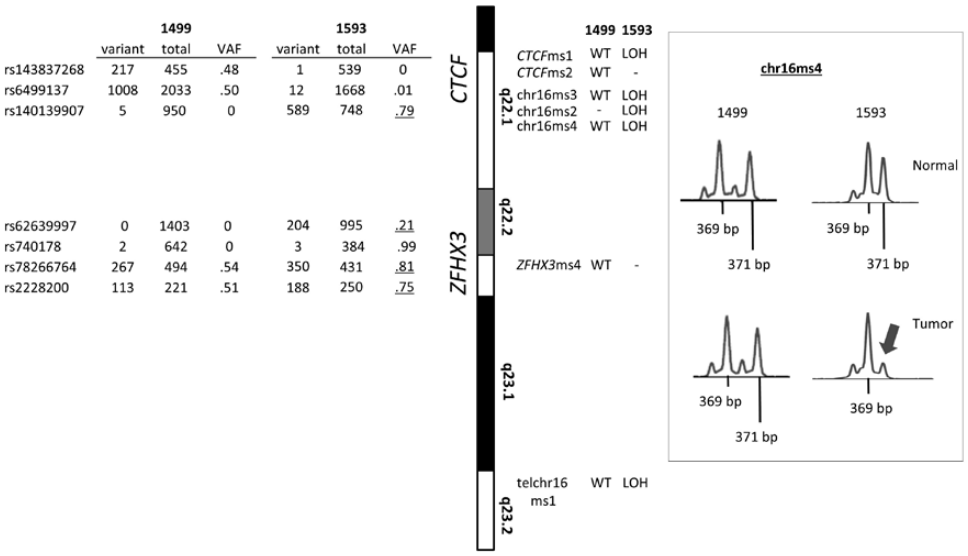

Figure 2. Representative example of allelic loss in the *CTCF*/*ZFHX3* region. **Left panel** shows variant allele frequencies (VAFs) for single-nucleotide polymorphisms (SNPs) in *CTCF* and *ZFHX3* in tumor 1593 (harboring allelic loss) and tumor 1499 (a typical two-copy tumor). Allelic imbalance is indicated by heterozygous SNPs with VAFs that deviate from the expected 0.50 (**underlined**). On the **right of the idiogram** for 16q region of interest are results for fluorescence-based microsatellite (MS) typing of paired tumor and normal DNA. MSs that revealed loss of heterozygosity (LOH) or showed no loss (WT) are given for both tumors. Uninformative markers indicated by "-". Traces for chr16ms4 typing are shown in the **right panel**. **Arrow** is used to emphasize the 371 bp allele. LOH = loss of heterozygosity; VAF = variant allele frequency; WT = wild-type.

Most instances of CNL included both *CTCF* and *ZFHX3* (16.6% of all tumors), with 1.5% showing loss of either gene alone (Supplementary Table 3, available online). CNL rates for *CTCF* and *ZFHX3* were 17.4% and 17.2%, respectively. CNL was statistically significantly more common in MSS than MSI tumors, with deletion in 68 and 19 case patients, respectively ($P < .001$ two-sided Fisher's exact test). A single instance of homozygous deletion was identified by greatly reduced normalized read depth within the *CTCF* coding sequence (Figure 3A). Long-range polymerase chain reaction (PCR) revealed a 1741 bp deletion spanning intron 2–3 (1157 bp) and exon 3 (594 bp), including the initiation codon (chr16:67,643,579-67,645,330) (Figure 3B). Homozygous deletion was confirmed by quantitative PCR of genomic DNA (data not shown).

Clinicopathologic Features Associated with *CTCF* and *ZFHX3* Mutation and Copy Number Status

The frequent codeletion and comutation of *CTCF* and *ZFHX3* observed (23.3%) suggests selection for inactivation of both genes. Patients with defects in both genes had higher-grade tumors (FDR $P = .001$), had more frequent lymphovascular space invasion (LVSI; FDR $P = .04$), and were older (>60 years, FDR $P = .04$) when compared with wild-type patients (Table 1). Tumors with defects in *CTCF* only and *ZFHX3* only were also seen at appreciable frequencies (16.5% and 9.2%, respectively) (Figure 4A). We present the relationship between clinicopathologic features and mutation and/or deletion stratified by the status of both genes (both genes defective, *CTCF* only defective, *ZFHX3* only defective, and both genes wild-type) in Supplementary Table 4 (available online). Given the differences between MSS and MSI tumors in mutation rates, the spectra of mutations, and the frequency of CNL, we also analyzed MSS and MSI patients separately. Compared with wild-type, MSS tumors with defects in both *CTCF* and *ZFHX3* were higher grade (FDR $P < .001$), and these patients were nonwhite (FDR $P = .06$). There were no statistically significant clinicopathologic associations in the MSI tumors (Supplementary Table 5, available online).

Women with tumors defective in both genes and those with defects in *ZFHX3* alone had reduced recurrence-free survival (Figure 4B). The hazard ratio (HR) was 2.29 for the group with both genes defective vs wild-type (95% confidence interval [CI] = 1.24 to 4.22, FDR $P = .03$) and was 2.32 for *ZFHX3*-only defective cases (95% CI = 1.02 to 5.23, FDR $P = .09$). OS was also statistically significantly reduced for case patients with defects in both genes (HR = 1.51, 95% CI = 1.06 to 2.15, FDR $P = .07$) (Supplementary Figure 2A, available online). In a multivariable model that included clinicopathologic factors statistically significantly associated with outcome (stage, grade, age, LVSI, and therapy) (38), the hazard ratio for mutation/CNL of both *CTCF* and *ZFHX3* on RFS was 1.54 (95% CI = .79 to 3.01), but the increase in risk did not reach statistical significance (Supplementary Table 6, available online). Early-stage (stage I/II) patients with mutation/CNL of both genes recurred statistically significantly more often than those without: There were 88 early-stage patients with defects in both genes and 13 recurred (14.8%), compared with only 12 recurrences in the 190 wild-type patients (6.3%) (odds ratio = 2.56, 95% CI = 1.03 to 6.46, FDR $P = .10$, two-sided Fisher's exact). Kaplan-Meier analysis for RFS confirmed this finding (Supplementary Figure 2B, available online).

The Kaplan-Meier plots presented in Figure 4B show patients with defects in *ZFHX3* only, and patients with both *CTCF* and *ZFHX3* defects have worse outcome compared with the *CTCF*-only group and wild-type patients, which strongly implicates *ZFHX3* as the factor contributing to reduced RFS. Therefore we compared cases with any *ZFHX3* defect to *ZFHX3* wild-type cases. Patients with *ZFHX3* defects had higher-grade tumors (FDR $P = .001$), had more frequent LVSI (FDR $P = .07$), and were older (FDR $P < .001$) (Supplementary Table 7, available online). These women had reduced RFS and OS (RFS: HR = 2.35, 95% CI = 1.38 to 3.99, FDR $P = .007$; OS: HR = 1.51, 95% CI = 1.11 to 2.07, FDR $P = .04$) (Figure 5, A and B). *ZFHX3* mutation remained a statistically significant factor associated with RFS in multivariable analysis ($P = .04$), and only tumor grade and stage were stronger predictors of recurrence in this model (Table 2).

ARTICLE

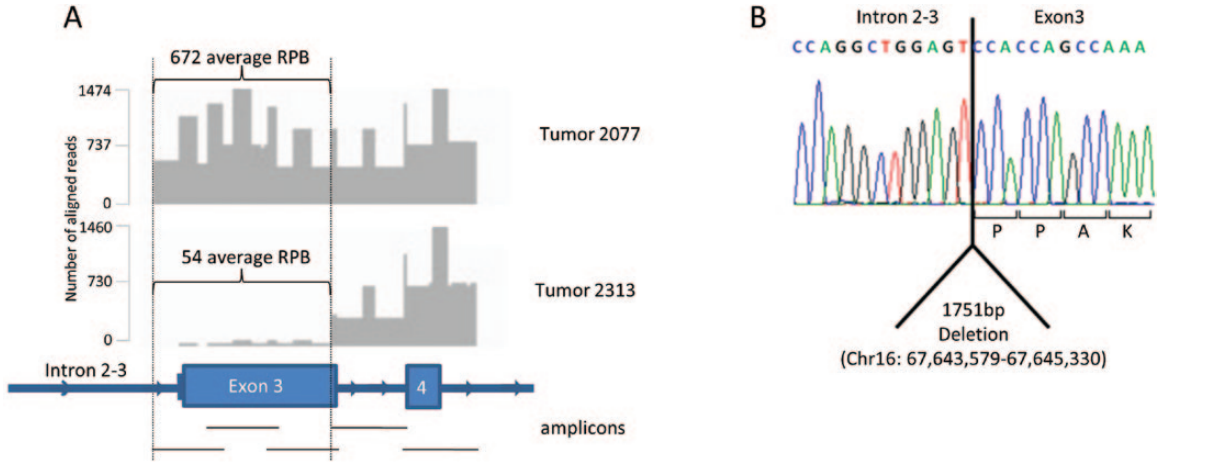

Figure 3. Homozygous deletion in *CTCF* in tumor 2313. **A)** Targeted sequencing of *CTCF* intron 2, exon 3, intron 3, and exon 4, in samples 2077 (representing a typical sample) and 2313. Five amplicons were sequenced using 250bp paired-end reads and are represented by **black lines**. **Y-axis** indicates number of reads across amplicons. **B)** Sequence trace of the breakpoint in sample 2313. RPB = reads per base.

Table 1. Clinicopathologic characteristics of patients with both *CTCF* and *ZFHX3* defects and women with no detectable abnormalities*

Clinicopathologic feature	*CTCF* and *ZFHX3* defects (23.3%) (n = 109) No. (%)	Wild-type (51.0%) (n = 238) No. (%)	P†
Grade			
1	43 (39.8)	139 (58.4)	.001
2	33 (30.6)	70 (29.4)	
3	32 (29.6)	29 (12.2)	
Age, y			
<60	32 (29.4)	105 (44.1)	.04
>60	77 (70.6)	133 (55.9)	
LVSI			
Present	46 (44.2)	70 (29.9)	.04
Absent	58 (55.8)	164 (70.1)	
Stage			
Early: I & II	90 (82.6)	194 (81.9)	ns
Advanced: III & IV	19 (17.4)	43 (18.1)	
Adjuvant therapy			
Yes	39 (36.1)	66 (27.7)	ns
No	69 (63.9)	172 (72.3)	
BMI, kg/m^2			
<30	35 (38.9)	71 (33.0)	ns
>30	55 (61.1)	144 (67.0)	
Race			
White	89 (81.7)	209 (87.8)	ns
Nonwhite	20 (18.3)	29 (12.2)	
Events‡			
Recur/progress	23 of 107 (21.5)	28 of 232 (12.1)	
Death	51 of 107 (47.7)	77 of 232 (33.2)	

* Defects include samples with mutation or loss of heterozygosity. Forty-six cases uninformative for mutation/loss of heterozygosity status and 29 cases with POLE mutations were excluded. BMI = body mass index; LVSI = lymphovascular space invasion; ns = not significant.
† Significance of grade was determined by chi-squared test. Significance of other variables was determined by Fisher's exact test. All *P* values are two-tailed and false discovery rate–adjusted.
‡ Excludes perioperative deaths.

We used TCGA for EC dataset as a validation cohort. RFS was compared between samples with mutation/CNL in *ZFHX3* (n = 45 EEC cases) and wild-type samples (n = 124 EEC cases). The RFS event rates are similar in our study and TCGA (12.6% and 13.0%). In TCGA, cohort patients with defects in *ZFHX3* had reduced RFS with a hazard ratio similar to our experimental cohort, but the difference did not reach statistical significance (HR = 2.00, 95% CI = .84 to 4.74, *P* = .12) (Figure 5C). However, considering the event rates, hazard ratios, number of individuals, and the proportion of individuals mutated in each cohort, TCGA cohort had only 40.2% power to detect the effect, whereas our experimental cohort achieved 83.6% power to detect the effect.

Discussion

Our data show *CTCF* and *ZFHX3* are among the most frequently altered genes in type I endometrial cancers (46): When both mutation and deletion are considered, 40% of EEC tumors have *CTCF* defects and 33% have abnormalities in *ZFHX3*. Although both *CTCF* and *ZFHX3* were classified as statistically significantly mutated genes by TCGA, co-occurring mutations, codeletion, and the relationship between gene defects and outcome were not addressed in the TCGA endometrial cancer report (5). Our combined sequencing and CNL analysis allowed us to fully interrogate the mutational status of both genes. Although it is well established that deep sequencing data can be used to detect copy number differences, accuracy depends on read depth (47,48). Mutation detection in coding sequence repeats can be challenging with exome capture sequencing methods. For tumor types such as EEC that frequently have defective DNA mismatch repair, the difficulty in sequencing coding repeats has likely resulted in underestimation of mutational frequency (17,49,50). The deep read depth we achieved and use of long (425 bp) amplicons made it possible to reliably detect both allelic imbalances/CNL and insertion/deletion mutations in coding repeats.

TCGA copy number alteration data (from GISTIC) (5) confirm deletion of *CTCF* and *ZFHX3* occurence in EEC at modest frequency and show that most of the deletions encompass large portions of 16q. TCGA exome sequencing data demonstrate that *CTCF* and *ZFHX3* are the only genes on chromosome 16q mutated at an appreciable frequency (>6%). Although many expressed genes map to the 16q deletion region we propose that selection for 16q deletion is driven by *ZFHX3* and *CTCF*, both of which have frequent LOF mutations in tumors that retain 16q heterozygosity.

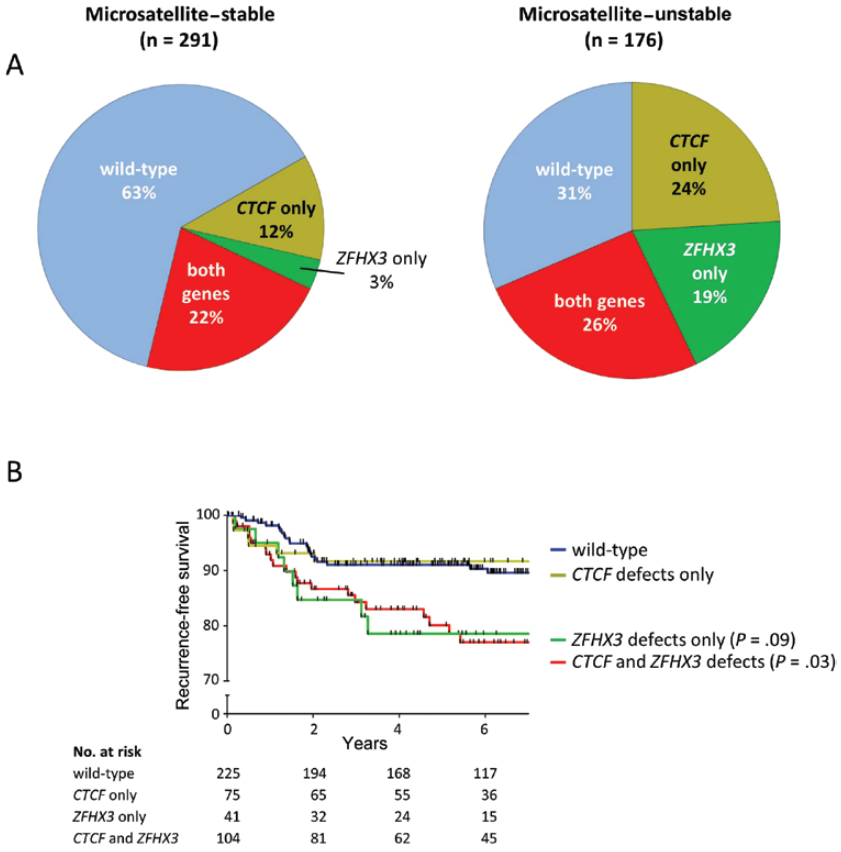

Figure 4. Patterns of *CTCF* and *ZFHX3* defects and associated outcomes. **A) Pie charts** show the percentage of samples harboring mutation and/or allelic loss of *CTCF* only, *ZFHX3* only, both genes and cases wild-type for both genes. Microsatellite-stable cases are shown on the **left**, and microsatellite-unstable cases are shown on the **right**. **B)** Kaplan-Meier estimates for recurrence-free survival. *P* values and hazard ratios for each group were calculated by Cox proportional hazard regression and false discovery rate adjusted.

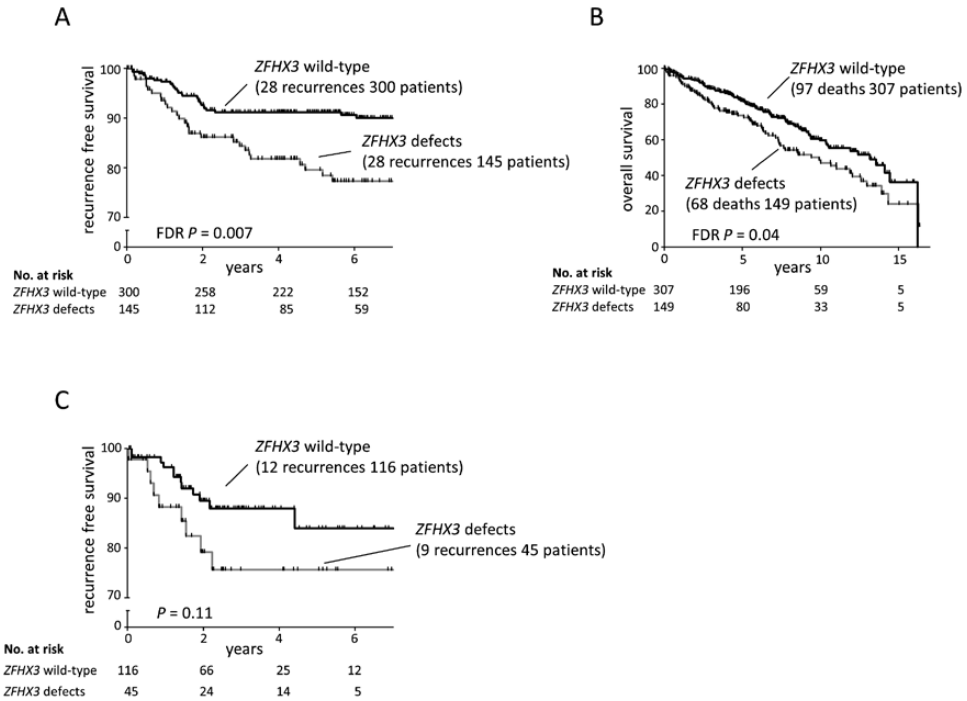

Figure 5. *ZFHX3* defects and association with survival. Kaplan-Meier estimates for recurrence-free **(A)** and overall survival **(B)** in our experimental cohort. **C)** Kaplan-Meier estimates for recurrence-free survival of endometrioid endometrial cancer patients analyzed by The Cancer Genome Atlas. *P* values obtained by log-rank test for univariate analyses and false discovery rate (FDR) adjusted as indicated.

Table 2. Multivariable model for outcome with *ZFHX3* mutation/CNL*

	RFS		OS	
Variable	HR (95% CI)	P†	HR (95% CI)	P†
Grade 3	5.17 (2.25 to 11.89)	<.001	3.21 (2.04 to 5.07)	<.001
Stage III/IV	3.30 (1.56 to 6.99)	.002	4.43 (2.65 to 7.41)	<.001
Grade 2	2.94 (1.38 to 6.27)	.005	1.60 (1.11 to 2.33)	.01
ZFHX3 mutation/CNL	1.82 (1.03 to 3.20)	.04	1.10 (0.79 to 1.54)	.57
Presence of LVSI	1.48 (0.77 to 2.83)	.24	1.55 (1.08 to 2.24)	.02
Use of adjuvant therapy	1.19 (0.56 to 2.51)	.65	0.57 (0.36 to 0.90)	.02
Age < 60 y	0.97 (0.53 to 1.77)	.92	0.41 (0.28 to 0.61)	<.001

* Model includes all clinicopathologic features listed and mutation and/or copy number loss of *ZFHX3* as variables. Reference groups are as follows: for grade 2 and grade 3, grade 1; for stage, stage I/II, for *ZFHX3* and *CTCF* defects, *ZFHX3* and *CTCF* wild-type; for presence of LVSI, absence of LVSI, for age < 60 years, age > 60 years; for use of adjuvant therapy, no adjuvant therapy. CI = confidence interval; CNL = copy number loss; HR = hazard ratio; LVSI = lymphovascular space invasion; OS = overall survival; RFS = recurrence-free survival.
† *P* values calculated using multivariable Cox proportional hazard tests. *P* values are two-sided.

Endometrial cancers are often categorized as having either an MSI or chromosomal instability (CIN) mutator phenotype (51–53). The genetic inactivation of *CTCF* and *ZFHX3* we observed in cancers with normal DNA mismatch repair (MSS tumors with CIN) and those with MSI implicates strong selection in EEC. MSS tumors frequently have chromosomal deletions involving the region containing *CTCF* and *ZFHX3*, whereas MSI tumors often harbor SNV/indels in both genes, including hotspot mutations in coding sequence repeats. Examination of the VAFs of these hotspot mutations showed that the vast majority are clonal (variant frequency > .3) and are, as such, unlikely to reflect mutational noise secondary to defective mismatch repair (data not shown). Three-fourths of *CTCF* mutations are LOF (frameshift, nonsense, and splice site) and, as we have previously demonstrated, subjected to nonsense mediated decay (17). Examination of the *CTCF* missense mutations shows that they cluster around the zinc finger domains, and condel (54) predictions indicate most are deleterious to protein function (Figure 1; Supplemental Table 3, available online). Although the evidence for LOF of *ZFHX3* is not as compelling, half of all mutations are LOF, and half of the missense mutations are predicted to be deleterious (Supplementary Table 3, available online). Thus we propose that allelic loss and mutations in EEC both lead to haploin sufficiency for CTCF and ZFHX3.

Recent studies have challenged the Knudsen two-hit model (55) for tumor suppressors and suggest that hemizygous deletions play critical roles in shaping the cancer genome (56,57). Our previous report on *CTCF* defects in EC strongly pointed to its function as a haploin-sufficient TSG, which has been confirmed in vivo (16,17). Knock-out mouse modeling has also implicated ZFHX3 as a haploin-sufficient TSG (35). Here, we found that only 9.6% of *CTCF*-mutated tumors and 19.1% of *ZFHX3*-mutated tumors harbor a second hit, further supporting haploin sufficiency. Although one specimen in our cohort has homozygous deletion of *CTCF*, the alternative initiation codon for isoform 2 is retained, which in principle could compensate for the loss of isoform 1.

Importantly, our data show that women with tumors defective in *ZFHX3* alone (no evidence for mutation in *CTCF*) have statistically significantly reduced RFS indistinguishable from that seen in cases with defects in both genes (Figure 4B). Although *CTCF* and *ZFHX3* are commonly inactivated together, our findings point to *ZFHX3* mutation as the key factor associated with poor outcome.

The correlative nature of the work described here is an obvious limitation of the study and inherent to many biomarker and molecular discovery efforts. In vitro and in vivo studies to functionalize the genetic inactivation of these two genes are the necessary next steps for better understanding the roles *ZFHX3* and *CTCF* play in endometrial tumorigenesis.

Overall, the frequency and spectra of mutations and deletions we identified in the closely linked tumor suppressors *CTCF* and *ZFHX3* suggests strong selection for inactivation of both genes in EEC. The statistically significantly reduced RFS for women with defects in *ZFHX3* highlights its underappreciated importance in EEC, specifically in biologic/clinical aggressiveness.

Funding

This work was supported by The National Institutes of Health (R21 CA155674 to PJG), the National Cancer Institute (P30 CA016058) supporting the Nucleic Acid and Biostatistics shared resources at the Ohio State University Comprehensive Cancer Center, and the Pelotonia Fellowship Program (CJW).

Notes

The sponsors had no role in the design of the study, the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication.

We would like to thank Alexis Chassen for manuscript editing and data curation, Sandya Liyanarachchi for assistance with data analysis, and Nima Esmaili Mokaram for assistance with computational methods. We would like to acknowledge Pearlly Yan and the Ohio State University Genomics Core Facility, and Tea Meulia and the Ohio State University Molecular and Cellular Imagine Center, a CFAES/OARDX core facility, in Wooster, OH.

We are very grateful to all of the patients who contributed specimens to this study and all of the attending physicians and staff at the Washington University School of Medicine Division of Gynecologic Oncology.

All authors declare they have no conflict of interest.

References

1. Hamilton CA, Cheung MK, Osann K, et al. Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br J Cancer*. 2006;94(5):642–646.

2. Hamilton CA, Kapp DS, Chan JK. Clinical aspects of uterine papillary serous carcinoma. *Curr Opin Obstet Gynecol*. 2008;20(1):26–33.
3. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9–29.
4. Berg A, Hoivik EA, Mjos S, et al. Molecular profiling of endometrial carcinoma precursor, primary and metastatic lesions suggests different targets for treatment in obese compared to non-obese patients. *Oncotarget*. 2015;6(2):1327–1339.
5. Cancer Genome Atlas Research N, Kandoth C, Schultz N, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497(7447):67–73.
6. Fukuchi T, Sakamoto M, Tsuda H, et al. Beta-catenin mutation in carcinoma of the uterine endometrium. *Cancer Res*. 1998;58(16):3526–3528.
7. Kohler MF, Berkholz A, Risinger JI, et al. Mutational analysis of the estrogen-receptor gene in endometrial carcinoma. *Obstet Gynecol*. 1995;86(1):33–37.
8. Koul A, Willen R, Bendahl PO, et al. Distinct sets of gene alterations in endometrial carcinoma implicate alternate modes of tumorigenesis. *Cancer*. 2002;94(9):2369–2379.
9. Tashiro H, Blazes MS, Wu R, et al. Mutations in PTEN are frequent in endometrial carcinoma but rare in other common gynecological malignancies. *Cancer Res*. 1997;57(18):3935–3940.
10. Xiao T, Wallace J, Felsenfeld G. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*. 2011;31(11):2174–2183.
11. Rubio ED, Reiss DJ, Welcsh PL, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*. 2008;105(24):8309–8314.
12. Parelho V, Hadjur S, Spivakov M, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*. 2008;132(3):422–433.
13. Stedman W, Kang H, Lin S, et al. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J*. 2008;27(4):654–666.
14. Wendt KS, Yoshida K, Itoh T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*. 2008;451(7180):796–801.
15. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113–1120.
16. Kemp CJ, Moore JM, Moser R, et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep*. 2014;7(4):1020–1029.
17. Zighelboim I, Mutch DG, Knapp A, et al. High frequency strand slippage mutations in CTCF in MSI-positive endometrial cancers. *Hum Mutat*. 2014;35(1):63–65.
18. Kondo K, Chilcote RR, Maurer HS, et al. Chromosome abnormalities in tumor cells from patients with sporadic Wilms' tumor. *Cancer Res*. 1984;44(11):5376–5381.
19. Maw MA, Grundy PE, Millow LJ, et al. A third Wilms' tumor locus on chromosome 16q. *Cancer Res*. 1992;52(11):3094–3098.
20. Devilee P, van Vliet M, van Sloun P, et al. Allelotype of human breast carcinoma: a second major site for loss of heterozygosity is on chromosome 6q. *Oncogene*. 1991;6(9):1705–1711.
21. Hansen LL, Jensen LL, Dimitrakakis C, et al. Allelic imbalance in selected chromosomal regions in ovarian cancer. *Cancer Genet Cytogenet*. 2002;139(1):1–8.
22. Kataoka H, Miura Y, Joh T, et al. Alpha-fetoprotein producing gastric cancer lacks transcription factor ATBF1. *Oncogene*. 2001;20(7):869–873.
23. Piao Z, Park C, Kim JJ, et al. Deletion mapping of chromosome 16q in hepatocellular carcinoma. *Br J Cancer*. 1999;80(5–6):850–854.
24. Snijders AM, Nowee ME, Fridlyand J, et al. Genome-wide-array-based comparative genomic hybridization reveals genetic homogeneity and frequent copy number increases encompassing CCNE1 in fallopian tube carcinoma. *Oncogene*. 2003;22(27):4281–4286.
25. Sun X, Frierson HF, Chen C, et al. Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nat Genet*. 2005;37(4):407–412.
26. Sun X, Zhou Y, Otto KB, et al. Infrequent mutation of ATBF1 in human breast cancer. *J Cancer Res Clin Oncol*. 2007;133(2):103–105.
27. Rakha EA, Green AR, Powe DG, et al. Chromosome 16 tumor-suppressor genes in breast cancer. *Genes Chromosomes Cancer*. 2006;45(6):527–535.
28. Kaspar P, Dvorakova M, Kralova J, et al. Myb-interacting protein, ATBF1, represses transcriptional activity of Myb oncoprotein. *J Biol Chem*. 1999;274(20):14422–14428.
29. Miura Y, Kataoka H, Joh T, et al. Susceptibility to killer T cells of gastric cancer cells enhanced by Mitomycin-C involves induction of ATBF1 and activation of p21 (Waf1/Cip1) promoter. *Microbiol Immunol*. 2004;48(2):137–145.
30. Morinaga T, Yasuda H, Hashimoto T, et al. A human alpha-fetoprotein enhancer-binding protein, ATBF1, contains four homeodomains and seventeen zinc fingers. *Mol Cell Biol*. 1991;11(12):6041–6049.
31. Nojiri S, Joh T, Miura Y, et al. ATBF1 enhances the suppression of STAT3 signaling by interaction with PIAS3. *Biochem Biophys Res Commun*. 2004;314(1):97–103.
32. Yasuda H, Mizuno A, Tamaoki T, et al. ATBF1, a multiple-homeodomain zinc finger protein, selectively down-regulates AT-rich elements of the human alpha-fetoprotein gene. *Mol Cell Biol*. 1994;14(2):1395–1401.
33. Dong XY, Sun X, Guo P, et al. ATBF1 inhibits estrogen receptor (ER) function by selectively competing with AIB1 for binding to the ER in ER-positive breast cancer cells. *J Biol Chem*. 2010;285(43):32801–32809.
34. Dong XY, Guo P, Sun X, et al. Estrogen up-regulates ATBF1 transcription but causes its protein degradation in estrogen receptor-alpha-positive breast cancer cells. *J Biol Chem*. 2011;286(16):13879–13890.
35. Sun X, Fu X, Li J, et al. Heterozygous deletion of Atbf1 by the Cre-loxP system in mice causes preweaning mortality. *Genesis*. 2012;50(11):819–827.
36. Sun X, Fu X, Li J, et al. Deletion of atbf1/zfhx3 in mouse prostate causes neoplastic lesions, likely by attenuation of membrane and secretory proteins and multiple signaling pathways. *Neoplasia*. 2014;16(5):377–389.
37. Zighelboim I, Goodfellow PJ, Gao F, et al. Microsatellite instability and epigenetic inactivation of MLH1 and outcome of patients with endometrial carcinomas of the endometrioid type. *J Clin Oncol*. 2007;25(15):2042–2048.
38. Billingsley CC, Cohn DE, Mutch DG, et al. Polymerase varepsilon (POLE) mutations in endometrial cancer: Clinical outcomes and implications for Lynch syndrome testing. *Cancer*. 2015;121(3):386–394.
39. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–498.
40. Forbes SA, Bhamra G, Bamford S, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet*. 2008;Chapter 10:Unit 10 11.
41. Bacher JW, Flanagan LA, Smalley RL, et al. Development of a fluorescent multiplex assay for detection of MSI-High tumors. *Dis Markers*. 2004;20(4–5):237–250.
42. Kupets R, Le T, Le T, et al. The role of adjuvant therapy in endometrial cancer. *J Obstet Gynaecol Can*. 2013;35(4):375–379.
43. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. *World Health Organ Tech Rep Ser*. 1995;854:1–452.
44. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet*. 2009;105(2):103–104.
45. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Stat Methodol*. 1995;57(1):289–300.
46. Bokhman JV. Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol*. 1983;15(1):10–17.
47. Boeva V, Popova T, Lienard M, et al. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics*. 2014;30(24):3443–3450.
48. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568–576.
49. Kim TM, Laird PW, Park PJ. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell*. 2013;155(4):858–868.
50. Novetsky AP, Zighelboim I, Thompson DM Jr, et al. Frequent mutations in the RPL22 gene and its clinical and functional implications. *Gynecol Oncol*. 2013;128(3):470–474.
51. Tritz D, Pieretti M, Turner S, et al. Loss of heterozygosity in usual and special variant carcinomas of the endometrium. *Hum Pathol*. 1997;28(5):607–612.
52. Pere H, Tapper J, Wahlstrom T, et al. Distinct chromosomal imbalances in uterine serous and endometrioid carcinomas. *Cancer Res*. 1998;58(5):892–895.
53. Le Gallo M, Bell DW. The emerging genomic landscape of endometrial cancer. *Clin Chem*. 2014;60(1):98–110.
54. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet*. 2011;88(4):440–449.
55. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68(4):820–823.
56. Davoli T, Xu AW, Mengwasser KE, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*. 2013;155(4):948–962.
57. Solimini NL, Xu Q, Mermel CH, et al. Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science*. 2012;337(6090):104–109.