Received: 16 October 2022 | Revised: 15 November 2022 | Accepted: 1 December 2022

DOI: 10.1002/ijc.34406

TUMOR MARKERS AND SIGNATURES

IJC INTERNATIONAL JOURNAL of CANCER

The WID-EC test for the detection and risk prediction of endometrial cancer

James E. Barrett[1,2,3] | Allison Jones[2] | Iona Evans[2] | Chiara Herzog[1] | Daniel Reisel[2] | Adeola Olaitan[2] | Tim Mould[2] | Nicola MacDonald[2] | Konstantinos Doufekas[2] | Claire Newton[4,5] | Emma J. Crosbie[6,7] | Line Bjørge[8,9] | Nicoletta Colombo[10,11] | Lukas Dostalek[12,13] | Laura Costas[14] | Paula Peremiquel-Trillas[14] | Jordi Ponce[15] | Xavier Matias-Guiu[16] | Michal Zikan[12,13] | David Cibula[13] | Jiangrong Wang[17] | Karin Sundström[17] | Joakim Dillner[17] | Martin Widschwendter[1,2,3]

Correspondence
Martin Widschwendter, European Translational Oncology Prevention and Screening (EUTOPS) Institute, Milser Str. 10, 6060 Hall, Austria.
Email: martin.widschwendter@uibk.ac.at; m.widschwendter@ucl.ac.uk

Funding information
European Union's Horizon 2020 European Research Council Programme, H2020 BRCA-ERC, Grant/Award Number: 742432; European Union's Horizon 2020 Research and Innovation Programme, H2020 FORECEE, Grant/Award Number: 634570; The Eve Appeal

Abstract

The incidence of endometrial cancer is rising. Measures to identify women at risk and to detect endometrial cancer earlier are required to reduce the morbidity triggered by the aggressive treatment required for advanced endometrial cancer. We developed the WID-EC (Women's cancer risk IDentification-Endometrial Cancer) test, which is based on DNA methylation at 500 CpG sites, in a discovery set of cervical liquid-based cytology samples from 1086 women with and without an endometrial cancer (217 cancer cases and 869 healthy controls) with a worse prognosis (grade 3 or ≥stage IB). We validated the WID-EC test in an independent external validation set of 64 endometrial cancer cases and 225 controls. We further validated the test in 150 healthy women (prospective set) who provided a cervical sample as part of the routine Swedish cervical screening programme, 54 of whom developed endometrial cancer within 3 years of sample collection. The WID-EC test identified women with endometrial cancer with a receiver operator characteristic area under the curve (AUC) of 0.92 (95% CI: 0.88-0.97) in the external set and of 0.82 (95% CI: 0.74-0.89) in the prospective validation set. Using an optimal cutoff, cancer cases were detected with a sensitivity of 86% and a specificity of 90% in the external validation set, and a sensitivity and specificity of 52% and 98% respectively in the prospective validation set. The WID-EC test can identify women with or at risk of endometrial cancer.

Abbreviations: AUC, area under the curve; CAH, complex atypical hyperplasia; CI, confidence interval; CIN, cervical intraepithelial neoplasia; DNAme, DNA methylation; EC, endometrial cancer; HPV, human papilloma virus; IC, immune cell content; PMR, percentage of fully methylated reference; ROC, receiver operating characteristic; tDNA, tumour DNA.

James E. Barrett, Allison Jones, Iona Evans, and Chiara Herzog contributed equally to our study.

For affiliations refer to page 1986

KEYWORDS
cervical sample, DNA methylation, endometrial cancer, epigenetics, risk prediction

What's new?

The endometrial cancer incidence is rising, and imaging-based screening is often not suitably accurate for endometrial cancers. Here, the authors show that a new DNA methylation signature (Women's cancer risk IDentification-Endometrial Cancer, or WID-EC) can both detect the presence and predict the risk of endometrial cancer. Together with three other recently-reported methylation indices that detect breast, ovarian, and cervical cancers (WID-BC, WID-OC, and WID-CIN), WID-EC completes a list of signatures that can detect/predict all four women-specific cancers using a single cervical sample and the same technology platform for methylation array-based tests.

1 | INTRODUCTION

Endometrial cancer has become the most common gynaecological cancer in developed countries. By 2030, it is expected that endometrial cancer will be the third most common cancer affecting women in the United States after that of both breast and thyroid.[1] Approximately 20% of women with endometrial cancer present with high-risk and/or more advanced disease characteristics with an increased incidence of distant metastases and cancer-related death and hence, in addition to surgery, require adjuvant chemo- and radiotherapy[2] which are associated with a high morbidity. Therefore, identification of a simple, painless and convenient tool that combines early detection and risk prediction of specifically high-risk endometrial cancers is a key research priority.[3,4]

Imaging-based screening for endometrial cancers is often not suitably accurate. It is primarily suitable for postmenopausal women, and even in this group, the performance characteristics of endometrial thickness (10 mm as cutoff) show a sensitivity of only 50% with a positive predictive value of 4.5% among women negative for bleeding symptoms.[5]

Assessing cervical samples or samples taken directly from the endometrium and assessing molecular alterations is an attractive concept for identifying women with or at risk for endometrial cancer.[6] A recent paper[7] demonstrated an overwhelming abundance of 'driver' mutations in cancer genes in the normal endometrium which originate during the first decades of life and subsequently progressively colonise the epithelial lining of the normal endometrium. Furthermore, an assay identifying mutations in 18 genes as well as aneuploidy in cervical brush samples was recently able to identify 81% of women presenting with endometrial cancer.[8] However, in this study, the average age of cases and controls was 62 and 34 years, respectively. The consistent observation of a high allele frequency of pathogenic driver mutations in DNA from nonmalignant normal endometrium with increasing age[9-11] combined with the fact that the cases were almost twice as old as the controls makes it difficult to judge the true specificity of a tool which assesses somatic mutations.

We along with others have shown that epigenetic alterations are likely to be causative for endometrial cancer development[12] and DNA methylation in vaginal fluid or cervical samples may identify women with endometrial[12-18] or other women's cancers.[19-22] In particular, in recent studies, we showed that a PCR-based assay evaluating methylation in the genes *ZSCAN12* and *GYPC* is capable of identifying endometrial[22] and cervical[23] cancer cases.

Differential DNA methylation levels of CpG sites (sites where a cytosine base is located next to a guanosine base in the human genome[24]) can be used as both a surrogate readout for factors which drive cancer formation and thereby predict cancer risk,[24] and a diagnostic tool that indicates the presence of a cancer.[25] The bulk of the DNA extracted from a cervical screening sample contains (i) DNA from normal cells (ie, hormone-sensitive cervical epithelial cells originating from the same embryological structure as the tissue at risk) that provides the *cancer risk component* of the signature, and (ii) DNA from cell-detritus draining from the endometrial cavity (the quantity of which is likely to be higher with respect to endometrial cancers) which provides the *diagnostic component* of the signature.

Here, we have developed and validated an array-based DNA methylation signature in cervical samples which is capable of identifying both women with, or at risk for, endometrial cancer. This test may be suitable in combination with other methylation array-based tests for women's cancers that we have recently developed, and could therefore identify the presence, or potentially risk of, four women's cancers — endometrial, breast,[19] ovarian,[21] and cervical[26] — using a single cervical sample and the same test platform.

2 | MATERIAL AND METHODS

2.1 | Study design and data and sample acquisition

The study was conducted as part of the multicentre 'FORECEE' study involving 15 recruitment sites across Europe, and study details have previously been published.[19,21,22] Briefly, women with symptoms suggestive of endometrial cancer (who were then subsequently diagnosed with endometrial cancer; cases) or a nonmalignant benign gynaecological condition (controls) were approached during outpatient hospital clinics, while healthy volunteers from the general population (population-based controls) were approached via outreach

campaigns, public engagement, and as part of cervical screening programmes. After signing an informed consent, participants completed an epidemiological questionnaire as well as a feedback form after their participation.

Cervical samples collection was conducted by trained staff using the ThinPrep system (Hologic Inc, cat #70098-002) as previously published.[19,21,22]

Women with (i) a current diagnosis of grade 3 and/or stage IB or above endometrial cancer and (ii) recruited prior to receiving any systemic treatment or surgery or radiotherapy were eligible as endometrial cancer cases. For the entire FORECEE Discovery set (which also included controls matched to breast and ovarian cancer cases), controls were initially matched one-to-one with cases based on menopausal status, age (5-year age ranges where possible), and recruitment centre/country. However, due to an imbalance in recruitment of cases and controls at some centres, a number of cases were matched on age and menopausal status alone.

Within the FORECEE sample collection, two sets of samples (Discovery and External validation) were collected. Samples within the External validation set were kept exclusively for validation of signatures.

2.2 | Prospective validation dataset

All cervical liquid-based cytology samples processed in the capital region of Stockholm in Sweden are biobanked through a state-of-the-art platform at the Karolinska University Laboratory, Karolinska University Hospital, as previously described.[27] Since 2013, virtually 100% of the ~150 000 LBCs per year are compacted and stored in a 600 μL, 96 well plate format at −27°C. This allows for preservation of whole cells, and analyses of DNA, RNA and protein content, among others. The biobank is linked to the Swedish health register infrastructure through the individually unique personal identification number (PIN).[28] A cohort was defined of all women participating in cervical screening, or clinically indicated testing, during the years 2011-2015 and linked to the National Cancer Register at the Swedish National Board of Health and Welfare, in order to identify all cases of endometrial cancer occurring in the sample collection during these years. Population-based control women were frequency matched in a 1:2 ratio to the cases, by age, sample type (screening-based or clinic-based) and calendar year of sample.

2.3 | Self and clinical collected samples

Five cases and five controls matched by age enrolled at the Bellvitge University Hospital in the Barcelona province provided a cervicovaginal sample using the self-collection device Evalyn Brush at the outpatient clinic after a brief explanation by medical staff. The Evalyn Brush was rotated 5 times through 360° once introduced in the vagina as indicated per fabricant protocol to maximise cell sampling. Participants were enrolled between October 2017 and February 2019. After using the Evalyn device, cervical samples were also collected by the gynaecologist using the Cervix brush as described above. Samples were stored using ThinPrep.

2.4 | DNA methylation array analysis

Array analysis was conducted as previously described.[19,21] Briefly, cervical DNA was extracted, normalised, bisulfite modified, and subjected to methylation analysis on the Illumina InfiniumMethylation EPIC BeadChip (Illumina, CA, USA) at UCL Genomics according to the manufacturer's standard protocol. The EPIC bead chip array encompasses over 850 000 CpG sites.[29]

2.5 | Methylation analysis

All methylation microarray data were processed through the same standardised pipeline to remove samples with low fluorescence intensities, failed probes, cross-reactive, and SNP probes, as previously described.[19,21] Additionally, probe bias correction and sample cell composition inference were conducted using the beta mixture quantile normalisation (BMIQ) and EpiDISH algorithm,[30] respectively.

2.6 | WID-EC index development

To derive a methylation-based classifier index capable of identifying endometrial cancer, we performed feature preselection prior to training, identifying those CpGs that showed the largest difference between cases and controls. In previous work, we have shown that contamination by immune cells (IC) can present a significant challenge with respect to the identification of differentially methylated positions (DMPs), as differential methylation that occurs solely in epithelial cells can be diminished in samples with high IC and vice versa.[19,21] Linear regression of beta values on IC for each CpG site, with separate models fitted to cases and controls, was used to overcome this. Intercepts were used to identify mean beta values for each group in either pure epithelial or pure immune cell populations (IC = 0, and IC = 1, respectively), with the difference between these intercept points able to provide epithelial- and immune-specific differences between cases and controls (delta-beta). A ranked list of CpGs was generated by taking the CpG with the largest epithelial delta-beta, followed by the CpG with the largest immune delta-beta, followed by the next largest epithelial delta-beta and so forth (any duplicates were removed). Training on the top n-ranked CpGs (max n = 30 000) was performed using the R package glmnet[31] with a mixing parameter value of alpha = 0 (ridge penalty) and alpha = 1 (lasso penalty) and a binomial response type. Data from the training dataset were used to fit the classifiers. Tenfold cross-validation was used inside the training set by the cv.glmnet function in order to determine the optimal value of the regularisation parameter lambda. The AUC was used as a metric of

classifier performance which was evaluated on the internal validation dataset as a function of n. The number of inputs for the final model was determined based on the classifier with the highest AUC in the internal validation set, following which the training and internal validation datasets were combined and the classifier was refitted using the entire discovery dataset with alpha and lambda fixed to their optimal values. This finalised classifier was then applied to the external validation dataset and the corresponding AUC was computed.

Denoting the top n CpGs as $x_1,\ldots,x_n$ and the regression coefficients from the trained classifier as $w_1,\ldots,w_n$ then WID-EC index $= \sum_{i=1}^{n}(w_i x_i - \mu)/\sigma$ where μ and σ are defined as the mean and SD of the quantity $\sum_{i=1}^{n} w_i x_i$ in the training dataset (ie, the index is scaled to have zero mean and unit SD in the training dataset).

2.7 | Estimation of endometrial tumour DNA proportion

The EpiDISH algorithm provides an estimate of cell type proportions within a given sample. A reference dataset consisting of CpGs that are unique to each cell type must be provided. In order to construct such a reference dataset for endometrial cancer, 11 epithelial, 7 fibroblast, 48 immune, and 9 endometrial cancer samples were downloaded from the GEO (Table S3). Each cell type was compared to the other three cell types (which were combined into one group) to identify CpGs that are unique to that cell type. A Wilcoxon rank sum test was used to test for differential methylation at each CpG. For epithelial cells, any CpGs with a *P*-value >.01 after false discovery rate (FDR) adjustment and an absolute difference in methylation >.56 were selected (212 in total). For fibroblasts, any CpGs with FDR adjusted *P*-values >.01 and differential methylation >.67 were selected (201 in total). For immune cells, any CpGs with FDR adjusted *P*-values >.01 and differential methylation >.84 were selected (218 in total). For endometrial cancer cells, any CpGs with FDR adjusted *P*-values >.01 and differential methylation >.69 were selected (217 in total). These were chosen such that in the reference panel there would be a roughly equal number of CpGs corresponding to each cell subtype (ie, approximately 200 CpGs for each cell type). The final reference dataset therefore consisted of 848 CpGs. It was observed that the inferred proportion of tumour DNA and epithelial cells were strongly associated in control samples from cancer-free women. Local polynomial regression fitting (using the loess R function) was used to regress the inferred tumour DNA proportion on the epithelial proportion (in control samples only) and the residuals were used as estimates for tumour DNA proportion.

2.8 | Statistical analyses

Odds ratios and 95% confidence intervals (CI) were calculated using the odds ratio function in the epitools R package, version 0.5.10. All statistical tests were two-sided and a *P*-value <.05 was considered statistically significant.

2.9 | Signal-to-noise inference on the Illumina EPIC array

Signal-to-noise inference was carried as previously described.[19]

3 | RESULTS

3.1 | General characteristics of the whole study cohort

We collected liquid-based cytology (LBC) samples from 217 women at the time of endometrial cancer diagnosis with a current diagnosis of grade 3 and/or stage IB or above for the Discovery Set (Figure 1). Samples were collected either at the time before an endometrial biopsy was taken for diagnostic purposes or before commencing a hysterectomy. Control LBC samples were obtained from 869 control women without endometrial cancer (n = 593 general population, n = 276 women attending hospital for benign women-specific conditions) from 15 European centres. Oversampling of younger women was deliberately performed to develop a risk predictor that would also be suitable in younger women.

3.2 | Sample heterogeneity and differential methylation

The cell-type distributions as assessed by our EPIDISH algorithm were similar between cancer cases and controls, although there were small but significant differences in some immune cell subtypes (Figure S1B). After Holm multiple testing adjustment, 116 658 CpGs showed significant differential methylation between cancer cases and controls (Figure S1A).

3.3 | Development of discriminatory index

We identified the top differentially methylated features (CpGs) in epithelial or immune cells between cases and controls and used the top n-ranked CpGs as input for ridge and lasso regression to derive a diagnostic methylation signature (women's cancer risk identification—endometrial cancer, WID-EC) that could detect endometrial cancer. The Discovery Set was split two thirds to one third into training and internal validation (n [training] = 572 cancer-free controls, 144 endometrial cancer cases; n [internal validation] = 297 controls, 73 cases). The internal validation set was used with the intention of evaluating the performance of the classifier as a function of the number of CpGs used to construct the index. The area under the receiver operator characteristic curve (AUC) was used as a measure of predictive performance.

The predictive performance of the index was evaluated as a function of the number of n CpGs used to train the classifier using the internal validation dataset. A maximum performance of 0.97 (95% CI:

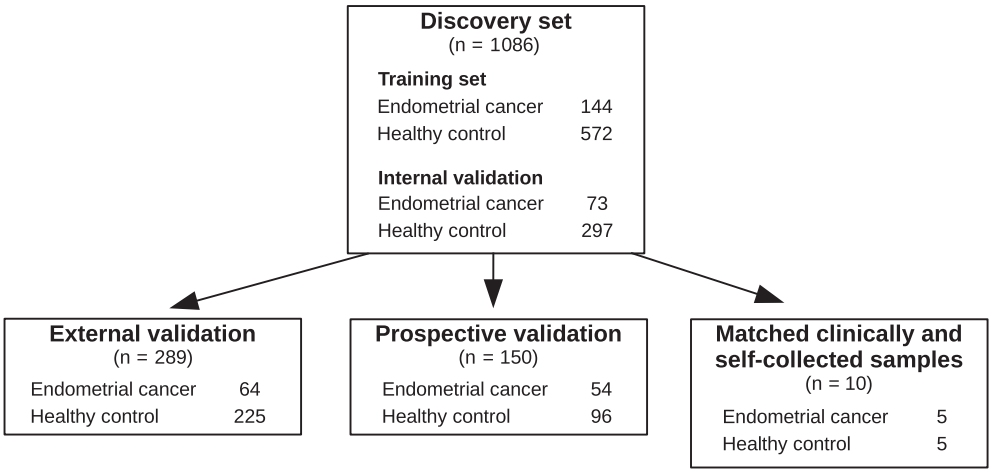

FIGURE 1 Experimental design

0.94-0.99) was achieved using 500 CpGs with ridge regression (Figure S1C; list of CpGs provided in Table S4). The discriminatory performance was broadly independent of immune cell (IC) proportion (Figure S1E), with AUCs of 0.98 (95% CI: 0.97-1.00) and 0.95 (95% CI: 0.91-0.99) in samples with and IC ≤0.5 and >0.5, respectively. The WID-EC index was slightly associated with IC fraction in controls (linear regression coefficient of 0.29, $P < .001$), with a negative trend in cancer cases (linear regression coefficient of −0.28, $P = 0.6$). An optimal WID-EC index cutoff value of 0.14, corresponding to a sensitivity of 86%, and specificity of 95% was defined using the internal validation set (Figure S1D).

3.4 | External validation

To validate the index, we evaluated its performance in a separate, independent external validation dataset consisting of 64 endometrial cancer cases and 225 controls (Figure 1; Table S1). The WID-EC index achieved an overall AUC of 0.92 (95% CI: 0.88-0.97) (Figure 2A), and 0.91 (95% CI: 0.8-1) and 0.9 (95% CI: 0.84-0.97) for samples from women aged below (≤60, n = 18 cases, n = 160 controls) or above 60 years of age (>60, n = 46 cases, n = 65 controls), respectively (Figure 2B). Endometrial cancer is less common in women aged ≤60 years of age, this indicated that index performance was strongly not dependent on age. Odds ratios, corresponding to quartiles defined on the internal validation set, reveal a 25.94-fold increased risk for women in the top compared to the bottom quartile of the WID-EC index (unadjusted), which corresponds to an OR of 13.52 after adjustment for age, menopausal status, and BMI (Table 1). Based on the WID-EC index cutoff of 0.14, cancer cases were identified with a sensitivity and specificity of 86% and 90%, respectively.

3.5 | Prospective validation

To assess whether the WID-EC index is able to predict future endometrial cancer risk, we analysed set of routine cervical screening samples collected from healthy women who subsequently developed endometrial cancer (cases; n = 54; average time between sample collection and cancer diagnosis was 304 days) or stayed cancer-free at our follow-up (controls, n = 96) (Figure 2C; Table S1). The WID-EC exhibited an AUC of 0.82 (95% CI: 0.74-0.89) overall. (Figure 2D). In women ≤60 years of age, the AUC was 0.77 (95% CI: 0.66-0.89), and in women >60 years of age was 0.89 (95% CI: 0.79-0.99) (Figure 2D), suggesting that the WID-EC risk predictive performance may be higher in women >60 years.

Based on the cutoff of 0.14, cancer cases were identified with a sensitivity and specificity of 52% and 98% respectively across all women, and with a sensitivity and specificity of 81% and 94% respectively in women >60 years. A Kaplan-Meier curve revealed excellent discrimination between cases and noncases over a five-year period (Figure S2G).

Interestingly, while the cell type composition of samples in our three datasets was similar (Figure S2A,B), we observed a systematic loss of methylation in cancer-free controls from the Prospective set in comparison to the Discovery set, a loss that predominantly occurred at CpG-sparse 'Open Sea' and 'Shore' regions of the genome (Figure S2C). Of note, the impact of storage was assessed in controls, because we expected control samples from both the training set and the prospective validation set to exhibit a similar overall distribution of methylation, as they we derived cancer-free women, and we did not do a direct comparison of methylation levels in cases between the two sets as they are likely not directly comparable (ie, cases form the Prospective validation set had yet to be diagnosed). In line with previous findings,[19] we hypothesised that these changes may be due to degradation associated with long-term storage between sample collection and transfer to the biobank (median 95 days, range 15 to 1001 days), and storage time within the biobank at −27°C in Preservcyt (median 2090 days, range 1250 to 2852 days). Indeed, there was an association of the WID-EC-index with the time in biobank, yet values for controls were generally below the threshold of 0.14 (Figure S2F). Due to the fact that almost half (ie, 228) of the 500 CpGs forming the WID-EC index are Open Sea CpGs (Figure S2D), the

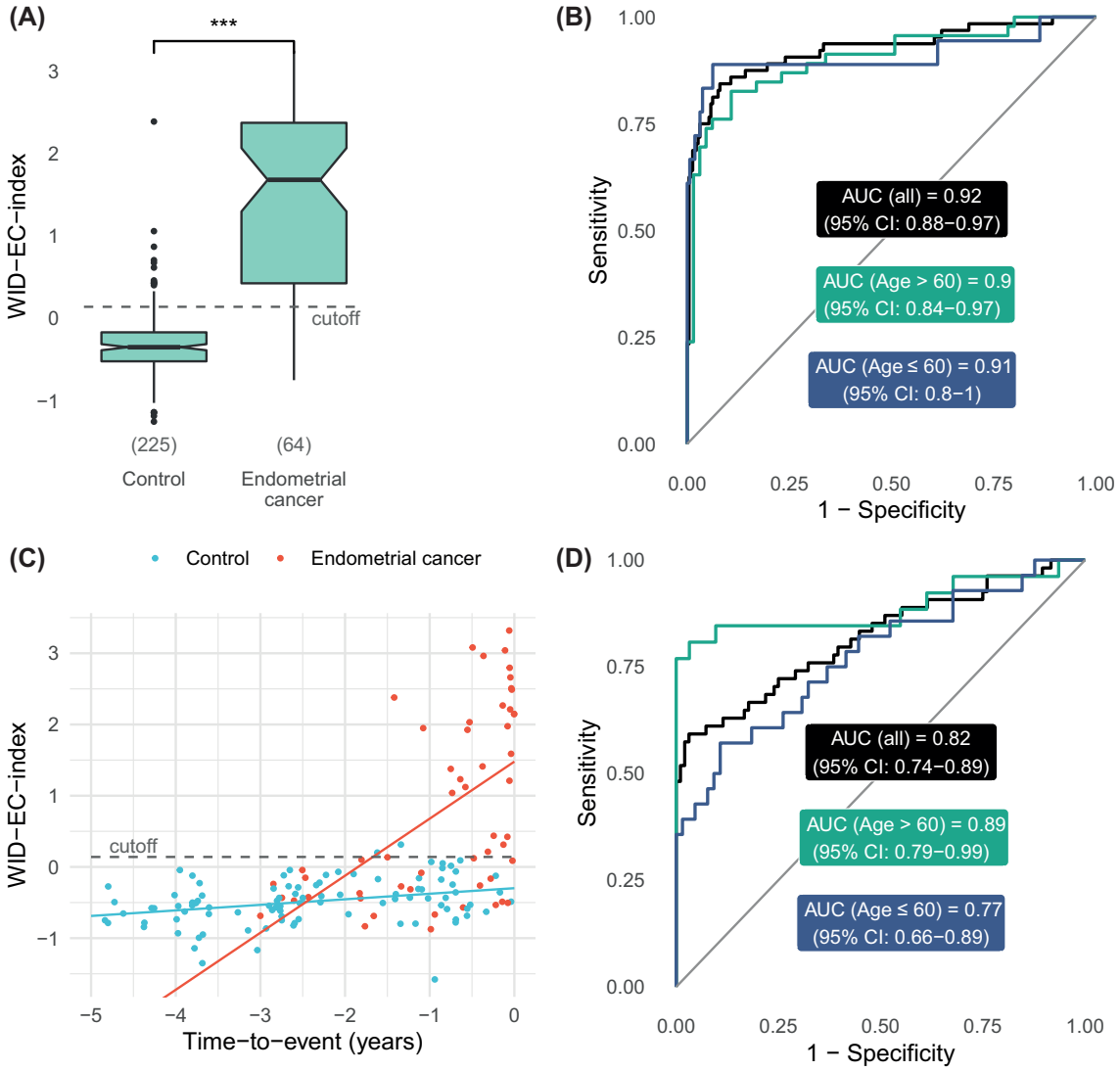

FIGURE 2 WID-EC index in endometrial cancer cases and healthy controls in the external validation dataset, using the cutoff that was defined based on the internal validation set (A). ROC curve in the external validation set (overall and in age groups <60 or ≥60 years of age); overall: n = 64 cases, n = 225 controls, thereof <60: n = 18 cases, n = 160 control and ≥60: n = 45 cases, n = 65 controls (B). WID-EC index vs time-to-event in prospectively collected validation samples (C). ROC curve corresponding to the prospective validation set overall: n = 54 cases, n = 96 controls, thereof <60: n = 28 cases, n = 65 control and ≥60: n = 26 cases, n = 31 controls (D)

overall performance of the WID-EC index in predicting endometrial cancer risk may have been compromised in this cohort-based setting. We applied a previously developed statistical technique to infer the signal-to-noise ratio of each sample using the global intensity profiles from the red and green channels and the level of background intensities,[19] which confirmed that biobanked samples suffered from a significantly lower signal-to-noise ratio (Figure S2E). Importantly, we moreover observed that the signal to noise was similarly reduced in both cases and controls and hence we expected that storage issues impact both cases and controls equally in the Prospective validation set.

Lastly, since the CpG islands were the least affected by this storage effect, we evaluated the AUC using the WID-EC-index using only CpGs from CpG Islands, which revealed a similar but slightly higher AUC than when considering all WID-EC-index CpGs in this set (0.85 compared to 0.82) (Figure S2H, comparison to Figure 2D).

Taken together, despite technical issues in the Prospective Validation set, the performance of the WID-EC test was highly promising in a population-based prospective cohort setting.

3.6 | Association of the WID-EC with epidemiological, clinical, and technical factors

We next explored the relationship between the WID-EC index and epidemiological and clinical variables in the internal and external validation sets (details in Tables S1 and S2). Perhaps unsurprisingly, a statistically significant association was found between the WID-EC index

TABLE 1 Odds ratios corresponding to quartiles defined using the internal validation dataset

Quantile	Controls	Cases	OR (unadjusted)	OR (adjusted)
External validation				
(−1.36, −0.66)	35	1	1.00 (reference)	1.00 (reference)
(−0.66, −0.43)	45	1	0.78 (0.02,31.16)	0.61 (0.01, 20.2)
(−0.43, −0.23)	77	4	1.65 (0.22,46.2)	0.66 (0.05, 16.74)
(−0.23, 1.03)	68	58	25.94 (5.37622.56)	13.52 (2.51, 251.76)
Prospective validation				
(−1.36, −0.66)	26	5	1.00 (reference)	1.00 (reference)
(−0.66, −0.43)	31	6	1 (0.26,3.97)	1.03 (0.28, 4)
(−0.43, −0.23)	18	7	1.98 (0.53,7.9)	2.31 (0.57, 10.21)
(−0.23, 1.03)	21	36	8.51 (3.01,28.84)	8.82 (2.7, 34.12)

Note: Adjustment was based on a logistic regression model with age, menopausal status, and BMI included as covariates for the external validation datasets. For the prospective validation dataset adjustment was made for age.

and age both in controls (correlation coefficient = .37, $P < 10^{-16}$; Figure 3A) and cancer cases (correlation coefficient = .16, $P = .06$), but, as shown previously, this did not affect the performance of the WID-EC test in the external validation dataset (Table 1, Figure 2B). The increase of the WID-EC index with age was also shown for the prospective validation set (Figure S3A). Moreover, we observed a significant correlation of .14 ($P < .01$) between the index and BMI in controls (Figure 3B), which was interesting as BMI constitutes a significant risk factor for endometrial cancer.[32] Although the WID-EC index increased with increasing BMI classification (normal, overweight, obese) in controls, the classification between controls and cases was largely unaffected by BMI group in the external validation set (Figure 3C). The WID-EC index was significantly elevated in postmenopausal controls (Figure 3E), reflecting the association with age. No significant association was observed with parity in postmenopausal controls (Figure 3D). The index was significantly elevated in stage III/IV cancers (Figure 3F) and in grade II cancers compared to grade I cancers (Figure 3G). No association was observed with histological subtypes (Figure 3H). The WID-EC was not associated with any technical parameters, including date of sample processing, plate number, or sentrix position. Lastly, we compared the 593 control samples from healthy volunteers to 276 control samples taken from women presenting with benign women-specific conditions, but did not find any significant differences (Figure S3B), indicating the signature is cancer-specific.

3.7 | Inferred proportion of tumour DNA

In line with the final finding presented above, and due to the anatomical proximity between the tissue of origin of the cancer (endometrium) and the area from which we sampled (cervix), we investigated whether the discriminatory signal might be driven by tumour DNA draining from the uterus to the cervix. Alternatively, the signal could be derived from a generic epigenetic endometrial cancer risk 'footprint' that may drive cancer in the endometrium but is retained in cervical epithelial cells. To infer the presence of tumour DNA in our cervical samples, we used 11 epithelial, 7 fibroblast, 42 immune cell, and 9 endometrial cancer tissue samples (Table S3) to develop a new reference panel for use with the EpiDISH algorithm (see Section 2 and Figure S4). The estimated proportion of tumour DNA was higher in cases than in controls; for the latter, the inferred proportion of tumour DNA was close to zero (Figure 4A); few control samples showed a proportion of tumour DNA higher than 0%, which might be explained by CpGs in the reference panel not exclusive to endometrial cancer. Forty-three per cent (43%) of cervical samples from endometrial cancer cases consisted of >10% tumour DNA. Although the WID-EC index strongly increased with the proportion of tumour DNA (correlation coefficient .70, $P < 10^{-16}$), a strong difference between cases and controls was present even in those cases estimated to have no tumour DNA present (Figure 4B). This indicated that the discriminatory signal did not depend entirely on the presence of tumour DNA.

3.8 | Utility of self-collected samples

As a proof of concept for the identification of women with a uterine cancer based on a self-collected sample, we analysed both healthcare professional-collected (ie, gynaecologists)- and self-collected samples from 10 women who presented with symptoms indicative of potential endometrial cancer and, subsequent to sample donation, underwent an endometrial biopsy which led to endometrial cancer diagnosis in 5 out of these 10 women (Figure 1). In 18/20 samples sufficient DNA was available. The WID-EC index showed a high correlation between self- and clinic-collected samples in those 8 women in which both samples were available for DNA methylation analysis (Figure 4C). Whereas both the WID-EC index and the tumour-DNA fraction were able to discriminate between cases and controls the performance of the WID-EC index and the tumour-DNA fraction was better in the clinic- and self-collected samples, respectively (Figure 4D,E).

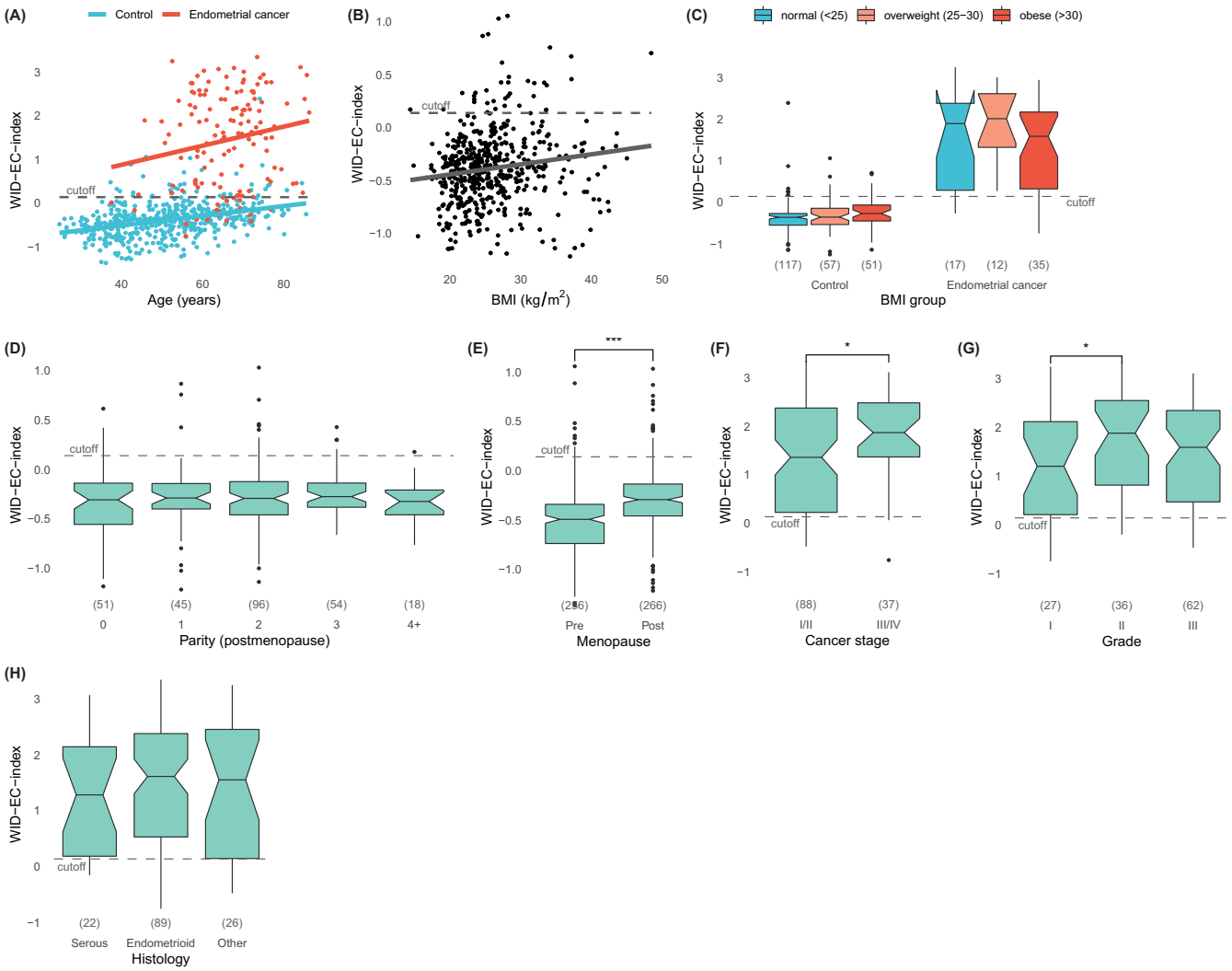

FIGURE 3 The WID-EC index vs age in samples from the internal and external validation datasets (A). Correlation with body mass index (BMI) in controls (B), and performance in the external validation set split by BMI group (<25 normal, 25-30 overweight, >30 obese) (C). Association with parity (D), menopause (E), stage (F), grade (G) and histology (H). Figures A-C and E are based on controls from both the internal and external validation datasets. Figures F to H are based on cases from the internal and external validation datasets

3.9 | Side-by-side comparison to the PCR-based WID-qEC test

We have recently developed and validated a PCR-based approach[22] that evaluates methylation in the genes *ZSCAN12* and *GYPC* (2 regions in the latter gene), using the same samples described in our study to identify endometrial cancer cases. For 132/370 samples of the internal validation set (n = 68 endometrial cancers, n = 64 healthy controls) and 82/150 samples in the prospective validation set (n = 31 future endometrial cancers, n = 51 controls), we had data for both array-based (WID-EC) or PCR-based (WID-qEC) index. In the internal validation, performance was similar (Figure S5A, Difference in AUC: 0.026, 95% CI: −0.022 to 0.075). In the prospective validation, the PCR-based WID-qEC outperformed the array-based WID-EC although this was nonsignificant (Figure S5B, Difference in AUC: 0.10, 95% CI: −0.039 to 0.247).

4 | DISCUSSION

Endometrial cancer is the most common gynaecological cancer in high human development index regions and among those cancers with the most rapidly increasing incidence rates.[33]

Using partially novel statistical approaches based on an epigenome-wide assay study of, to our knowledge, the largest scale so far in the literature, we propose a novel test—the WID-EC test—that is based on a DNA methylation signature in cervical samples and that can identify women both with, and at risk for, endometrial cancer.

The WID-EC test identifies 70% of women with an endometrial cancer with a specificity of 97%, and 90% of women with an endometrial cancer with a specificity of 78%. We leveraged an ambitious second validation cohort using biobanked samples from a real-world screening programme. The sensitivity in this setting was only 52%

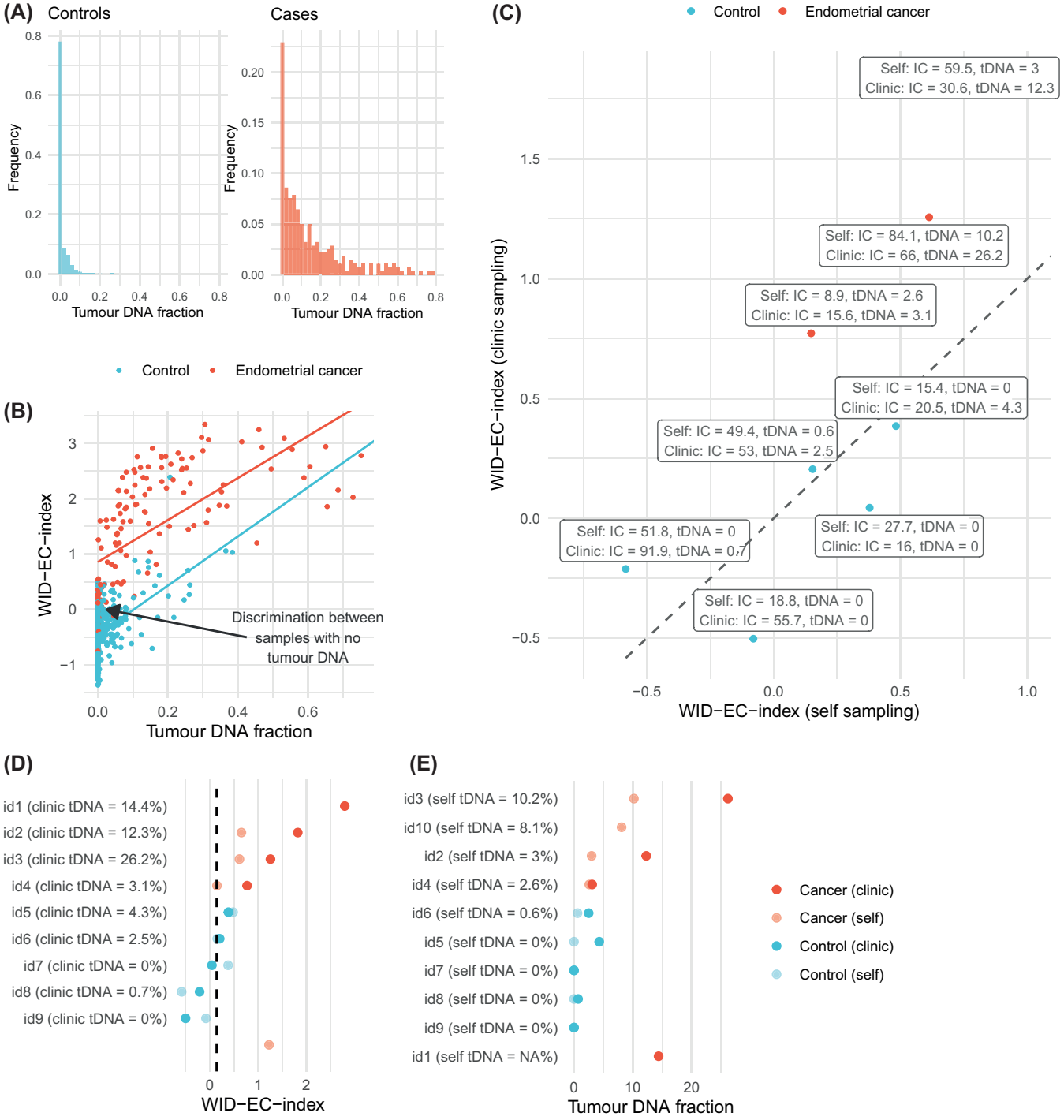

FIGURE 4 The estimated proportion of tumour DNA in each cervical smear sample in the EC external validation set as estimated using the EpiDISH algorithm (A). The WID-EC index vs the estimated proportion of tumour DNA in the EC external validation set (B). The WID-EC index in matched clinically collected and self-collected samples (C). The WID-EC index (D) and the estimated fraction of tumour DNA (E) in clinically and self-collected samples

(but 81% in >60-year-old women), this slightly lower performance (which was isolated to the prospective validation set) was at least in part explained by long-term storage issues that impacted negatively on the signal-to-noise ratio. We showed that the WID-EC test specificity for endometrial cancers was also continuously excellent in this setting, at 98%. We would thus emphasise that detecting even half of endometrial cancer cases through a method which is easily integrated into cervical cancer screening would be of great clinical potential and earlier detection of cases could potentially reduce surgical complexity (eg, no pelvic and para-aortic node dissection as done in many centres and triggered by stage) and the need for adjuvant treatment (eg, radiotherapy +/− chemotherapy).[2]

We previously demonstrated that DNAme signature assessment using methylation array analysis in cervical screening samples is able to detect and likely predict the risk for breast,[19] ovarian,[21] and cervical[26] cancer. As a next step, we propose to validate the WID-EC test in a large prospective clinical trial which aims to detect or predict the risk for all the above women's-specific cancers, using a single sample and test platform. Depending on risk-harm and cost-effectiveness analyses in such trials, the combined and/or single screening tests could then be made available to either the entire population of women or only subgroups at elevated risk for certain cancers. We suggest that women in the top 10th percentile of the test index should be classified as 'high-risk'; in these women, the WID⁻qEC[22,23] test can be subsequently carried out using the same initial sample. The WID⁻qEC assesses DNA methylation patterns in three regions via real time PCR. Primers and probes are designed to solely amplify regions where all linked CpGs are methylated. The high performance of the PCR-based assay in particular in prospective samples which suffer from a reduced methylation signal-to-noise ratio is likely caused by the more stable signal from island regions covered by PCR, as opposed to single CpG signals across the genome. Women with a positive WID-EC (highest 10th percentile) followed by a positive WID-qEC, based on a threshold optimised for high specificity, would be referred to dilatation and curretage for further diagnostic investigation.

Inherent limitations of case-control studies also apply to our study, in particular in the Discovery and External validation sets. We were also limited by the reduced signal to noise-ratio in the Prospective validation set. Nonetheless, the current study has several strengths, including the application of an epigenome-wide discovery approach to identify biomarkers of (risk of) endometrial cancer in a surrogate tissue. Many current biomarker studies investigate the use of cell free DNA for earlier cancer detection. By contrast, our approach sets itself apart from early detection strategies as the WID-EC signature does not (solely) rely on the presence of tumour DNA (Figure 4B), and instead may also enable detection of future cancers and thereby indicate future disease risk. A further strength is the inclusion of samples from a well-defined population-based screening cohort in samples predating endometrial cancer diagnosis in the Prospective validation set. Despite the abovementioned issues with sample quality, to our knowledge only a handful of population-based prospective sample collections of cervical samples exist worldwide, and virtually all might face the same storage issues. We offer an in-depth exploration of the association of storage factors which may negatively impact on DNA methylation signatures. The fact that a predictive signal is retained despite signal to noise ratio issues is promising but warrants further future prospective validation. Lastly, we provide proof of concept for utilisation of self-collected samples in a sample set comprised of matched healthcare professional- and self-collected cervicovaginal samples. Self-collection may reduce barriers encountered with cervical screening and sample collection, and could improve access to healthcare and diagnostics at times when face to face access is limited.

AUTHOR CONTRIBUTIONS

Martin Widschwendter conceived, designed and supervised the study and received funding. James E. Barrett and Chiara Herzog carried out the statistical analyses and produced the display items. Martin Widschwendter with support from James E. Barrett and Chiara Herzog interpreted the data and drafted the manuscript. Allison Jones and Iona Evans carried out the wet-laboratory work. Daniel Reisel, Adeola Olaitan, Tim Mould, Nicola MacDonald, Konstantinos Doufekas, Claire Newton, Emma J. Crosbie, Line Bjørge, Nicoletta Colombo, Lukas Dostalek, Laura Costas, Paula Peremiquel-Trillas, Jordi Ponce, Xavier Matias-Guiu, Michal Zikan, David Cibula, Jiangrong Wang, Karin Sundström and Joakim Dillner contributed to the acquisition of the data. All authors had access to the study data, contributed to data interpretation, critically reviewed the manuscript, and reviewed and approved the final manuscript. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

AFFILIATIONS

[1]European Translational Oncology Prevention and Screening (EUTOPS) Institute, Hall in Tirol, Austria
[2]Department of Women's Cancer, UCL EGA Institute for Women's Health, University College London, London, UK
[3]Research Institute for Biomedical Aging Research, Universität Innsbruck, Innsbruck, Austria
[4]University Hospitals Bristol NHS Foundation Trust, Bristol, UK
[5]University of Bristol, Bristol, UK
[6]Division of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[7]Division of Gynaecology, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
[8]Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway
[9]Centre for Cancer Biomarkers CCBIO, Department of Clinical Science, University of Bergen, Bergen, Norway
[10]Istituto Europeo di Oncologia, Milan, Italy
[11]University of Milano-Bicocca, Milan, Italy
[12]Hospital Na Bulovce, Prague, Czech Republic
[13]Department of Obstetrics and Gynecology, General University Hospital in Prague, First Faculty of Medicine, Charles University, Prague, Czech Republic
[14]Unit of Molecular Epidemiology and Genetics in Infections and Cancer, Cancer Epidemiology Research Programme, IDIBELL, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain
[15]Department of Gynecology and Obstetrics, Hospital Universitari de Bellvitge, IDIBELL. Hospitalet de Llobregat, Barcelona, Spain
[16]Department of Pathology, Hospital Universitari de Bellvitge, IDIBELL, CIBERONC. Hospitalet de Llobregat, Barcelona, Spain
[17]Department of Laboratory Medicine, Division of Pathology, Karolinska Institutet, Stockholm, Sweden

ACKNOWLEDGEMENTS

We are most grateful to the participants of the FORECEE and BRCA-ERC studies and the management team, research nurses, interviewers,

research assistants and other staff who have taken part in gathering the data for our study. The authors also wish to acknowledge the support of the National Institute for Health Research (NIHR) and the University College London Hospitals (UCLH) Biomedical Research Centre (BRC). We wish to thank Yoo-Na Kim for assistance in preparing the manuscript and Richard Gunu for help with the sample processing.

FUNDING INFORMATION

Our study was funded by the European Union's Horizon 2020 Research and Innovation Programme, H2020 FORECEE under Grant Agreement No. 634570 and by the charity, The Eve Appeal (https://eveappeal.org.uk/). Martin Widschwendter was also supported by the European Union's Horizon 2020 European Research Council Programme, H2020 BRCA-ERC under Grant Agreement No. 742432. Emma J. Crosbie was supported by a National Institute for Health Research (NIHR) Advanced Fellowship (NIHR300650) and the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007).

CONFLICT OF INTEREST

James E. Barrett, Allison Jones, Iona Evans and Martin Widschwendter are named as inventors on a patent (submitted by UCLB, UCL's technology transfer organisation) which partly covers aspects described in this paper. James E. Barrett, Chiara Herzog and Martin Widschwendter are shareholders of Sola Diagnostics GmbH, which holds an exclusive licence to the intellectual property that protects the commercialisation of the WID-EC test. Laura Costas received supplies from Integrated DNA Technologies—IDT—and Roche Diagnostics at a 50% discount for a research project on endometrial cancer, and received a speaker's honoraria from Roche. Idibell and Roche signed a contract to collaborate in the development of the bioinformatics pipeline of a research project to early detect endometrial cancer. Laura Costas has received a competitive grant from Novosanis/European Association for Cancer Research—EARC—and received Colli-Pee devices (Novosanis) for a research project free of charge. Nicoletta Colombo has associations with Astra Zeneca, Clovis, Eisai, GSK, Immunogen, Mersana, MSD, Nuvation Bio, Onxerna, Pfizer, Pieris, Roche, and Novartis. All other authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The raw Illumina HumanMethylationEPIC BeadChip array data generated in our study have been uploaded to EGA under the accession number EGAS00001005033 (https://ega-archive.org/studies/EGAS00001005033). Other data that support the findings of our study are available from the corresponding author upon request.

ETHICS STATEMENT

The study is a substudy of the FORECEE (4C) Programme, which has ethical approval from the UK Health Research Authority (REC 14/LO/1633) and all other contributing centres. Participants for the self-sampling study were recruited in Bellvitge University Hospital and ASSIR Delta between 2017-2021 and signed an informed consent (ethical approvals PR128/16 and PR348/19). For the PMB cohort women provided written informed consent (ethical approval by the Joint University College London/University College London Hospital Committees on the Ethics of Human Research No. 06/Q0502/89).

ORCID

Chiara Herzog https://orcid.org/0000-0002-1572-498X
Emma J. Crosbie https://orcid.org/0000-0003-0284-8630
Line Bjørge https://orcid.org/0000-0002-0240-2770
Laura Costas https://orcid.org/0000-0003-2755-302X
Paula Peremiquel-Trillas https://orcid.org/0000-0002-8636-1725
Jiangrong Wang https://orcid.org/0000-0002-0001-503X
Karin Sundström https://orcid.org/0000-0002-6865-0224
Joakim Dillner https://orcid.org/0000-0001-8588-6506
Martin Widschwendter https://orcid.org/0000-0002-7778-8380

REFERENCES

1. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res*. 2014;74(11):2913-2921.
2. de Boer SM, Powell ME, Mileshkin 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*. 2019;20(9):1273-1285.
3. Wan YL, Beverley-Stevenson R, Carlisle D, et al. Working together to shape the endometrial cancer research agenda: the top ten unanswered research questions. *Gynecol Oncol*. 2016;143(2):287-293.
4. Badrick E, Cresswell K, Ellis P, Detecting Cancer Early Priority Setting Partnership Steering Group, Renehan AG, Crosbie EJ. Top ten research priorities for detecting cancer early. *Lancet Public Health*. 2019;4(11):e551.
5. Jacobs I, Gentry-Maharaj A, Burnell M, et al. Sensitivity of transvaginal ultrasound screening for endometrial cancer in postmenopausal women: a case-control study within the UKCTOCS cohort. *Lancet Oncol*. 2011;12(1):38-48.
6. Costas L, Frias-Gomez J, Guardiola M, et al. New perspectives on screening and early detection of endometrial cancer. *Int J Cancer*. 2019;145(12):3194-3206.
7. Moore L, Leongamornlert D, Coorens THH, et al. The mutational landscape of normal human endometrial epithelium. *Nature*. 2020;580(7805):640-646.
8. Wang Y, Li L, Douville C, et al. Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Sci Transl Med*. 2018;10(433):eaap8793.
9. Lac V, Nazeran TM, Tessier-Cloutier B, et al. Oncogenic mutations in histologically normal endometrium: the new normal? *J Pathol*. 2019;249(2):173-181.
10. Salk JJ, Loubet-Senear K, Maritschnegg E, et al. Ultra-sensitive TP53 sequencing for cancer detection reveals progressive clonal selection in normal tissue over a century of human lifespan. *Cell Rep*. 2019;28(1):132-44 e3.
11. Nair N, Camacho-Vanegas O, Rykunov D, et al. Genomic analysis of uterine lavage fluid detects early endometrial cancers and reveals a prevalent landscape of driver mutations in women without histopathologic evidence of cancer: a prospective cross-sectional study. *PLoS Med*. 2016;13(12):e1002206.
12. Jones A, Teschendorff AE, Li Q, et al. Role of DNA methylation and epigenetic silencing of HAND2 in endometrial cancer development. *PLoS Med*. 2013;10(11):e1001551.

13. Fiegl H, Gattringer C, Widschwendter A, et al. Methylated DNA collected by tampons—a new tool to detect endometrial cancer. *Cancer Epidemiol Biomarkers Prev*. 2004;13(5):882-888.
14. Doufekas K, Zheng SC, Ghazali S, et al. DNA methylation signatures in vaginal fluid samples for detection of cervical and endometrial cancer. *Int J Gynecol Cancer*. 2016;23(6):1050-1055. doi: 10.1097/IGC.0000000000000739
15. De Strooper LM, van Zummeren M, Steenbergen RD, et al. CADM1, MAL and miR124-2 methylation analysis in cervical scrapes to detect cervical and endometrial cancer. *J Clin Pathol*. 2014;67(12):1067-1071.
16. Huang RL, Su PH, Liao YP, et al. Integrated Epigenomics analysis reveals a DNA methylation panel for endometrial cancer detection using cervical scrapings. *Clin Cancer Res*. 2017;23(1):263-272.
17. Liew PL, Huang RL, Wu TI, et al. Combined genetic mutations and DNA-methylated genes as biomarkers for endometrial cancer detection from cervical scrapings. *Clin Epigenetics*. 2019;11(1):170.
18. Sangtani A, Wang C, Weaver A, et al. Combining copy number, methylation markers, and mutations as a panel for endometrial cancer detection via intravaginal tampon collection. *Gynecol Oncol*. 2020;156(2):387-392.
19. Barrett JE, Herzog C, Jones A, et al. The WID-BC-index identifies women with primary poor prognostic breast cancer based on DNA methylation in cervical samples. *Nat Commun*. 2022;13(1):449.
20. Barrett JE, Herzog C, Kim YN, et al. Susceptibility to hormone-mediated cancer is reflected by different tick rates of the epithelial and general epigenetic clock. *Genome Biol*. 2022;23(1):52.
21. Barrett JE, Jones A, Evans I, et al. The DNA methylome of cervical cells can predict the presence of ovarian cancer. *Nat Commun*. 2022;13(1):448.
22. Herzog C, Marin F, Jones A, et al. A simple cervicovaginal epigenetic test for screening and rapid triage of women with suspected endometrial cancer: validation in several cohort and case/control sets. *J Clin Oncol*. 2022;40(33):3828-3838.
23. Schreiberhuber L, Herzog C, Vavourakis CD, et al. The WID(TM)-qEC test—performance in a hospital-based cohort and feasibility to detect endometrial and cervical cancers. *Int J Cancer*. 2022:1-6. doi:10.1002/ijc.34275.
24. Widschwendter M, Jones A, Evans I, et al. Epigenome-based cancer risk prediction: rationale, opportunities and challenges. *Nat Rev Clin Oncol*. 2018;15(5):292-309.
25. Widschwendter M, Fiegl H, Egle D, et al. Epigenetic stem cell signature in cancer. *Nat Genet*. 2007;39(2):157-158.
26. Barrett JSK, Jones A, Evans I, et al. The WID-CIN test identifies women with, and at risk of, cervical intraepithelial neoplasia grade 3 and invasive cervical cancer (CIN3+). *Genome Med*. 2022;14(1):1-11.
27. Perskvist N, Norman I, Eklund C, Litton JE, Dillner J. The Swedish cervical cytology biobank: sample handling and storage process. *Biopreserv Biobank*. 2013;11(1):19-24.
28. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol*. 2016;31(2):125-136.
29. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 2016;17(1):208.
30. Zheng SC, Webster AP, Dong D, et al. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics*. 2018;10(7):925-940.
31. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
32. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N. Endometrial cancer. *Lancet*. 2022;399(10333):1412-1428.
33. Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. *Lancet*. 2016;387(10023):1094-1108.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Barrett JE, Jones A, Evans I, et al. The WID-EC test for the detection and risk prediction of endometrial cancer. *Int J Cancer*. 2023;152(9):1977-1988. doi:10.1002/ijc.34406