CANCER GENETICS AND EPIGENETICS

INTERNATIONAL
JOURNAL of CANCER

HPV-induced host epigenetic reprogramming is lost upon progression to high-grade cervical intraepithelial neoplasia

Chiara Herzog[1,2] | Charlotte D. Vavourakis[1,2] | James E. Barrett[1,2] |
Gerlinde Karbon[3] | Andreas Villunger[3] | Jiangrong Wang[4] |
Karin Sundström[4,5] | Joakim Dillner[5] | Martin Widschwendter[1,2,6,7]

[1]European Translational Oncology Prevention and Screening (EUTOPS) Institute, Universität Innsbruck, Hall in Tirol, Tirol, Austria

[2]Institute for Biomedical Aging Research, Universität Innsbruck, Innsbruck, Tirol, Austria

[3]Institute for Developmental Immunology, Biocenter, Medical University of Innsbruck, Innsbruck, Austria

[4]Department of Laboratory Medicine, Division of Pathology, Karolinska Institutet, Stockholm, Sweden

[5]Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden

[6]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

[7]Department of Women's Cancer, UCL EGA Institute for Women's Health, University College London, London, UK

Correspondence
Martin Widschwendter, Universität Innsbruck, European Translational Oncology Prevention and Screening (EUTOPS) Institute, Milser Straße 10, 6060 Hall in Tirol, Austria.
Email: martin.widschwendter@uibk.ac.at

Abstract

The impact of a pathogen on host disease can only be studied in samples covering the entire spectrum of pathogenesis. Persistent oncogenic human papilloma virus (HPV) infection is the most common cause for cervical cancer. Here, we investigate HPV-induced host epigenome-wide changes prior to development of cytological abnormalities. Using cervical sample methylation array data from disease-free women with or without an oncogenic HPV infection, we develop the WID (Women's cancer risk identification)-HPV, a signature reflective of changes in the healthy host epigenome related to high-risk HPV strains (AUC = 0.78, 95% CI: 0.72-0.85, in nondiseased women). Looking at HPV-associated changes across disease development, HPV-infected women with minor cytological alterations (cervical intraepithelial neoplasia grade 1/2, CIN1/2), but surprisingly not those with precancerous changes or invasive cervical cancer (CIN3+), show an increased WID-HPV index, indicating the WID-HPV may reflect a successful viral clearance response absent in progression to cancer. Further investigation revealed the WID-HPV is positively associated with apoptosis ($\rho = 0.48$; $P < .001$) and negatively associated with epigenetic replicative age ($\rho = -0.43$; $P < .001$). Taken together, our data suggest the WID-HPV captures a clearance response associated with apoptosis of HPV-infected cells. This response may be dampened or lost with increased underlying replicative age of infected cells, resulting in progression to cancer.

KEYWORDS
apoptosis, cervical cancer, DNA methylation, human papillomavirus, replicative aging

What's new?

HPV infection is the most common cause of cervical cancer. Studies have found that cervical cancer changes DNA methylation (DNAme) in viral and patient DNA collected from cervical samples. Here, the authors looked for epigenome-wide changes in patient DNA caused by HPV infection that occur before cytological abnormalities could be detected. They created a DNA methylation signature called the WID-HPV, which reflects activation of a host defense response that helps prevent disruption of cell proliferation and promotes apoptosis. Women with CIN1 and CIN2 showed increases in the WID-HPV, while those with CIN3+ had WID-HPV similar to HPV negative samples.

1 | INTRODUCTION

Microbes, including viruses, bacteria, and fungi, are frequently involved in the development of chronic disease, importantly cancer. The impact of microbes on cancer formation has often been assessed using samples from healthy and diseased individuals but studying samples from individuals with and without a microbial infection at different stages of disease progression could improve our understanding of how host-microbe interactions drive disease progression. Owing to cervical cancer screening strategies in place in developed countries, cervical cancer is one of the few microbial-mediated human diseases for which samples from the site of origin are readily available prior to the onset of disease.

Cervical cancer is the fourth most common cancer in women worldwide.[1] The disease progresses slowly through stages of cervical intraepithelial neoplasia (CIN grade 1-2), most of which are likely to regress, with only a small portion of women further developing precancerous CIN3 or invasive cervical cancer (CC). Most cases are caused by persistent infection with oncogenic human papilloma virus (HPV).[2] HPV infects epithelial cells in the basal layer of the cervical epithelium and may ultimately overexpress the viral oncogenic proteins E6 and E7 that promote cellular proliferation, prolong the cell cycle, and prevent apoptosis of infected cells, leading to cervical cancer (reviewed by reference 3). Importantly, the HPV virus adopted a low-rate replication strategy tightly synchronized to the developmental stages in the maturing of epithelial cells, allowing it to effectively evade the host immune system and persistently infect epithelial tissues. Overexpression of the oncogenic proteins E6 and E7 is an exclusive feature of high-grade lesions. It remains debated what triggers this viral switch in expression, but it seemingly co-occurs with integration of the virus into the host genome and disruption of specific viral genes (reviewed by reference 4).

Previous studies have investigated aberrant methylation (DNAme) of host or viral DNA in cervical samples from healthy women and women with cervical cancer, revealing that DNAme analysis is suitable for detecting precancerous lesions and invasive cancer.[5-12] E6 and E7 have been shown to elicit substantial changes in the host epigenome which may contribute to viral immune evasion, genomic instability, and immortalization of malignant cells (reviewed by reference 13). However, relatively little is known about DNAme changes induced upon initial HPV infection in the genomes of women with no or low-grade lesions, when the virus is not (yet) overexpressing these oncoproteins.

Recently, a large-scale tumor sequencing study investigating viral pathogens found that key mutations were lacking specifically in HPV-positive head and neck cancers, and alternatively proposed impaired antiviral defense mechanisms as key driver forces for cervical, bladder, and head and neck cancers.[14] A separate study found a higher mutation burden in HPV16 genomes induced by the antiviral restriction factor APOBEC3 in control vs precancerous/cancer cases, suggesting that impaired viral clearance by the host innate immune response may drive progression to cancer.[15] Putatively, early E6/E7-independent host epigenetic responses may hold additional clues as to how in most women HPV infection is naturally cleared and why few women progress to malignant disease.

Here, we revisited Illumina EPIC methylation array data probing DNAme at ~850.000 CpGs available from a cervical screening cohort[12] to dissect apart the initial host epigenetic response to an oncogenic HPV infection from epigenetic changes that appear upon progression to cervical cancer. This unique dataset comprises 1040 cervical screening samples from HPV negative (HPV−) or high-risk HPV (hrHPV) positive (HPV+) healthy women lacking cervical cytological abnormalities, women with low-grade cervical intraepithelial neoplasia (CIN1 or 2), and women high-grade CIN or invasive cervical cancer (CIN3+; Table S1). We used the subgroup of healthy women to develop the WID-HPV (*W*omen's cancer risk *id*entification-*HPV*), a DNAme-based signature that captures hrHPV infection status and proposedly the initial host defense response. We further linked this host defense with cell proliferation and apoptosis at different stages of CIN, applying pcgtAge,[16] a previously developed epigenetic mitotic clock indicative of replicate age, and a new apoptosis DNAme signature, derived from a separately generated in vitro dataset.

2 | MATERIALS AND METHODS

2.1 | Cytology data set

We collated a cohort-based nested-control dataset and a diagnostic validation set used previously to construct and validate a DNAme classifier predicting CIN3+ from cervical samples.[12] This combined

dataset is unique in that it entails both associated metadata on HPV status and HPV type, and that it covers the entire spectrum of CIN. Human cervical samples used in our study were obtained from the Stockholm Cervical Cytology Bank. Briefly, 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.[17] Since 2013, for virtually all of the ~150 000 liquid-based cytology samples collected in PreservCyt each year, 600 μL are biobanked in a 96 well format at −27°C. This allows for preservation of intact cells and subsequent analyses of DNA, RNA, and protein content, among others. The samples in the biobank can be linked to the Swedish health register infrastructure for cytology/HPV results, histopathology test and results, as well as cervical cancer diagnoses, through the individually unique personal identification number (PIN).[18] High-performance HPV testing and type determination as part of the screening program was conducted using the cobas 4800 HPV test,[19] which detects following hrHPV strains: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68 (Roche Diagnostics).

We selected biobanked cervical samples from women resident in Stockholm that participated in cervical screening or clinically indicated testing during the years 2013-2015 (Figure S5). Participants data were linked to the National Cancer Register at the Swedish National Board of Health and Welfare, and the Swedish National Cervical Screening Registry, to identify all current cases of CIN3/Adenocarcinoma in situ (AIS) or invasive cervical cancer (CIN3+), as well as low-grade lesions (CIN1/2). Selected samples were shipped to UCL and DNA was isolated with the Nucleo-Mag Blood 200 μL kit (Macherey Nagel, #744501.4) on a Hamilton Star liquid handling platform. Concentration and quality absorbance ratios were measured using a Nanodrop-8000 (Thermo Scientific, Inc) and the extracted DNA was stored at −80°C until further analysis.

2.2 | Induction of apoptosis with BH3 mimetics in cancer cell lines

The BH3 mimetic dataset was generated from four cancer cell lines: HelaS3 (CVCL_0058; DSMZ ACC 161), Nalm6 (CVCL_0092; DSMZ ACC 128), HT1080 (CVCL_0317; DSMZ ACC 315), Cal51 (CVCL_1110; DSMZ ACC 302) treated with a combination of the BH3 mimetics S63845 (1 μM; Selleck Chemicals, S8383) and ABT-737 (1 μM; Selleck Chemicals, S1002), DMSO was used as control, for indicated time points. All human cell lines have been authenticated using STR profiling within the last 3 years. All experiments were performed with mycoplasma-free cells. Cells were kept at 37°C and 5% CO_2. Cal51 and HeLaS3 were cultured in DMEM (Sigma Aldrich, D5671) supplemented with 10% FCS (Invitrogen, 10 270 106), 1% L-Glutamin (Sigma G7513), and 100 U/mL Penicillin, and 100 μg/mL Streptomycin (Lonza 17-602 E). Nalm6 were grown in RPMI-1640 (Sigma Aldrich, R0883) supplemented with 10% FCS, 1% L-Glutamin, 100 U/mL Penicillin, and 100 μg/mL Streptomycin. HT1080 were grown in MEM (9.6 g/L; Sigma, M0643) and Sodium Bicarbonate (1.5 g/L; Sigma S5761) solved in A.d., supplemented with 10% FCS,

100 U/mL Penicillin, and 100 μg/mL Streptomycin. Cells were harvested, washed with PBS and either frozen in FCS (Invitrogen, 10270106) with 10% DMSO or analyzed for cell death. For the latter cells were washed after treatment with PBS and resuspended in PBS with propidium iodide (1 μg/mL; Sigma-Aldrich, 81 845). Samples were measured with an Attune Nxt flow cytometer (Thermo Fisher Scientific, Waltham) and analyzed with FlowJo software (VersionX, FlowJo LLC, Ashland) (Figure S3). DNA was isolated from frozen samples with the Allprep DNA/RNA/protein mini kit (Qiagen, #80004). For controls, isolates from each cell line were combined. DNA was further cleaned up and concentrated with the Genomic DNA Clean & Concentrator-10 kit (Zymo, #D4011). Concentrations were measured with the QuantiFluor dsDNA System (Promega, #E2670), and the extracted DNA was stored at −80°C until further analysis.

2.3 | Illumina EPIC methylation arrays

DNA methylation analysis and quality control of the raw data was performed using a standardized pipeline as described previously.[20,21] In brief, between 300 and 500 ng DNA were bisulfite-modified (Zymo Research Corp., EZ DNA Methylation-Lightning Automation, #D5049), of which 200 ng were evaluated using the Illumina Human MethylationEPIC methylation profiling array (Illumina, # 20042130). Quality control was performed using the R package minfi, version 1.40.0.[22] Low quality samples, that is, those with a median methylated or median unmethylated intensity <9.5 and/or > 10% failed probes (detection P-value >.01), were removed before background intensity and dye bias correction. BMIQ normalization[23] was performed using the champ. norm function within the R package ChAMP, version 2.24.0.[24,25] Visual inspection of the beta distribution was performed for all samples. Additionally, 537 probes that mapped to the Y chromosome, 82 108 SNPs flagged by reference [26], 6102 SNPs flagged by references [20,21], 2932 non-CpG probes, and probes that failed in >10% of the samples were removed. Remaining beta-values with a detection P-value >.01 were removed and imputed using the R package impute, version 1.68.0,[27] leaving a total of 777 005 CpG probes in the cleaned beta matrix. For each sample the fraction of immune cells and cell subtypes were estimated using the HepiDISH algorithm.[28] CpGs in final indices were annotated using the IlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package.[29]

2.4 | WID-HPV index development

To construct the linear WID-HPV index, diagnostic for the presence of an HPV infection, samples from cytology negative (cyt−) women within the cytology data set were divided into a subset for training (2/3: n = 146 HPV-, n = 202 HPV+) and a holdout test data set for validation (1/3: n = 68 HPV-, n = 111 HPV+). The top CpGs discerning HPV+ and HPV− samples were identified via the following approaches: (i) biggest difference in epithelial cells (n = 10 000, ranked by the difference in intercepts for ic = 0 in the linear model cg ~ HPV

+ age + ic), (ii) biggest difference in immune cells (n = 10 000, ranked by the difference in intercepts for ic = 1 in the linear model cg ∼ HPV + age + ic), (iii) merged epithelial and immune difference (n = 10 000, ranked by P-value of linear model cg ∼ HPV + age + ic), (iv) biggest overall difference, regardless of cell type and age (ranked by F-statistic; HPV status ∼ cg methylation). For each feature selection approach, model training was conducted using the R package glmnet, version 4.1.4. Lasso, Ridge, and Elastic Net regularized logistic regressions were performed with 10-fold cross validation to tune the hyperparameter λ, gradually adding more input features (CpG). The best performing classifier was manually selected using the out-of-bag estimate of the area under the ROC curve (top 500 CpGs with the biggest epithelial difference, L1 regularization) and returned 27 nonzero coefficients. For further analyses, the resulting linear index, the WID-HPV, was adjusted for chronological age and inferred immune cell composition of the sample. Classification performance was tested in the holdout test set.

2.5 | Gene ontology analysis

To identify pathways affected by DNAme in HPV+ women compared to HPV-, we conducted gene ontology enrichment analysis on the differentially methylated probes (DMPs) that had a P-value <.05 after FDR adjustment corresponding to the response variable HPV in the linear regression model (cg ∼ HPV + age + ic). We used the gometh function from the R package missMethyl, version 1.28.0[30] to compare the ontology of the DMPs against all probes that passed QC as a background. Significantly enriched GO terms, and the genes and methylation status of the CpGs within, were visualized using modified source code from the package GOplot[31] and with the package ComplexHeatmap version 2.12.0.[32]

To identify pathways covered by the subset of DMPs in the WID-HPV, we compared ontology of the index CpGs against all probes that passed QC as a background, only keeping terms that had a P-value <.05 without FDR adjustment and two or more gene hits. The package GOxploreR version 1.2.6[33] was used to remove intermediate-level terms in connected paths of a simplified DAG, retaining only highest and lowest level terms for each combination of genes.

2.6 | Inference of replicative age

To estimate the proportion of cells that have accumulated a high replicative age, we calculated a pcgtAge score, a previously established epigenetic mitotic clock evaluating methylation at 385 Polycomb group target (PCGT) promoter CpGs.[16]

2.7 | Epigenetic evaluation of apoptosis

The list of *Homo sapiens* Hallmark apoptosis genes was retrieved from MSigDB[34] using the R msigdbr package,[35] version 7.5.1, querying for the gene set ID "M5902." Median methylation at promoters of these apoptosis-related genes was calculated using the median methylation of

CpGs located within 1500 bp of the transcription start site (TSS200 and TSS1500). Out of 161 genes in the Apoptosis Hallmark set, 103 were represented in the final list by CpGs within 1500 bp of the transcription start site on the Illumina Human MethylationEPIC array.

For development of the apoptosis index, samples from the BH3 mimetics dataset were split into training (n = 9 samples) and holdout test sets (n = 3 samples, including 1 control cell line). Elastic Net logistic regression with 3-fold cross validation was performed, gradually adding more features from four different subsets: (i) top 10 000 CpGs with the largest absolute mean delta-beta between control cell lines and apoptotic cell lines, (ii) top 10 000 CpGs that were significantly differentially methylated after linear regression with the largest absolute mean delta-beta between control and apoptotic cell lines, (iii) top 10 000 CpGs that were significantly differentially methylated after linear regression with the largest absolute mean delta-beta between control and apoptotic cell lines, but were not significantly different between the immune cell vs epithelial cell lines, (iv) significantly differentially methylated CpGs as determined by limma function from the R package ChAMP, version 2.24.0. The best performing classifier was chosen manually from performance profiles based on slope, intercept and mean deviance of the logistic regression fit. The final classifier, termed apoptosis index, used a pool of 3000 input CpGs from subset (ii), resulting in 114 nonzero regression coefficients. The apoptosis index was further tested on a publicly available Illumina Human MethylationEPIC dataset from high grade serous ovarian cancer cell lines in which apoptosis was induced.[36]

2.8 | Statistical analysis

All statistical analyses were conducted in R, version 4.1.2. WID-HPV and pcgtAge were adjusted for age and inferred immune cell composition using residuals from linear models: index/clock ∼ age + ic.

3 | RESULTS

3.1 | The WID-HPV captures substantial HPV-induced epigenetic reprogramming in healthy women

We sought to characterize host epigenetic changes following an oncogenic HPV infection in healthy individuals, with the aim to develop a DNAme-based classifier for HPV infection allowing us to investigate whether early epigenetic changes in the host genome are retained during disease progression (Figure 1A). We aimed to specifically identify methylation changes associated with hrHPV but not CIN or cervical cancer, and therefore excluded samples from women with cytological alterations in this initial step. The EPIC methylation array data from HPV+ and HPV− individuals without CIN were split into a training and a test set (70% and 30% of the samples). In the training set, we identified 8386 significantly differentially methylated positions (DMPs, P < .05 after FDR correction), indicating substantial epigenetic changes upon viral infection in healthy individuals. The DMPs were enriched pathways for DNA
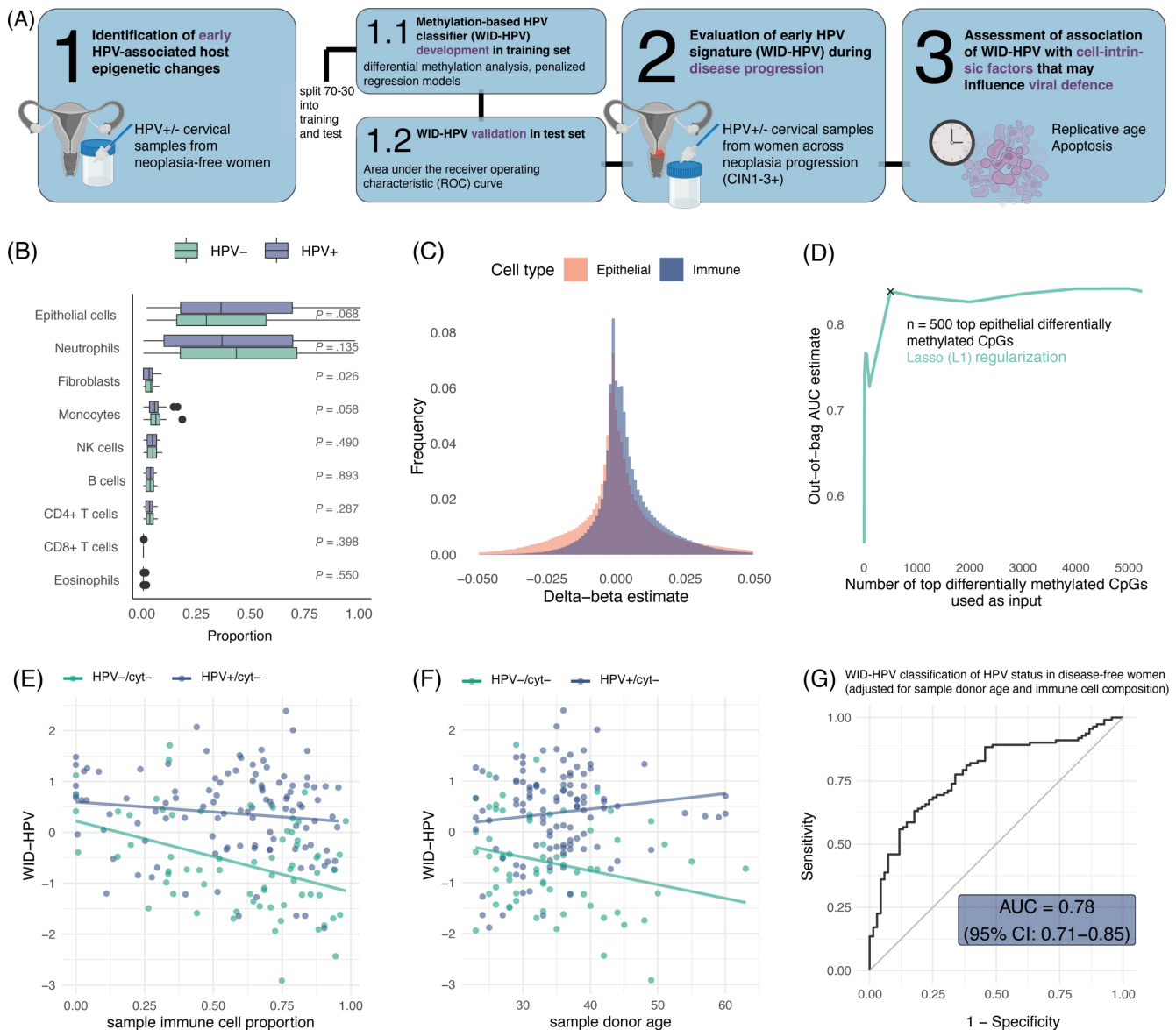
FIGURE 1 Study overview and WID-HPV training and evaluation. (A) Study overview. Early HPV-associated host epigenome changes in the Cervix uteri were evaluated in a sample set of (cervical) disease-free, cytology-negative HPV+ or HPV− samples. Samples were split 70-30 into a training and test set and differential methylated probes determined. A HPV signature, called WID-HPV (Women's cancer risk-human papillomavirus) was derived using penalized regression models. Performance was validated in the holdout test set. The WID-HPV was then further evaluated in association with cervical intraepithelial neoplasia progression (CIN1-3+), and its relation with cell-intrinsic factors that may influence defense mechanisms was investigated via the association with replicative age and apoptosis signatures. (B) Sample cell type composition of HPV+ and HPV− cytology negative samples. (C) Histogram of differential methylation in epithelial and immune cells, the major cell subtypes in cervical samples. (D) Out-of-bag area under the curve (AUC) estimate for the receiver operating characteristic during training of the linear WID-HPV index using penalized regression. Providing the top 500 differentially methylated epithelial CpGs as input for an L1 norm classifier (Lasso) resulted in the highest out-of-bag AUC. (E) WID-HPV vs sample immune cell proportion in the test set. (F) WID-HPV vs sample donor age in the test set. (G) Receiver operating characteristic and area under the curve (AUC) for the WID-HPV (adjusted for sample donor age and immune cell proportion, using HPV− cyt− samples) in the test set.

damage repair and chromosomal organization (Figure S1). DNAme is strongly cell type dependent, and hence DMPs may arise from changes in underlying sample composition rather than genuine methylation changes. The estimated proportion of epithelial and immune cells was broadly similar across HPV+ and HPV− samples, with a small but significant alteration in the proportion of fibroblasts (Figure 1B). Given that fibroblasts make up less than 10% of cells in cervical samples, we assumed that the observed DNAme changes were not primarily caused by differences in cell type composition after infection.

The final WID-HPV index is a linear combination of DNAme at 27 CpG sites selected from the DMPs that are either hypo- or hyper-methylated in HPV positive women in the test set (Table 1, Figure 1C,D).

TABLE 1   Annotation of the 27 WID-HPV-index CpGs.

Name	Chromosome	Gene	Infinium type	CpG context	Gene context	db_epithelial
cg05105016	chr15	PIAS1	II	OpenSea	Body	0.091
cg18773260	chr17	HOXB7	II	Island	Body	0.091
cg16901621	chr8		II	OpenSea	IGR	0.098
cg19770292	chr5		I	OpenSea	IGR	0.089
cg26296769	chr6	KLC4	II	N_Shore	TSS1500	0.089
cg25092881	chr10	NRG3	II	OpenSea	Body	0.098
cg00779638	chr9	AGTPBP1	II	N_Shore	5′UTR	0.088
cg03499324	chr9	PSMD5-AS1	II	S_Shore	Body	0.109
cg14683065	chr10	LRRC27	II	S_Shelf	Body	0.139
cg04308838	chr18	TTC39C	II	OpenSea	Body	0.103
cg11952595	chr13	ARL11	I	OpenSea	TSS1500	0.091
cg15074838	chr6	HLA-DRA	II	OpenSea	TSS1500	0.099
cg14804593	chr4		II	OpenSea	IGR	0.091
cg05840412	chr13		II	OpenSea	IGR	0.103
cg08355045	chr6		II	OpenSea	IGR	0.089
cg19045773	chr6		II	OpenSea	IGR	−0.091
cg22139830	chr2		II	OpenSea	IGR	−0.088
cg13893555	chr9	FBXO10	II	N_Shore	5′UTR	−0.136
cg12284518	chr11	VEGFB	II	N_Shelf	Body	−0.110
cg19585676	chr6	HLA-A	II	S_Shore	3′UTR	−0.111
cg26955290	chr17	ASGR2	II	OpenSea	Body	−0.095
cg12770425	chr6		II	N_Shelf	IGR	−0.092
cg23265105	chr5	PCDHB7	II	N_Shore	1stExon	−0.087
cg10473907	chr6	HCG27	II	S_Shore	Body	−0.088
cg02767777	chr11		II	N_Shelf	IGR	−0.090
cg15071481	chr18	KATNAL2	II	N_Shore	TSS1500	−0.093
cg18922214	chr17	SLC26A11	II	OpenSea	Body	−0.102

Note: CpGs were annotated using the IlluminaHumanMethylationEPICanno.ilm10b4.hg19 R package (Bioconductor). db_epithelial is the mean delta-beta of the epithelial component (inferred immune cell composition = 0) in HPV− compared to HPV+ cytology negative samples in the training set. The mean delta-beta is positive when the respective CpG is on average hypermethylated in the HPV+ group, negative when on average hypomethylated.

For subsequent analyses, the WID-HPV was adjusted for chronological age and inferred immune cell composition of the sample, to account for small changes other than HPV status that might influence the WID-HPV index (Figure 1E,F, see Methods for details). As a classifier discriminating for HPV status, the adjusted WID-HPV achieved an AUC of 0.78 in the holdout test set (95% CI: 0.72-0.85; Figure 1G).

Some of the selected DMPs that constitute the final WID-HPV are in genes that putatively play a role in the clearance of infected cells following the initial host immune response to the HPV infection (Tables 1 and S2). Altered DNA methylation in the human leukocyte antigen system may affect antigen processing and presentation (*HLA-DRA* gene) and downstream T-cell mediated apoptosis (*HLA-A* gene), and several variants of the HLA-locus have been previously linked to cervical cancer susceptibility (reviewed by reference 37). Additionally, altered DNA of the *PIAS1* gene may affect sumoylation of the latent cytoplasmic transcription factor STAT1 and in this way modulate the host's innate immune response.[38]

3.2 | HPV-induced epigenetic reprogramming is lost upon progression to cancer and the WID-HPV is inversely associated with replicate age

We next evaluated the WID-HPV in samples from women diagnosed with CIN1-3+, the majority of which were also positive for oncogenic (high-risk) HPV strains (251/257, 97.7%; Figure 2A, Table S1). As expected, the WID-HPV was significantly increased in women with CIN1 and CIN2, but surprisingly exhibited scores not significantly different from HPV− samples without cytological alterations in samples from women with CIN3+, including 16 women diagnosed with adenocarcinoma in situ or invasive cervical cancer (Figure S2A). Although the proportion of HPV16 and/or HPV18 positive samples was far larger in the CIN3+ group compared to those with CIN2 and lower (Figures 2B and S2B), the unexpected drop in the WID-HPV for CIN3+ was independent of these oncogenic HPV strains (Figures 2C and S2C).
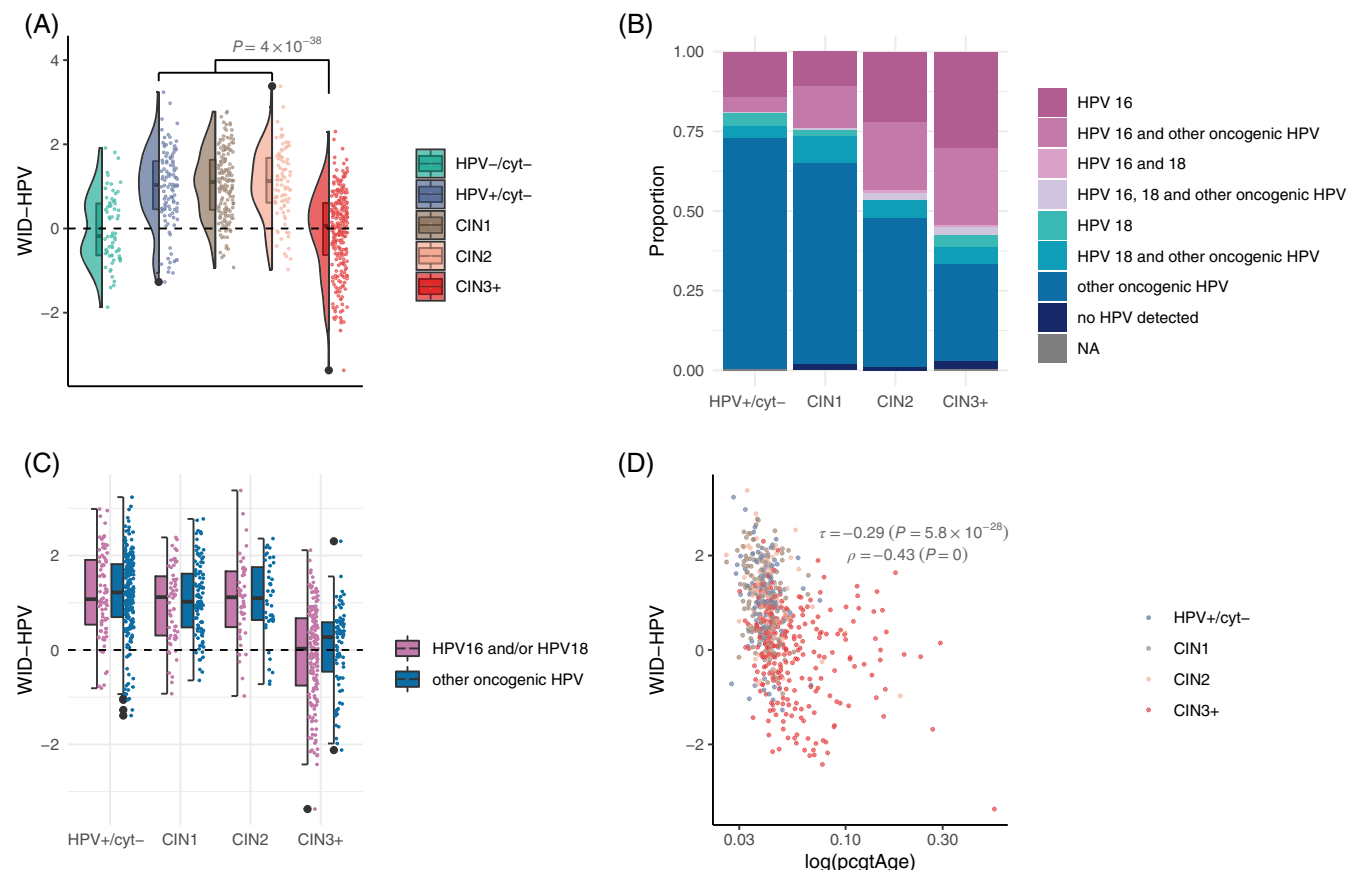
FIGURE 2 WID-HPV in relation to CIN, HPV oncogenic subtype and replicative age. (A) Overlayed dot-, box- and violin plots showing WID-HPV in the groups of different samples. Samples used for index training were excluded. Box plots show median and interquartile range, with whiskers showing 1.5 times the 25th and 75th percentile and dots the outliers. Violin plots give the kernel probability density of the data in each group. P value is derived from Wilcoxon test comparing all HPV+ cytology negative (cyt−) samples, CIN1, and CIN2 samples vs CIN3+. P values for difference to HPV− cyt− samples (reference) were as follows: HPV+, $P = 2 \times 10^{-9}$; CIN1, $P = 1.7 \times 10^{-19}$; CIN2, $P = 3.1 \times 10^{-11}$; CIN3, $P = 1$ (not shown in the figure). P values were adjusted for multiple comparisons using the Bonferroni method. (B) Proportion of oncogenic HPV subtypes detected by qPCR for each sample group defined in the CIN dataset. Samples used for training the WID-HPV index were excluded. Other oncogenic HPV subtypes detected were 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, or a combination thereof. (C) Performance of the WID-HPV in HPV16/18 positive samples vs other oncogenic HPV genotypes in the CIN dataset. Samples used for index training and those where no HPV was detected were excluded. Box plots give median and interquartile range, with whiskers showing 1.5 times the 25th and 75th percentile and dots the outliers. (D) Kendall (τ) and Spearman (ρ) correlations of the WID-HPV-index with the logarithm of the pcgtAge score for the subset of HPV+ women with respect to CIN.

Given this surprising finding, we evaluated how the WID-HPV is associated with replicative age, a factor that has previously been proposed to indicate cancer risk. We observed a moderate but significant inverse correlation between pcgtAge,[16] a marker of stem/progenitor cell's replicative age, and the WID-HPV in all HPV-positive women, suggesting that for this group, a low WID-HPV is correlated with increased cellular replicative aging and cancer risk (Figures 2D and S2D).

3.3 | The WID-HPV correlates with DNAme based apoptotic signatures

To investigate the possible link between WID-HPV and apoptosis as an intrinsic defense mechanism clearing infected cells, we next

evaluated signatures of apoptosis in cervical samples from HPV-/HPV+ controls and women with CIN1-3+ (Figure 3A). Promoter methylation at MSigDB hallmark genes associated with apoptosis (M5902)[34] was decreased in HPV+ and CIN1/2 samples compared to controls, but not in CIN3+ samples. Higher and lower methylation levels of promoter sites are frequently associated with reduced and increased gene transcription, respectively. This indicated potentially higher expression levels of apoptosis-related genes in HPV+, CIN1, and CIN2 samples, with epigenetic silencing of apoptotic pathways in CIN3 and cervical cancer (Figures 3B and S4A).

Promoter methylation only provides indirect evidence of function, hence we more closely investigated epigenetic changes linked with apoptosis by treating four different cancer cell lines with BH3 mimetics, a powerful proapoptotic stimulus. We developed a linear index from the DMPs with the largest mean absolute difference in methylation in controls vs
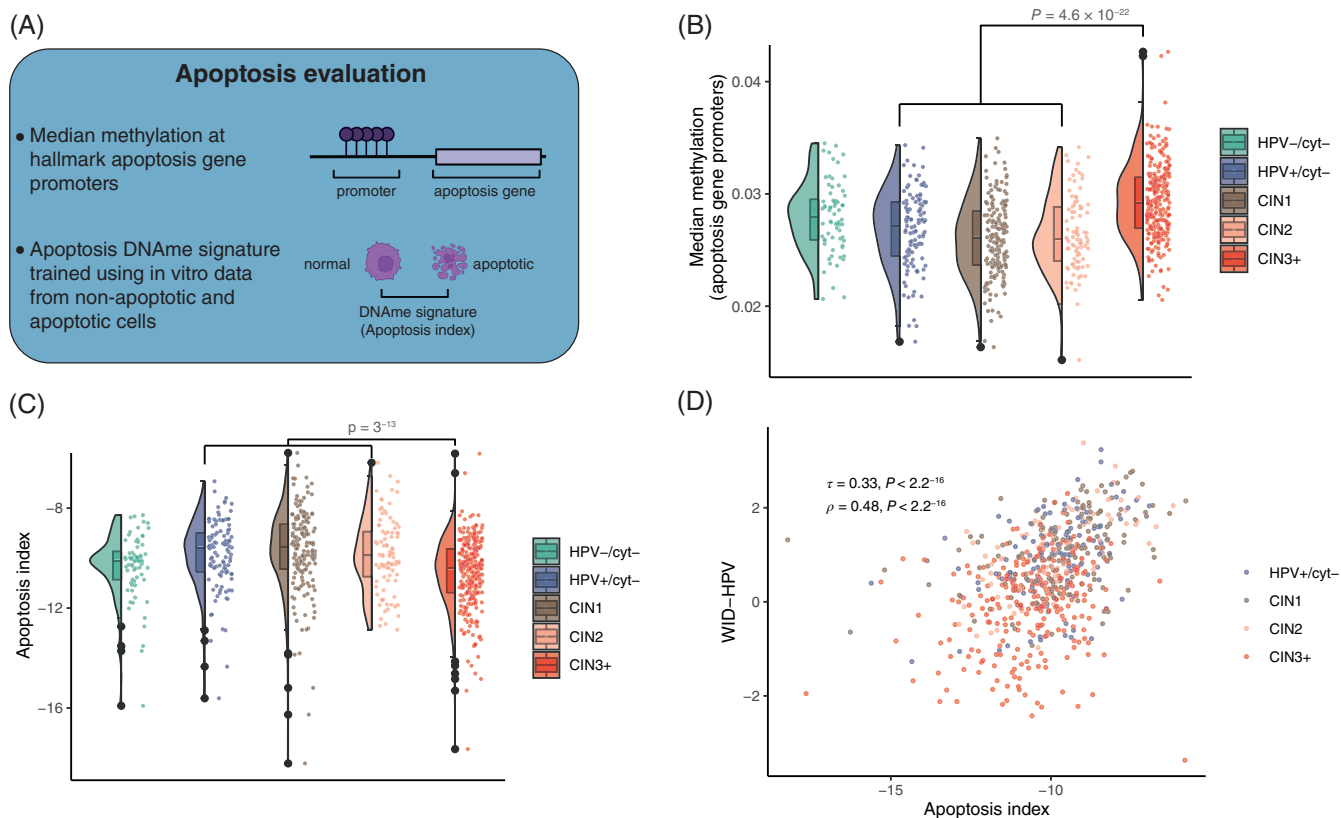
FIGURE 3 WID-HPV correlation with apoptosis. (A) Overview methods for evaluating apoptosis. (B) Overlayed dot-, box- and violin plots showing median methylation of CpGs in promotor regions from the Apoptosis Hallmark gene set (MSigDB) vs HPV and cytology status in the CIN dataset. *P* value derived from Wilcoxon test comparing HPV+, CIN1, and CIN2 with CIN3+ samples. (C) Overlayed dot-, box- and violin plots showing the apoptosis index constructed using BH3 mimetics on cancer cell lines vs HPV and cytology status in the CIN dataset. *P* value derived from Wilcoxon test comparing HPV+, CIN1, and CIN2 with CIN3+ samples. (D) Kendall (τ) and Spearman (ρ) correlations of the WID-HPV-index with the apoptosis-index for the subset of women that were HPV+ or had CIN1-3+. Box plots give median and interquartile range, with whiskers showing 1.5 times the 25th and 75th percentile and dots the outliers. Violin plots give the kernel probability density of the data in each group. Samples used for training the WID-HPV were excluded from all plots.

apoptotic cells across the four cell lines (Figure S3; Table S3, see Methods for details). Similar to the WID-HPV, the resulting apoptosis index, probing 114 CpGs, was elevated in all HPV+ sample groups compared to the HPV− control group except for samples derived from women with CIN3+ (Figures 3C and S4B), and there is a significant moderate-to-strong correlation between the WID-HPV and the apoptosis index (Figures 3D and S4C). This corroborates the hypothesis that the WID-HPV index captures DNAme changes in the host genome associated with the activation of apoptosis upon initial HPV infection and that this apoptotic defense mechanism is impaired upon progression to cervical cancer.

4 | DISCUSSION

Our data are consistent with a model suggesting that epigenetic reprogramming and apoptosis of infected epithelial cells contributes to the clearance of the virus in HPV infections. The WID-HPV index captures host epigenetic responses to an HPV infection which we hypothesize to contribute to or be a result of, at least in part, the

prevention of cell proliferation deregulation and the induction of programmed cell death of infected cells in initial disease stages. Tissues that have accumulated cells with a high replicative age, as evidenced from a high pcgtAge score in a cervical sample, could exhibit a compromise in certain antiviral responses such as HPV-mediated apoptosis. Impaired antiviral responses may contribute to HPV persistence, development of high-grade CIN, and ultimately progression to cervical cancer.

An important limitation of our findings is that our approach does not provide mechanistic insights into the biology of the host upon HPV infection and progression to cervical cancer, nor can we distinguish cause from consequence. For example, reduced clearance of infected cells can lead to the accumulation of cells with higher replicative age in HPV-infected tissue. On the other hand, like chronological age, higher replicative cellular aging has been suggested to result in reduced priming and refractoriness for apoptosis.[39] It is not yet clear which factors contribute to a high replicative age, but these could include additional risk factors that have previously been suggested to accelerate aging, for instance smoking.[40] Moreover, it is not clear

whether an impaired viral host defense occurs gradually, or rather as an epigenetic field defect. Furthermore, the majority of the detected DMPs here comprise open sea CpGs in gene bodies or noncoding genomic regions. It is well understood that DNA methylation of CpG islands in classical promotor regions may repress gene transcription, but we are only beginning to appreciate the link between gene expression and DNA methylation of CpGs outside of this well-studied context, including CpGs in noncoding genomic elements (eg, enhancers) and their importance for cancer.[41,42] Hence, the functional downstream consequences of alterations observed in DNA methylation at these sites cannot be inferred without further studies. Future studies considering (i) the viral load, (ii) specific cell subtypes rather than bulk tissue (ie, differentiated or stem/progenitor epithelial cells and immune cells), (iii) sequential samples from women with different HPV clearance, and (iv) functional work considering the impact of HPV-infection of epithelial cells with different replicative ages can enable further elucidation of which factors inherent to epithelial cells contribute to HPV-mediated carcinogenesis.

Taken together, we show that substantial DNAme changes in the host genome occur early upon infection with hrHPV, when the virus is not yet overexpressing E6/E7 oncoproteins and no cytological abnormalities in the cervix have formed. As it remains unknown why most women achieve natural clearance of infection while others experience persistence and formation of cancerous lesions, these early DNAme changes warrant further investigation. We conclude that HPV triggers an initial host defense response for which associated DNAme changes in the host genome are captured by the WID-HPV. Independent of the oncogenic HPV subtype, an impaired host defense concurs with increased replicative aging, reduced apoptosis-mediated clearance of HPV-infected cells, and the subsequent development of viral-associated cervical malignancy.

AUTHOR CONTRIBUTIONS

The work reported in the paper has been performed by the authors, unless clearly specified in the text. *Conceptualization*: Chiara Herzog, Charlotte D. Vavourakis, James E. Barrett and Martin Widschwendter. *Methodology*: Chiara Herzog, James E. Barrett, Charlotte D. Vavourakis, Gerlinde Karbon and Andreas Villunger. *Software, Formal Analysis and Validation*: Charlotte D. Vavourakis, Chiara Herzog and James E. Barrett. *Data curation, Validation and Vizualization*: Charlotte D. Vavourakis and Chiara Herzog. *Investigation and Resources*: Jiangrong Wang, Karin Sundström, Joakim Dillner, Gerlinde Karbon and Andreas Villunger. *Writing—Original Draft*: Charlotte D. Vavourakis, Chiara Herzog and Martin Widschwendter. *Writing—Review & Editing*: all authors. *Supervision*: Chiara Herzog and Martin Widschwendter. *Project Administration and Funding Acquisition*: Martin Widschwendter.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw methylation files from cervical screening samples analyzed using the Illumina Human MethylationEPIC are deposited at the European Phenome-Genome Archive (EGA) under accession codes EGAS00001005078. Raw methylation files from cancer cell lines treated with BH3 mimetics are deposited at the NCBI Gene Repository Omnibus (GEO) repository under accession code GSE207224. All original code, including code to reproduce all graphs and tables for our study, is provided in the eutops/WID-HPV github repository (https://github.com/eutops/WID-HPV). Further details and other data that support the findings of our study are available from the corresponding author upon request.

ETHICS STATEMENT

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. Ethical approval for use of samples and linked disease status information in the current study was granted by the Karolinska Ethical Committee (Dnr 2014/1242-31/4).

ORCID

Chiara Herzog https://orcid.org/0000-0002-1572-498X
Charlotte D. Vavourakis https://orcid.org/0000-0002-5027-2309
Gerlinde Karbon https://orcid.org/0000-0002-6149-3342
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

TWITTER

Chiara Herzog @chiara_herzog

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209-249.
2. Sundström K, Eloranta S, Sparén P, et al. Prospective study of human papillomavirus (HPV) types, HPV persistence, and risk of squamous cell carcinoma of the cervix. *Cancer Epidemiol Prev Biomark*. 2010;19: 2469-2478.
3. Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. *Lancet*. 2013;382:889-899.
4. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Pathog*. 2017;13: e1006211.
5. Fiano V, Trevisan M, Fasanelli F, et al. Methylation in host and viral genes as marker of aggressiveness in cervical lesions: analysis in 543 unscreened women. *Gynecol Oncol*. 2018;151:319-326.

6. Apostolidou S, Hadwin R, Burnell M, et al. DNA methylation analysis in liquid-based cytology for cervical cancer screening. *Int J Cancer*. 2009;125:2995-3002.

7. Teschendorff AE, Jones A, Fiegl H, et al. Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med*. 2012;4:24.

8. Doufekas K, Zheng SC, Ghazali S, et al. GALR1 methylation in vaginal swabs is highly accurate in identifying women with endometrial cancer. *Int J Gynecol Cancer*. 2013;23(6):1050-1055. doi:10.1097/IGC.0b013e3182959103

9. Snoek BC, van Splunter AP, Bleeker MCG, et al. Cervical cancer detection by DNA methylation analysis in urine. *Sci Rep*. 2019;9:3088.

10. Siegel EM, Riggs BM, Delmas AL, Koch A, Hakam A, Brown KD. Quantitative DNA methylation analysis of candidate genes in cervical cancer. *PLoS One*. 2015;10:e0122495.

11. Widschwendter A, Gattringer C, Ivarsson L, et al. Analysis of aberrant DNA methylation and human papillomavirus DNA in cervicovaginal specimens to detect invasive cervical cancer and its precursors. *Clin Cancer Res*. 2004;10:3396-3400.

12. Barrett J, Sundström K, Jones A, 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;14:116.

13. Da Silva MLR, De Albuquerque BHDR, Allyrio TADMF, et al. The role of HPV-induced epigenetic changes in cervical carcinogenesis (review). *Biomed Rep*. 2021;15:60.

14. Zapatka M, Borozan I, Brewer DS, et al. The landscape of viral associations in human cancers. *Nat Genet*. 2020;52:320-330.

15. Zhu B, Xiao Y, Yeager M, et al. Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance. *Nat Commun*. 2020;11:886.

16. Yang Z, Wong A, Kuh D, et al. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol*. 2016;17:205.

17. Perskvist N, Norman I, Eklund C, Litton JE, Dillner J. The Swedish cervical cytology biobank: sample handling and storage process. *Biopreserv Biobank*. 2013;11:19-24.

18. Ludvigsson JF, Almqvist C, Bonamy A-KE, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol*. 2016;31:125-136.

19. Hortlund M, Mühr LSA, Lagheden C, Hjerpe A, Dillner J. Audit of laboratory sensitivity of human papillomavirus and cytology testing in a cervical screening program. *Int J Cancer*. 2021;149:2083-2090.

20. Barrett JE, Jones A, Evans E, et al. The DNA methylome of cervical cells can predict the presence of ovarian cancer. *Nat Commun*. 2022;13:448.

21. 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;13:449.

22. Aryee MJ, Jaffe AE, Corrado-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363-1369.

23. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189-196.

24. Morris TJ, Butcher LM, Feber A, et al. ChAMP: 450k Chip analysis methylation pipeline. *Bioinformatics*. 2014;30:428-430.

25. Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017;33:3982-3984.

26. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017;45:e22.

27. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001;17:520-525.

28. 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;10:925-940.

29. Hansen KD. *IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC Methylation Arrays*. R Package Version 0.6.0; 2017.

30. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2016;32:286-288.

31. Walter W, Sánchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015;31:2912-2914.

32. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847-2849.

33. Manjang K, Tripathi S, Yli-Harja O, Dehmer M, Emmert-Streib F. Graph-based exploitation of gene ontology using GOxploreR for scrutinizing biological significance. *Sci Rep UK*. 2020;10:16672.

34. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst*. 2015;1:417-425.

35. Dolgalev I. *msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format*. R Package Version 7.5.1; 2022.

36. Wong-Brown MW, van der Westhuizen A, Bowden NA. Sequential azacitidine and carboplatin induces immune activation in platinum-resistant high-grade serous ovarian cancer cell lines and primes for checkpoint inhibitor immunotherapy. *BMC Cancer*. 2022;22:100.

37. Ramachandran D, Dörk T. Genomic risk factors for cervical cancer. *Cancer*. 2021;13:5137.

38. Shuai K, Liu B. Regulation of gene-activation pathways by PIAS proteins in the immune system. *Nat Rev Immunol*. 2005;5:593-605.

39. Sarosiek KA, Fraser C, Muthalagu N, et al. Developmental regulation of mitochondrial apoptosis by c-Myc governs age- and tissue-specific sensitivity to cancer therapeutics. *Cancer Cell*. 2017;31:142-156.

40. Klopack ET, Carroll JE, Cole SW, Seeman TE, Crimmins EM. Lifetime exposure to smoking, epigenetic aging, and morbidity and mortality in older adults. *Clin Epigenetics*. 2022;14:72.

41. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13:484-492.

42. Murtha M, Esteller M. Extraordinary cancer epigenomics: thinking outside the classical coding and promoter box. *Trends Cancer*. 2016;2:572-584.

SUPPORTING INFORMATION

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