Molecular progression to cervical precancer, epigenetic switch or sequential model?

Running title: Molecular progression of cervical precancerous lesions

Belinda Nedjai[1], Caroline Reuter[1*], Amar Ahmad[1*], Rawinder Banwait[1], Rhian Warman[1], James Carton[2], Sabrina Boer[3], Jack Cuzick[1], Attila Lorincz[1, 4]

[1] Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Bart's and the London School of Medicine, Charterhouse Square, London EC1M 6BQ, UK.

[2] Department of Histopathology, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK.

[3] Department of Urology, Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands.

[4] Corresponding author: a.lorincz@qmul.ac.uk.

* Authors contributed equally

Conflict of interest statement: none

Word count from Introduction to Discussion: 4274

Brief description of the novelty and impact of the article:

For the first time, we explored DNA methylation signatures in adjacent cervical intraepithelial neoplasia from the same cervices. Our study indicates that distinct histopathological grades of CIN have characteristic epigenetic patterns. A clearer understanding of molecular changes leading to CIN3 could provide more accurate biomarkers to improve diagnosis and therapy, facilitating a reduction in healthcare costs and unnecessary treatments in HPV-positive women.

Abstract

The evolution of precancerous cervical lesions is poorly understood. A widely held model of cervical intraepithelial neoplasia grade 3 (CIN3) development is sequential progression from normal through CIN1 and CIN2 to CIN3. Another hypothesis, the "molecular switch" model, postulates that CIN3 can evolve directly from human papillomavirus (HPV)-infected normal epithelium without progressing through CIN1 and CIN2. To shed light on this process, we compared DNA methylation of selected human biomarkers and HPV types in two groups of CIN1: CIN1 that were near or adjacent to CIN3 (adjacent-CIN1) and CIN1 that were the principal lesions with no CIN3 detected (principal-CIN1). 354 CIN (CIN1 and CIN3) and normal tissue areas were dissected and typed for HPV from 127 women who underwent loop electrosurgical excision procedures (LEEP). Methylation of genes *EPB41L3* and the viral regions of HPV16-L1/L2, HPV18-L2, HPV31-L1 and HPV33-L2 were determined by a highly accurate quantitative pyrosequencing of bisulfite converted DNA. There was a significant trend of increased methylation with disease grade comparing normal to CIN1 and CIN3 ($p<0.0001$). Adjacent-CIN1 predominantly shared the same HPV types as the CIN3, however, methylation differed substantially between adjacent-CIN1 and CIN3 ($p=0.008$). In contrast diagnostically principal-CIN1 had an indistinguishable methylation distribution compared to adjacent-CIN1 (*EPB41L3:* $p=0.49$; HPVme-All: $p=0.11$). Our results suggest that progression from normal epithelium to CIN1 or CIN3 is usually promoted by the same HPV type but occurs via distinct DNA epigenotypes, thus favouring the "molecular switch" model.

Word count 235

KEYWORDS:

Methylation; epigenotype; human papillomavirus; biomarker; cervical cancer; cervical intraepithelial neoplasia (CIN); HPV typing; molecular switch; sequential progression.

Abbreviations: Adj.: adjacent; AVER: average; CIN: cervical intraepithelial neoplasia; CpG: 5'-Cytosine-phosphate-Guanine-3'; CTT χ^2: Cuzick Test for Trend statistic; DNA: deoxyribonucleic acid; hr: high-risk; E2BS: E2 binding site; EDTA: Ethylenediaminetetraacetic acid; FFPE: formalin-fixed paraffin-embedded; HCl: hydrochloric acid; HPV: human papillomavirus; HPV16: human papillomavirus type 16; HPV18: human papillomavirus type 18; HPV31: human papillomavirus type 31; HPV33: human papillomavirus type 33; HPVme: HPV methylation; IMP: data imputation; KWT χ^2: Kruskal-Wallis Test statistic, LEEP: loop electrosurgical excision procedure; µL; microliter; n: number of samples; ng: nanogram; Norm: normal; qPCR: quantitative polymerase chain reaction; Prin: principal; SDS: sodium dodecyl sulphate; TSS: transcription start site; URR: upstream regulatory region; vs: versus.

Introduction

Human papillomavirus (HPV) infections account for an estimated 530,000 new cervical cancers and 270,000 deaths annually, most of which occur in developing countries [1-3]. More than 25 types of HPV are transmitted through sexual contact but most infections do not cause symptoms and are cleared after a short time. When an infection is persistent with one or more high-risk HPV (hrHPV) there is an increased chance of cervical intraepithelial neoplasia (CIN2 or CIN3) [4-8]. It was originally thought that cervical cancer evolved from HPV infected normal cervical epithelium via a "sequential progression" model (Figure 1). In this model, a long-lasting HPV infection would cause cervical intraepithelial changes in a consecutive way, from HPV infected normal tissue to CIN1, CIN2, CIN3 and finally cancer [9]. However, an alternative idea is that CIN1 may not be necessary for the development of CIN3 [10, 11] and that CIN3 could evolve directly from normal epithelium infected by HPV following a "molecular switch" model (Figure 1). This means clinically relevant CIN3 may develop fairly rapidly following HPV infection [12]. As a consequence, CIN1 lesions in general and most CIN2 may not be precursor stages of cervical cancer, but rather the visible effects of a productive HPV infection. It may

then take 10–12 years to develop invasive cervical cancer from CIN3 [13, 14]. A clearer understanding of molecular changes leading to CIN3 could provide more accurate biomarkers to improve diagnosis and therapy.

Looking at the methylation level of certain genes is a way to shed light on the molecular progression of precancerous cervical lesions. DNA methylation is a reproducible physical epigenetic change involved in a variety of cellular processes that plays an important role in cancer progression [15]. Numerous studies have highlighted the importance of host and viral gene methylation in the development of cervical cancer and potential use as biomarkers to triage HPV positive patients [16-24]. A strong association has been observed between DNA methylation patterns and CIN of various grades in both cervical scrapes and biopsies. Typically, higher methylation is observed in patients with advanced intraepithelial lesions and carcinoma.

In this study we investigated whether methylation data could bring evidence in support of either the molecular switch or the sequential progression model or both. We used epigenetic markers and HPV typing to investigate the two main models of precancerous cervical disease progression in areas of cervical lesions with discrete coexisting foci of different grades. We sought to determine whether levels of DNA methylation were a characteristic of lesion grade. Our primary aims were to contrast methylation in CIN1 in two possible configurations: a) adjacent-CIN1 vs CIN3 from the same women and b) adjacent-CIN1 vs principal-CIN1 lesion from different women. We hypothesised that if methylation levels were different between adjacent-CIN1 and CIN3 on the same cervix but similar between adjeacent-CIN1 and principal-CIN1, this would add support to the "molecular switch" model. On the other hand, if the methylation levels overlapped substantially between grades and adjacent-CIN1 showed higher methylation levels than principal-CIN1 then; "the sequential progression" model would be favoured. A secondary aim was to determine whether adjacent lesions were usually infected by the same or different HPV types. The study was made possible by our collection

of carefully annotated Loop Electrosurgical Excision Procedure (LEEP) cervical tissues (Figure 2).

Material and methods

Our goal was to compare HPV types and DNA methylation patterns in different kinds of cervical lesions that we obtained from the LEEP surgical tissues. Tissue areas to dissect were identified by microscope-based pathology review and annotation.

Study population and clinical specimens

Archived material from the Predictors studies (Ethics no. 05/Q0406/57) collected from 2005 to 2009 was used [25, 26]. All patients with moderate and higher dyskaryosis in the baseline liquid-based cytology (ThinPrep; Hologic, Bedford, USA) and those with persistent mild abnormalities were sent for colposcopy. Before colposcopic examination an exfoliated cell specimen was taken from the cervix into PreservCyt (Hologic Inc.) for HPV typing. Additionally, punch biopsies were taken from areas with abnormal appearance to determine the histological diagnosis of the patients. Final histopathological diagnoses were based on reviews by at least two pathologists. The highest grade of abnormality seen in the biopsy was used. If the pathologist panel review confirmed CIN2+, patients underwent treatment by LEEP and the surgical tissues were fixed in formalin and embedded in paraffin (FFPE). We used FFPE punch biopsies from 104 women and FFPE LEEP tissues from 127 women (Figure 2). We excluded women who had a final diagnosis of CIN2 from the study to minimize confusion from inaccurate pathology diagnoses; this was because of strong evidence that CIN2 in particular is a less reliable diagnosis than CIN1 or CIN3 and may not be a distinct biological entity [27-31].

Lesion dissection, annotation and DNA extraction

For the biopsies, twelve 5µm FFPE sections of normal, CIN1 and CIN3 were cut on a microtome using a new blade for each block. Sections were stored at -70°C until DNA

extraction. Entire sections were scraped from the slides using a scalpel and deparaffinised using three washes in xylene and one wash with 100% ethanol. DNA was extracted using QIAamp DNA FFPE Tissue Kit (QIAGEN) according to manufacturer's instructions with an initial incubation step at 56°C for 16-18 hours with proteinase K. The DNA concentration was measured using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

For the LEEP blocks, sections were cut using a sandwich cutting procedure as described in by [32] using a new blade for each block (Supplementary Figure 1) [32]. Negative controls (breast tissue) were used after every 10 blocks to rule out contamination. The negative controls all tested negative for HPV. Twelve 5um FFPE sections of the 127 LEEP specimens were cut. The first and last sections were H&E stained and the remaining ten sections were used for DNA extraction. The H&E slides were carefully marked by an expert histopathologist (J Carton) to indicate different areas of CIN and normal tissues and used as guides for subsequent precise scalpel-based dissection of the corresponding unstained sections under low power magnification. Sections were separated into two categories (CIN1 and CIN3 cases) depending on the highest-grade lesion found. Where possible, normal squamous epithelium adjacent to CIN1 were also dissected. The CIN3 cases also included additional tissue (where possible) extracted from adjacent normal and CIN1 areas. Only one (if any) adjacent normal area was dissected per CIN1 or CIN3 case. However, if sections contained multifocal CIN1 or CIN3 lesions, additional areas were dissected and processed separately. Scraped areas were separated into different tubes for DNA extraction. Because of the small amount of starting material, dissected areas were deparaffinised using 160µL of hexadecane followed by a 5-minute incubation at 56°C. Two hundred microliters of universal extraction buffer containing 50mM Tris-HCl pH 8.0, 1mM EDTA and 0.05% SDS was added to tissues along with 400mg of Proteinase K (QIAGEN) and incubated overnight at 56°C followed by a one-hour incubation at 90°C. The lower phase was then transferred to a new tube and stored at -20°C before PCR.

HPV type detection

The 354 dissected areas from LEEP samples were tested using the PapType High Risk HPV Detection and Genotyping kit (PapType kit, Genera Biosystems Ltd) according to the manufacturer's instructions (Figure 2). The kit is able to detect 13 hrHPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68), one possibly hrHPV type (HPV66) and two low-risk types (HPV6 and 11). The PapType test was performed with 10µL of DNA in a final reaction volume of 20µL with the addition of 0.4µL of Tween 20 (2%). The PCR reaction amplifies a variable region of the L1 gene. A fragment of the human cardiac myosin light chain gene (MLC-1) was co-amplified in the same reaction vessel as a quality and quantity control. The results of HPV typing from the LEEP samples were compared to each other and their corresponding cervical scrape samples taken before colposcopy (Figure 2). The exfoliated cells were genotyped by Linear Array (Roche Diagnostics, Rotkreuz, Switzerland) for Predictors 1 and by BD HPV test (BD Diagnostics, Franklin Lakes, USA) for Predictors 2 [25, 33].

DNA methylation assays

DNA methylation was measured by a highly accurate pyrosequencing assay as previously described [34]. Sodium bisulfite conversion of the genomic DNA was done using 200ng DNA obtained from the punch biopsies and 20µL of the DNA extract from the LEEP sections using the EZ DNA Methylation Kit (Zymo Research, CA, USA) following the manufacturer's instructions.

The punch biopsies were used to determine the most informative genes for the main model development study using the LEEP sections (Figure 2). Analysed human regions on the biopsies were based on our and others' previous work and included *EPB41L3* (CpG sites 425, 427 and 438 relative to transcription start site (TSS)) and *MAL* (CpG sites 529, 533, 535, 539, and 542 relative to TSS) [24, 35]. Viral regions included the HPV16-URR (CpG sites 31, 37, 43, 52, 58, 7428, 7434, 7455 and 7461, which comprises E2BS1, 3 and 4), HPV16-L1 (CpG

sites 6367 and 6389), HPV16-L2 (CpG sites 4238, 4247, 4259 4268, 4275) and HPV18-L2 (CpG sites 4256, 4261, 42656, 4269, 4275, 4282)[21, 22]. None of the samples were excluded on the basis of their HPV type. Amplification of CpG positions were done using the PyroMark PCR kit (QIAGEN, Germany) with 10ng of converted DNA (except for HPV18-L2 PCR, for which 20ng of DNA was used) in a 25µL volume with final concentration of reagents of 1x for Coral Load and PyroMark mix, 0.2µM of PCR primers. PCR cycling conditions were 15 minutes at 94°C, followed by 45 cycles of 94°C, 54°C (51°C for HPV16-L2, 55°C for HPV16-URR), 72°C each for 30 seconds and a final extension at 72°C for 10 minutes. PCR products were pyrosequenced using a PyroMark™Q96 ID (Qiagen) instrument as previously described [36]. Pyrosequencing runs included positive controls of known methylation level (0%, 50% and 100%) to allow standardised direct comparisons between different primer sets and a negative control. For each gene and viral region, the methylation percentage was averaged over all the CpG positions investigated since we have previously shown that these CpGs are always similarly methylated within a particular sample [21, 22].

For the main study on the LEEP sections, based on the pilot results obtained from the punch biopsies, we decided to concentrate on human gene *EPB41L3*, the L1 and L2 regions of HPV16 and the L2 region of HPV18 (Figure 2). To the assays above, we also added newly developed assays for HPV31 and HPV33 [37, 38]. These were two CpG positions in the L1 region of HPV31 (6352 and 6354) and four positions in the L2 region of HPV33 (5557, 5560, 5566 and 5572). Assays were run as previously described [22]. We used 8µLof bisulfite converted DNA in the PCR using the PyroMark PCR kit as described above. Percentage methylation was averaged over all the CpG positions investigated and methylation levels were further averaged over all HPV types to create a single variable for HPV methylation called HPVme-All (Figure 2) because there were insufficient data for HPV18, HPV31 and HPV33 separately. For women with multiple samples of the same type (for example, when two CIN1 or three CIN3 areas were dissected), the percentage methylation was averaged to produce a single value per lesion type per woman (Figure 2).

Assays were performed blinded by the technicians with cases and controls randomly intermixed, thereby minimizing concerns of biasing batch effects. We used a pre-specified study design and the statistical analyses were done after molecular testing was finished and were independent of the team that produced the laboratory results.

Statistical analysis

Analyses were performed according to a pre-specified statistical analysis plan, which was blinded to the methylation data. For the pilot study using punch biopsies, Kruskal-Wallis rank sum tests were used to compare averaged percentage methylation data between normal, CIN1 and CIN3 cases.

For the main study using LEEP samples, paired and unpaired Wilcoxon tests were performed as appropriate to compare DNA methylation levels between groups using *EPB41L3* and HPVme-All as predictors. DNA methylation missing values of HPVme-All were imputed with the value of zero for any hrHPV negative sample and by a median regression with age as a predictor and DNA methylation as an outcome for hrHPV positive samples. Missing values for *EPB41L3* were imputed by a median regression with age as a predictor and DNA methylation as an outcome independently of their HPV infection status (Figure 2).

All p-values were two-sided with significance set at $\alpha < 0.05$. No adjustments were made for multiple comparisons. Analyses were undertaken using R statistical software version 3.3.1.

Results

Pilot and main study: specimens and strategy

A pilot study was performed on 104 punch biopsies to select a set of informative biomarkers for the main study. The methylation of candidate biomarkers *EPB41L3, MAL*, HPV16-L1, HPV16-L2, HPV16-E2 binding sites (1, 3, and 4), and HPV18-L2 is presented in supplementary Figure 2 and supplementary Table 1. *MAL* and the HPV16-E2 binding sites were dropped from further study due to inadequate effect size. Although the L2 region of HPV18 was not significant in the pilot study due to small sample size (Supplementary Figure

2), methylation of this region was assayed in the main study because of the importance of HPV18 as a carcinogenic HPV. To these assays we added biomarkers in the L1 region of HPV31 and the L2 region of HPV33, which were validated as informative in different studies [22, 37].

For the main study we used dissected LEEP tissues from 127 women (Figure 2). The women were subdivided into CIN1 and CIN3 cases depending on the highest grade intraepithelial lesion diagnosed in the available tissues. There were 49 women classed as principal-CIN1 (where no CIN3 was present on the cervix) and 78 as CIN3 cases. A majority (n=53) of the CIN3 cases were classed as multifocal because they presented adjacent lesions of different grades (adjacent-CIN1 or adjacent-CIN3). The process gave us a total of 354 dissected CIN or normal areas (Figure 2).

Nature of methylation patterns in multifocal CIN3

A paired analysis on tissue isolated from the same cervices indicated that methylation of the normal tissue and CIN1 adjacent to CIN3 were significantly different from the CIN3 lesions. This was true for both *EPB41L3* (normal vs CIN3: p<0.0001 and CIN1 vs CIN3: p=0.008) and HPVme-All (normal vs. CIN3: p<0.0001 and CIN1 vs. CIN3: p=0.0011; Figure 3). The normal tissue samples were also different from the CIN1 lesions (*EPB41L3*: p=0.005 and HPVme-All: p=0.0004). Another way to view the data is to test the differences in methylation levels between the paired histopathological groups on a per-cervix basis. In figure 4, we plotted percentage differences in methylation between all paired measurements for *EPB41L3* and HPVme-All markers. A large majority of methylation differences were positive, suggesting an increase of methylation as diagnoses changed from CIN1 to CIN3 in each woman.

Normal tissue and CIN1 lesions have characteristic methylation levels regardless of whether they are from unifocal or multifocal lesional cervices

Methylation levels of normal tissues located near CIN1 or CIN3 were similar for both *EPB41L3* (p=0.21) and HPVme-All (p=0.16) (Figure 5). Of greater interest, the pattern was the same for

CIN1. Methylation levels of principal-CIN1 were not different from adjacent-CIN1 in CIN3 for *EPB41L3* (p=0.49) nor HPVme-All (p=0.11) (Figure 5). In fact the median of *EPB41L3* for adjacent-CIN1 appeared slightly (but non-significantly) lower than for primary-CIN1 and supports a model where *EPB41L3* DNA methylation levels may exist as discrete haplotypes, characterizing and possibly controlling the morphological appearance of pre-cancerous lesions.

Association of high-risk HPV positivity and severity of lesions

354 dissected areas from LEEP samples were HPV typed (Figure 6). A majority of principal-CIN1 (87%), adjacent-CIN1 (97%) and CIN3 (99%) were high-risk HPV positive. Most of the CIN1 and CIN3 were infected by a single hrHPV type (69% of principal-CIN1 and 81% adjacent-CIN1). HPV infection was also found in normal tissues, but at a much lower rate (Figure 6).

Comparison of HPV types found in adjacent lesions and between LEEP samples and exfoliated cells

In multifocal CIN3, a majority of adjacent-CIN1 were infected by the same hrHPV as the CIN3 (87%, supplementary Figure 3). HPV type results were classed as compatible when some HPV types matched but one or the other paired sample also presented additional types. Normal tissues showed a somewhat lower similarity to their adjacent-CIN1 (41% with same or compatible types to adjacent-CIN1 and 45% compatible to principal-CIN1) and adjacent CIN3 (51%). Furthermore, we noticed that when multiple lesions of the same grade were present in a multifocal CIN3, they were generally infected by the same HPV type (13/16=81% in CIN3 and 11/12=91% in CIN1), which is compatible with a single infection event from which multiple lesions arose. However, in principal-CIN1 only 56% (9/16) of adjacent CIN1 were infected by the same HPV type, consistent with greater diversity of low-grade lesions, perhaps due to multiple infection events combined with both transient and short-term persistent HPV

infections.

HPV types in macro-dissected LEEP samples were also compared to their corresponding exfoliated cells from cervical scrapes at colposcopy and a good agreement was found between the two (Supplementary table 2). The HPV types found in 99% (77 out of 78) of the CIN3 and 88% (43 out of 49) of CIN1 lesions matched or were compatible (i.e. additional types were found either in cervical scrapes or in the dissected areas from LEEP) with the corresponding exfoliated cell genotyping results.

Discussion

The main goal of our study was to characterize DNA methylation events associated with progression to CIN3. We show in our detailed methylation and HPV typing analyses that the topography of cervical HPV infection, the resultant intraepithelial neoplasia and the molecular patterns are heterogeneous and quite complex. We report here for the first time that, although adjacent CIN of different grades usually contain the same hrHPV type(s), methylation patterns are significantly different and characteristic of the lesion grade.

Previously published evidence of genetic alterations that accumulate during cervical tumorigenesis indicate a common cellular origin for multifocal lesions, suggesting that different intraepithelial lesions arise from the same or clonally related progenitor cells [11]. The earlier study taken together with our results suggests that intraepithelial lesions of different grades on a multifocal cervix can arise simultaneously and therefore the "molecular switch" model (different methylation haplotype switches triggered at the same time) might be more likely than the "sequential progression" model (Figure 1). If CIN3 predominantly arose directly from adjacent-CIN1, we would expect to see methylation in the adjacent-CIN1 having a bridging pattern. Methylation of the adjacent-CIN1 should be very close to or at the same average level as the nearby CIN3 and such adjacent-CIN1 should be distinctly different than primary-CIN1; however, this was not what we observed. Although we cannot prove the actual overall secular

mechanism of CIN3 origin from cross-sectional data, our results do suggest a model where CIN3 can emerge directly from normal epithelium, possibly simultaneously to CIN1. After that, lesions that persist can enlarge over time but retain their characteristic epigenotypes. This model is also consistent with what we see with respect to HPV type concordance in different lesions on the same cervix. An initial rapid progression to CIN3 may perhaps be driven by a series of epigenetic changes happening both before and during HPV infection. We propose that distinct epigenotypes underpin distinct morphologic patterns such as CIN1 or CIN3, explaining partially the different progression characteristics of these lesions. The drivers of epigenetic changes could be multiple. One possibility is that epigenetic changes, such as *EPB41L3* methylation, may occur before hrHPV infection and facilitate the virus genome amplification and genetic instability phase, allowing in some women relatively rapid progression to CIN3. Furthermore, although most normal tissues have 0% methylation for *EPB41L3*, a sizeable proportion have modest elevation of *EPB41L3* methylation of up to 10% [19, 24].

As regards to HPV types, we found that most CIN1 and CIN3 lesions (69% in unifocal CIN1 and 81% in multifocal CIN3 cases) were infected by a single hrHPV type. Our data partly confirm a similar study by Quint [32] where an association of a single HPV type with a discrete area of CIN was found for 93% (372/399) of laser capture micro-dissected tissue fragments they analysed by PCR. Van der Marel et al have suggested that multiple high-grade lesions on the cervix are often caused by a single carcinogenic genotype while other carcinogenic HPV types detected in cervical smears of the same patients are related to independent transient infections [39]. They support the "one virus-one lesion" hypothesis and discount biological interactions of multiple HPV infections on the lesion level [32].

We also showed that when multiple lesions of the same grade were present in a multifocal CIN3, they were very often infected by the same HPV type (81% in CIN3 and 91% in CIN1). However, in principal-CIN1 only 56% of the multiple CIN1 dissected were infected by the same

HPV type. This indicates that when lower-grade lesions are the primary diagnostic manifestation of HPV infection the distinct lesions are often infected by different and presumed transient HPV types.

There were consecutively large differences and an overall significant Cuzick test trend of increased methylation with increasingly severe intraepithelial histopathology, going from normal to CIN1 to CIN3. These effects were seen for both *EPB41L3* and HPVme-All, and were very similar to results we obtained earlier on exfoliated cervical specimens. The correlation between biopsies and scrapes allows a generalization of our findings to specimens obtained by cervical scraping and vaginal self-sampling [24, 38].

Our study has some limitations. While the epigenetic switch model can underpin added insights for understanding epidemiologic and molecular distinctions between disease transitions in cervical pathogenesis a limitation is our inability to formally prove the methylation haplotype switch. However, the same caveat applies to the sequential model, conception of which developed over decades, predominantly on the basis of logical but unproven assumptions. A definitive study would require lengthy (possibly decades long) follow-up of large groups of women with HPV infection. The cervical lesions would require detailed topological mapping with frequent sampling of changing tissues areas by biopsies (which may affect the course of disease) and carefully directed scrapes, coupled to extensive molecular investigations. Such an *in vivo* mapping would no doubt quickly run into ethical concerns. We need to bear in mind that the two models are compatible and possibly both are correct in certain situations. Although our data support the idea that most CIN3 arise via a haplotype epigenetic switch some CIN3 may develop via a sequential progression.

A strength of our study is the use of an expertly rendered consensus CIN3 histopathological diagnosis as the primary endpoint rather than CIN2/3 or routine practice CIN3. Previous studies have shown that CIN2 is an equivocal diagnosis of precancer and more importantly in

this context CIN2 has very poor inter-rater agreement [27, 29, 40, 41] and is much more likely to regress than CIN3 [42, 43]. Therefore, the grading of CIN2/3 may vary depending on the person rendering the diagnosis or on the follow-up protocol, making comparisons of risks between studies including CIN2 as an endpoint challenging to interpret. In addition, we used a highly accurate pyrosequencing method of DNA methylation measurement, shown to have high precision and low bias, which provided quantitative results expressed as a percentage from 0 to 100. We performed a molecular examination of multifocal intraepithelial lesions within the same patients, with direct quantitative contrasts to adjacent lesions and lesions from other women.

While our data suggest at least two discrete epigenetic states (CIN1 and CIN3), we cannot say whether there are more epigenotypes. It is unlikely that our results were biased by the contamination of dissected lesions with other CIN or surrounding normal cells. All lesions were expertly marked by our team histopathologist (JC) and then dissected, carefully avoiding areas of other CIN and normal tissue. While the isolated CIN DNA probably had a low level of contamination from normal DNA the latter tissues were predominantly either not methylated or were methylated at lower levels than the CIN. Furthermore normal contamination on average would have been similar for principal-CIN1 and adjacent-CIN1, which were our most important comparison groups.

There are many remaining questions and it is clear that relatively little is known about specific genetic and epigenetic mutations that underpin the evolution of CIN3, nor exactly which steps have a required sequence that then further facilitate the transformation of CIN3 to malignancy. A lot is known about molecular and mechanistic aspects of oncogenes, tumour suppressor genes, regulatory RNA pathways, multi-protein pathways etc. and also that many such pathways may become abnormal during carcinogenesis. However, for any given woman it is still not possible to predict accurately whether she will develop cervical cancer, what specific

constellation of pathway abnormalities would lead to the cancer and over what timeframe this will occur.

In conclusion, we show for the first time that, although adjacent CIN of different grades usually contain the same hrHPV type(s), methylation patterns are significantly different and characteristic of the lesion grade, supporting a "molecular switch" model of progression to CIN3 that may be characterizable by distinct methylation haplotype patterns. Methylation testing may be an effective triage tool to detect and characterize women at high risk of developing CIN3 and cancer. Our results may influence the screening process in many ways, for example by providing an objective method to reach more accurate prognoses or by helping to avoid overtreatment of women with non-progressive lesions. Reducing the numbers of younger women treated by surgical excisions would have an important effect in preserving the ability of women to have uncomplicated pregnancies and to overall improve childbearing at a later date.

Statement of author contributions

Study conception and design: CR, AL, BN

Acquisition of data: CR, SB, JC, RB, RW

Analysis and interpretation of data: BN, CR, AL, AA, JCK

Drafting of manuscript: BN, CR, AL

Critical revision: BN, CR, AL, AA, JCK

Authors list: Belinda Nedjai (BN), Amar Ahmad (AA), Caroline Reuter (CR), Rawinder Banwait (RB), Rhian Warman (RW), James Carton (JC), Sabrina Boer (SB), Jack Cuzick (JCK) , Attila Lorincz (AL)

References

1. Ferlay, J., Forman, D., Mathers, C. D. and Bray, F. Breast and cervical cancer in 187 countries between 1980 and 2010. Lancet. 2012;379(9824):1390-1.
2. Fidler, M. M., Soerjomataram, I. and Bray, F. A global view on cancer incidence and national levels of the human development index. International journal of cancer. 2016;139(11):2436-46.
3. de Martel, C., Plummer, M., Vignat, J. and Franceschi, S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. International journal of cancer. 2017;141(4):664-70.
4. Bosch, F. X., Lorincz, A., Munoz, N., Meijer, C. J. and Shah, K. V. The causal relation between human papillomavirus and cervical cancer. Journal of clinical pathology. 2002;55(4):244-65.
5. Kjaer, S. K., Frederiksen, K., Munk, C. and Iftner, T. Long-term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus infection: role of persistence. Journal of the National Cancer Institute. 2010;102(19):1478-88.
6. Thomsen, L. T., Frederiksen, K., Munk, C., Junge, J., Iftner, T. and Kjaer, S. K. Long-term risk of cervical intraepithelial neoplasia grade 3 or worse according to high-risk human papillomavirus genotype and semi-quantitative viral load among 33,288 women with normal cervical cytology. International journal of cancer. 2015;137(1):193-203.
7. Hoffman, S. R., Le, T., Lockhart, A., Sanusi, A., Dal Santo, L., Davis, M., McKinney, D. A., Brown, M., Poole, C., Willame, C. and Smith, J. S. Patterns of persistent HPV infection after treatment for cervical intraepithelial neoplasia (CIN): A systematic review. International journal of cancer. 2017;141(1):8-23.
8. Xi, L. F., Schiffman, M., Koutsky, L. A., Hughes, J. P., Hulbert, A., Shen, Z., Galloway, D. A. and Kiviat, N. B. Variant-specific persistence of infections with human papillomavirus Types 31, 33, 45, 56 and 58 and risk of cervical intraepithelial neoplasia. International journal of cancer. 2016;139(5):1098-105.
9. Snijders, P. J., Steenbergen, R. D., Heideman, D. A. and Meijer, C. J. HPV-mediated cervical carcinogenesis: concepts and clinical implications. The Journal of pathology. 2006;208(2):152-64.
10. Castle, P. E., Gage, J. C., Wheeler, C. M. and Schiffman, M. The Clinical Meaning of a Cervical Intraepithelial Neoplasia Grade 1 Biopsy. Obstet Gynecol. 2011;118(6):1222-9.
11. Larson, A. A., Liao, S. Y., Stanbridge, E. J., Cavenee, W. K. and Hampton, G. M. Genetic alterations accumulate during cervical tumorigenesis and indicate a common origin for multifocal lesions. Cancer research. 1997;57(19):4171-6.
12. Winer, R. L., Kiviat, N. B., Hughes, J. P., Adam, D. E., Lee, S. K., Kuypers, J. M. and Koutsky, L. A. Development and duration of human papillomavirus lesions, after initial infection. The Journal of infectious diseases. 2005;191(5):731-8.
13. Wallin, K. L., Wiklund, F., Angstrom, T., Bergman, F., Stendahl, U., Wadell, G., Hallmans, G. and Dillner, J. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. The New England journal of medicine. 1999;341(22):1633-8.
14. Zielinski, G. D., Snijders, P. J., Rozendaal, L., Voorhorst, F. J., van der Linden, H. C., Runsink, A. P., de Schipper, F. A. and Meijer, C. J. HPV presence precedes abnormal cytology in women developing cervical cancer and signals false negative smears. British journal of cancer. 2001;85(3):398-404.

15. Robertson, J. T., Papadopoulos, S. M. and Traynelis, V. C. Assessment of adjacent-segment disease in patients treated with cervical fusion or arthroplasty: a prospective 2-year study. J Neurosurg Spine. 2005;3(6):417-23.

16. Badal, S., Badal, V., Calleja-Macias, I. E., Kalantari, M., Chuang, L. S., Li, B. F. and Bernard, H. U. The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. Virology. 2004;324(2):483-92.

17. Turan, T., Kalantari, M., Calleja-Macias, I. E., Cubie, H. A., Cuschieri, K., Villa, L. L., Skomedal, H., Barrera-Saldana, H. A. and Bernard, H. U. Methylation of the human papillomavirus-18 L1 gene: a biomarker of neoplastic progression? Virology. 2006;349(1):175-83.

18. Wentzensen, N., Sherman, M. E., Schiffman, M. and Wang, S. S. Utility of methylation markers in cervical cancer early detection: appraisal of the state-of-the-science. Gynecologic oncology. 2009;112(2):293-9.

19. Lorincz, A. T. Virtues and Weaknesses of DNA Methylation as a Test for Cervical Cancer Prevention. Acta Cytol. 2016;60(6):501-12.

20. Mirabello, L., Sun, C., Ghosh, A., Rodriguez, A. C., Schiffman, M., Wentzensen, N., Hildesheim, A., Herrero, R., Wacholder, S., Lorincz, A. and Burk, R. D. Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. Journal of the National Cancer Institute. 2012;104(7):556-65.

21. Lorincz, A. T., Brentnall, A. R., Vasiljevic, N., Scibior-Bentkowska, D., Castanon, A., Fiander, A., Powell, N., Tristram, A., Cuzick, J. and Sasieni, P. HPV16 L1 and L2 DNA methylation predicts high-grade cervical intraepithelial neoplasia in women with mildly abnormal cervical cytology. International journal of cancer. 2013;133(3):637-44.

22. Brentnall, A. R., Vasiljevic, N., Scibior-Bentkowska, D., Cadman, L., Austin, J., Szarewski, A., Cuzick, J. and Lorincz, A. T. A DNA methylation classifier of cervical precancer based on human papillomavirus and human genes. International journal of cancer. 2014;135(6):1425-32.

23. Hesselink, A. T., Heideman, D. A., Steenbergen, R. D., Gok, M., van Kemenade, F. J., Wilting, S. M., Berkhof, J., Meijer, C. J. and Snijders, P. J. Methylation marker analysis of self-sampled cervico-vaginal lavage specimens to triage high-risk HPV-positive women for colposcopy. International journal of cancer. 2014;135(4):880-6.

24. Vasiljevic, N., Scibior-Bentkowska, D., Brentnall, A. R., Cuzick, J. and Lorincz, A. T. Credentialing of DNA methylation assays for human genes as diagnostic biomarkers of cervical intraepithelial neoplasia in high-risk HPV positive women. Gynecologic oncology. 2014;132(3):709-14.

25. Szarewski, A., Ambroisine, L., Cadman, L., Austin, J., Ho, L., Terry, G., Liddle, S., Dina, R., McCarthy, J., Buckley, H., Bergeron, C., Soutter, P., Lyons, D. and Cuzick, J. Comparison of predictors for high-grade cervical intraepithelial neoplasia in women with abnormal smears. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17(11):3033-42.

26. Szarewski, A., Mesher, D., Cadman, L., Austin, J., Ashdown-Barr, L., Ho, L., Terry, G., Liddle, S., Young, M., Stoler, M., McCarthy, J., Wright, C., Bergeron, C., Soutter, W. P., Lyons, D. and Cuzick, J. Comparison of seven tests for high-grade cervical intraepithelial neoplasia in women with abnormal smears: the Predictors 2 study. Journal of clinical microbiology. 2012;50(6):1867-73.

27. Carreon, J. D., Sherman, M. E., Guillen, D., Solomon, D., Herrero, R., Jeronimo, J., Wacholder, S., Rodriguez, A. C., Morales, J., Hutchinson, M., Burk, R. D. and Schiffman, M. CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists. 2007;26(4):441-6.

28. Gage, J. C., Schiffman, M., Hunt, W. C., Joste, N., Ghosh, A., Wentzensen, N., Wheeler, C. M. and Steer, N. M. H. P. R. Cervical Histopathology Variability Among Laboratories A

Population-Based Statewide Investigation. American journal of clinical pathology. 2013;139(3):330-5.

29. Stoler, M. H., Ronnett, B. M., Joste, N. E., Hunt, W. C., Cuzick, J., Wheeler, C. M. and Steeri, N. M. H. P. R. The Interpretive Variability of Cervical Biopsies and Its Relationship to HPV Status. Am J Surg Pathol. 2015;39(6):729-36.

30. Castle, P. E., Stoler, M. H., Solomon, D. and Schiffman, M. The relationship of community biopsy-diagnosed cervical intraepithelial neoplasia grade 2 to the quality control pathology-reviewed diagnoses: an ALTS report. American journal of clinical pathology. 2007;127(5):805-15.

31. Stoler, M. H., Schiffman, M. and Undetermin, A. S. C. Interobserver reproducibility of cervical cytologic and histologic interpretations - Realistic estimates from the ASCUS-LSIL triage study. Jama-J Am Med Assoc. 2001;285(11):1500-5.

32. Quint, W., Jenkins, D., Molijn, A., Struijk, L., van de Sandt, M., Doorbar, J., Mols, J., Van Hoof, C., Hardt, K., Struyf, F. and Colau, B. One virus, one lesion-individual components of CIN lesions contain a specific HPV type. The Journal of pathology. 2012;227(1):62-71.

33. Szarewski, A., Mesher, D., Cadman, L., Austin, J., Ashdown-Barr, L., Ho, L., Terry, G., Liddle, S., Young, M., Stoler, M., McCarthy, J., Wright, C., Bergeron, C., Soutter, W. P., Lyons, D. and Cuzick, J. Comparison of Seven Tests for High-Grade Cervical Intraepithelial Neoplasia in Women with Abnormal Smears: the Predictors 2 Study. Journal of clinical microbiology. 2012;50(6):1867-73.

34. Mirabello, L., Schiffman, M., Ghosh, A., Rodriguez, A. C., Vasiljevic, N., Wentzensen, N., Herrero, R., Hildesheim, A., Wacholder, S., Scibior-Bentkowska, D., Burk, R. D. and Lorincz, A. T. Elevated methylation of HPV16 DNA is associated with the development of high grade cervical intraepithelial neoplasia. International journal of cancer. 2013;132(6):1412-22.

35. Hesselink, A. T., Heideman, D. A. M., Steenbergen, R. D. M., Coupe, V. M. H., Overmeer, R. M., Rijkaart, D., Berkhof, J., Meijer, C. J. L. M. and Snijders, P. J. F. Combined Promoter Methylation Analysis of CADM1 and MAL: An Objective Triage Tool for High-Risk Human Papillomavirus DNA-Positive Women. Clinical Cancer Research. 2011;17(8):2459-65.

36. Vasiljevic, N., Wu, K., Brentnall, A. R., Kim, D. C., Thorat, M. A., Kudahetti, S. C., Mao, X., Xue, L., Yu, Y., Shaw, G. L., Beltran, L., Lu, Y. J., Berney, D. M., Cuzick, J. and Lorincz, A. T. Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing. Dis Markers. 2011;30(4):151-61.

37. Brentnall, A. R., Vasiljevic, N., Scibior-Bentkowska, D., Cadman, L., Austin, J., Cuzick, J. and Lorincz, A. T. HPV33 DNA methylation measurement improves cervical pre-cancer risk estimation of an HPV16, HPV18, HPV31 and *EPB41L3* methylation classifier. Cancer biomarkers : section A of Disease markers. 2015;15(5):669-75.

38. Vasiljevic, N., Scibior-Bentkowska, D., Brentnall, A., Cuzick, J. and Lorincz, A. A comparison of methylation levels in HPV18, HPV31 and HPV33 genomes reveals similar associations with cervical precancers. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2014;59(3):161-6.

39. van der Marel, J., Quint, W. G., Schiffman, M., van de Sandt, M. M., Zuna, R. E., Terence Dunn, S., Smith, K., Mathews, C. A., Gold, M. A., Walker, J. and Wentzensen, N. Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16. International journal of cancer. 2012.

40. Castle, P. E., Schiffman, M., Wheeler, C. M., Wentzensen, N. and Gravitt, P. E. Impact of Improved Classification on the Association of Human Papillomavirus With Cervical Precancer. American journal of epidemiology. 2010;171(2):155-63.

41. Robertson, A. J., Anderson, J. M., Beck, J. S., Burnett, R. A., Howatson, S. R., Lee, F. D., Lessells, A. M., Mclaren, K. M., Moss, S. M., Simpson, J. G., Smith, G. D., Tavadia, H. B. and Walker, F. Observer Variability in Histopathological Reporting of Cervical Biopsy Specimens. Journal of clinical pathology. 1989;42(3):231-8.

42. Castle, P. E., Kreimer, A. R., Wacholder, S., Wheeler, C. M., Koutsky, L. A., Rydzak, G., Buckman, D. W., Graubard, B. and Schiffman, M. Influence of Loop Electrosurgical Excision Procedure on Subsequent Acquisition of New Human Papillomavirus Infections. Journal of Infectious Diseases. 2009;199(11):1612-20.

43. Trimble, C. L., Piantadosi, S., Gravitt, P., Ronnett, B., Pizer, E., Elko, A., Wilgus, B., Yutzy, W., Daniel, R., Shah, K., Peng, S. W., Hung, C. F., Roden, R., Wu, T. C. and Pardoll, D. Spontaneous regression of high-grade cervical dysplasia: Effects of human papillomavirus type and HLA phenotype. Clinical Cancer Research. 2005;11(13):4717-23.

Figure Legend

Figure 1 Current alternative cervical cancer progression models. In the sequential progression model (A), a long-lasting HPV infection causes cervical intraepithelial neoplasia (CIN) changes in consecutive steps from low grade CIN1 to high grade CIN3 and finally cancer. An alternative concept, the molecular switch model B), assumes that CIN1 is not necessary for the development of CIN3. CIN3 could evolve straight from normal epithelium infected by HPV. Here, we propose that distinct epigenotypes trigger distinct morphologic changes such as CIN1 or CIN3 independently, explaining partially the versatile nature of these lesions. Double-sided arrows indicate possible regression.

Figure 2. Flow chart showing the number of cases and dissected areas used for the main study including loop electrosurgical excision procedure (LEEP; tissues, centre main panel) and punch biopsies for the pilot study (top main panel). For some cases several CIN1 or CIN3 areas were dissected per cervix[A]. Sections of five CIN1 and eight CIN3 cases had no normal epithelium [B]. Sections of ten CIN3 cases had no adjacent- CIN1 lesions [C]. Mean methylation levels were averaged per lesion type[D]. HPV genotyping was done separately for all lesions. Our study also included a comparison of HPV typing data from the 127 LEEP to corresponding exfoliated cell specimens (right panel). Abbreviations; Adj.: adjacent, AVER: data averaged per lesion type, CIN: cervical intraepithelial neoplasia, IMP: data imputation, Prin.: principal, vs.: versus,

Figure 3. A paired analysis comparison of methylation levels between adjacent cervical tissue samples from the same women. Methylation levels of sample pairs from unifocal principal-CIN1 cases are shown on the left panel and sample pairs from multifocal CIN3 cases on the right panel for human gene *EPB41L3* (A) and HPVme-All (B). Adjacent normal tissue samples were significantly different from their concurrent CIN1 in both unifocal and multifocal cases (*EPB41L3* unifocal: p=0.0007, and multifocal: p=0.005; HPVme-All unifocal: p=0.0002 and multifocal cases: p=2e-04). In multifocal CIN3 cases, the CIN3 lesions were significantly different from the adjacent-CIN1 lesions (*EPB41L3*: p=0.008; HPVme-All: p=0.011). All pairwise comparisons were tested with Wilcoxon matched-pairs signed rank tests. Abbreviations; Adj.: adjacent, Prin.: principal.

Figure 4. Differences in DNA methylation levels between adjacent lesions from the same cervix for *EPB41L3* (A) and HPVme-All (B). The median differences were all positive and a large majority of individual differences were also positive, suggesting an increase of methylation with the severity of the lesion in cervical tissues of each woman. As expected the biggest difference was found between normal tissue samples and their adjacent CIN3 lesions in the multifocal CIN3 cases. Abbreviations; Adj.: adjacent, Prin.: principal

Figure 5. DNA methylation levels of normal (white), CIN1 (light grey) and CIN3 (dark grey) samples, analysed in an unpaired manner. Methylation levels of CIN1 specimens were similar whether originating from principal-CIN1 cases (indicated by CIN1<CIN1) or from adjacent-CIN1 near to CIN3 (CIN1<CIN3) for both A) *EPB41L3* (p=0.497) and B) HPVme-All (p= 0.110). Methylation patterns were also similar for normal tissues taken from CIN1 cases (normal<CIN1) and CIN3 cases (normal<CIN3) for *EPB41L3* (p=0.212) and HPVme-All (p=0.163). All comparisons tested with Mann-Whitney tests.

Figure 6. Percentage of samples infected by hrHPVs. In all sample types, single HPV infections were more frequent than multiple infections. Adjacent normal tissue samples presented a higher number of hrHPV negative samples (23% and 27% in unifocal principal-CIN1 and multifocal CIN3 cases respectively) compared to the intraepithelial lesions: principal-CIN1 (4%), adjacent-CIN1 (3%) and CIN3 (1%). Most CIN1 lesions were infected by hrHPV types (87% of principal-CIN1 and 97% in adjacent-CIN1). All CIN3 samples, except one, were infected by hrHPV types. Abbreviations; Adj.: adjacent, Norm: normal, Prin.: principal.

Supplementary Figures legends
Supplementary figure 1. Methodology for sectioning, annotating and dissecting LEEP tissue blocks. Twelve 5um sections per block were cut. The first and last sections were used for haematoxylin and eosin staining (H&E) and sent for histopathological review and annotation. The annotated slides were used to dissect the CIN lesions and normal tissues of the remaining ten unstained sections. This material was used to purify DNA for subsequent analysis. Abbreviations; Adj: adjacent; Norm: normal; Prin: principal.

Supplementary figure 2. DNA methylation levels of normal, CIN1 and CIN3 punch biopsies for human genes *EPB41L3* (A) and *MAL* (B), HPV16 L1 region (C), L2 region (D), E2 binding site 1 (E), E2 binding sites 3 and 4 (F) and HPV 18 L2 region (G). Significance level is shown with the horizontal bars above the boxplots; KWT χ^2: Kruskal-Wallis test statistic, CTT χ^2: Cuzick test for trend statistic, ns: non-significant, *: $p<0.05$, **: $p<0.001$, ***: $p<0.0001$.

Supplementary figure 3. Comparison of HPV types found between adjacent tissue samples of different lesion grades. In the CIN3 cases, 87% of the adjacent-CIN1 lesions presented the same (or compatible) HPV type(s) as their nearby CIN3 lesions. A compatible type was defined as samples containing a matching HPV type in addition to other HPV types. Normal tissue samples showed fewer similarities when compared to adjacent-CIN1 (39%) or CIN3 samples (33%) in the multifocal CIN3 cases. This number amounted to 45% in principal-CIN1. A substantial number of normal samples were HPV negative and no comparison could be made for these samples. Abbreviations; Adj: adjacent; Norm: normal; Prin: principal.

What's new? (Lay abstract)
A better understanding of molecular mechanisms in skin tissue leading from human papillomavirus (HPV) infection to precancerous lesions is key to efficient prevention and treatment of HPV disease. Despite best efforts, accurate prediction of cancer progressive potential in HPV-infected patients remains inconsistent and elusive. We studied DNA methylation patterns to investigate models of HPV disease progression in areas of the uterine cervix with complex multi-focal precancerous cervical lesions. Carefully isolated disease foci of different grades were tested for molecular signatures. Methylation levels were much higher in women with high grade disease than low grade disease. Low grade lesions adjacent to high grade lesions predominantly shared the same HPV types whereas their DNA methylation patterns differed substantially; however, in women where only low grade disease was found on the cervix the methylation in these lesions was indistinguishable from low grade lesions adjacent to high grade lesions. We propose that progression from normal HPV-infected epithelium to low grade lesions or high grade lesions is generally promoted by the same persistent HPV type infections but occurs via distinct DNA methylation processes. Our data support a "molecular switch" model of precancer development which can lead to improved quantitation functions for cancer risk.