

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
04 June 2020 (04.06.2020)

(10) International Publication Number
WO 2020/109866 A1

(51) International Patent Classification:

CI2Q 1/70 (2006.01) *CI2Q 1/6806* (2018.01)

(21) International Application Number:

PCT/IB2019/001313

(22) International Filing Date:

29 November 2019 (29.11.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/773,831 30 November 2018 (30.11.2018) US

(71) Applicant: **GENOMIC VISION** [FR/FR]; 80-84, rue des Meuniers, 92220 Bagneux (FR).

(72) Inventors: **BENSIMON, Aaron**; 5 rue du Clos de l'Abbaye, 92160 Antony (FR). **BOUCHILLOUX, Stéphanie**; 78, rue Anatole France, 92290 Chatenay-Malabry (FR). **FER, Frédéric**; 9 rue Paul Bert, 2 allée Traversière, 92700 Colombes (FR). **CAVAREC, Laurent**; 21, rue Crébillon, 94300 Vincennes (FR). **MAHE, Florence**; 3 rue des ormes, 94320 Thiais (FR).

(74) Agent: **ERNEST GUTMANN - YVES PLASSERAUD**
SAS et al.; 66, rue de la Chaussée d'Antin, 75009 Paris (FR).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

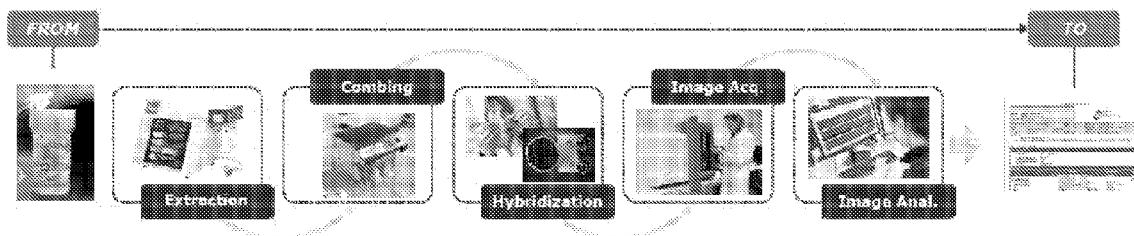
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: ASSOCIATION BETWEEN INTEGRATION OF VIRAL AS HPV OR HIV GENOMES AND THE SEVERITY AND/ OR CLINICAL OUTCOME OF DISORDERS AS HPV ASSOCIATED CERVICAL LESIONS OR AIDS PATHOLOGY

FIG. 4



(57) Abstract: The invention concerns the detection and the quantification of integrated nucleic acids of viruses and thus the detection and follow-up of reservoir cells harbouring such integrated viral genomes. One aspect of the invention is directed to a method for detecting a level of integrated viral DNA that includes removing episomal viral or vector nucleic acids from genomic DNA in a cell sample, and quantifying a number of integrations of viral DNA into the genomic DNA of the cells by a method of amplification of an integration region in the DNA sample; thereby detecting a level of integrated viral DNA in the genomic DNA from a cell sample, such as a biological sample containing HPV virus and DNA.

ASSOCIATION BETWEEN INTEGRATION OF VIRAL AS HPV OR HIV GENOMES
AND THE SEVERITY AND/OR CLINICAL OUTCOME OF DISORDERS AS HPV
ASSOCIATED CERVICAL LESIONS OR AIDS PATHOLOGY

CROSS-REFERENCE TO RELATED APPLICATIONS

This application falls within a similar technical field as U.S. Application No.15/976,758 filed May 10, 2018 (or in WO2018/207022) and entitled *Association between Integration of High-Risk HPV Genomes Detected by Molecular Combing and the Severity and/or Clinical Outcome of Cervical Lesions*; Mahiet, et al., US 2016/0047006 A1, filed March 4, 2015, entitled *Diagnosis of Viral Infections by Detection of Genomic and Infectious Viral DNA by Molecular Combing*; Lebofsky, et al., U.S. Patent No. 7,985,542 B2, filed September 7, 2006 entitled *Genomic Morse Code*; and Lebofsky, et al., U.S. Patent No. 8,586,723 B2, filed September 5, 2007 entitled *Genomic Morse Code*. Each of these documents is incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Field of the Invention. The invention falls within the fields of virology and molecular biology, especially as applied to medical diagnostics, prognostics and therapy.

Description of the Related Art. Cervical cancer was the tenth most common cancer in terms of frequency in Czech women in 2013 there were 895 new cases for a standardized incidence rate of 11/100,000 person-years (105th position in the world) with 388 deaths for a standardized mortality rate of 3.76/100,000 person-years (130th position in the world).

The onset of this cancer is associated with persistent infection by one or several high-risk human papillomaviruses (HR-HPV) and has been recognised by the WHO as attributable in nearly 100% of cases to these viruses. The most common genotypes associated with cervical cancer are HPV genotypes 16, 18, 31, 33 and 45, which are responsible for more than 80% of these cancers. Because of its slow progression, cervical cancer can be prevented by screening

and the treatment of the precancerous lesions that precede it. Ministry of Health of the Czech Republic initiated the National programme of the cervical cancer screening in 2008. The aim of the programme is early detection of cervical cancer. The screening is available for all women older than 15 years. The screening is currently based on the cytological examination of Pap smears once per year. In case of positive finding the cytological examination is performed again after 4 months in ASC-US or after 7 months for LSIL.

Recently, some European countries (Netherlands, Belgium) have planned to abandon cytological screening for the HPV screening test, which is considered to be more sensitive (> 90%) than cytology in detecting precancerous and cancerous lesions of the uterine cervix. The slightly lower specificity of HPV testing requires the use of a triage test to avoid excessive use of colposcopy. The procedure that has been best assessed to date is a cytological screening examination. However, other triage tests that would specifically identify patients at risk of lesion progression may have their place in this screening strategy.

Integration of the HPV genome/ Viral integration, a major event in tumour progression.

HR-HPV infection is considered to be the major cause of cervical cancer (zur Hausen 2002). However, most infections are spontaneously cleared by the immune system, while some persist for several years and sometimes progress to cancer (Crosbie, Einstein et al. 2013). HR-HPV infection is therefore a necessary but not sufficient cause of cervical cancer. Integration of the high-risk HPV genome in the host genome is considered to be a key event in the development of cervical cancer and as one of its most important risk factors (Pett and Coleman 2007). During the initial infection phase, HPV is present as a nuclear episome; however the integration of HR-HPV DNA into the host genome is a major step in the progression of cervical neoplasia (Wentzensen, Vinokurova et al. 2004). The integration of the HR-HPV genome in the host cell genome gives these cells a strong selective advantage promoting the clonal expansion of this cell population. Three major mechanisms play an important role in this process: Firstly, it has

been reported that the integration of the HR-HPV genome frequently and preferentially causes a deletion in the open reading frame (ORF) of the viral E2 gene (Choo, Pan and al. 1987). The E2 protein is a negative regulator of the viral E6/E7 promoter. One of the consequences of this integration is the loss of control of expression of viral oncoproteins E6 and E7, leading to their stable over-expression. E6 and E7 target the p53 and pRb tumour suppressors respectively, negatively regulating their anti-tumour function (Dyson, Howley et al. 1989, Scheffner, Werness et al. 1990). Secondly, the integration of HR-HPV increases the stability of the E6 and E7 viral transcripts derived from integrated viral genome copies (von Knebel Doeberitz, Bauknecht et al. 1991), thus increasing their level of expression and their oncogenicity (Jeon, Allen-Hoffmann et al. 1995, Jeon and Lambert 1995). Finally, the integrated viral genes may activate the cellular oncogenes or inactivate tumour suppressor genes close to their integration sites. This is the case, for example, for the c-Myc oncogene, the protein expression of which increases when HR-HPV is integrated in the adjacent regions (Wentzensen, Ridder et al. 2002, Ferber, Thorland et al. 2003, Peter, Rosty et al. 2006, Hu, Zhu et al. 2015). A molecular combing study showed that integration of HR-HPV at this locus was associated with a strong genetic instability of this genomic region (Herrick, Conti et al. 2005) sometimes leading to malignant progression.

Viral integration: a potential diagnostic and prognostic marker. From a morphological point of view, nothing distinguishes a lesion that will regress or persist from one that will progress to invasive cancer. In most cervical carcinomas, the HPV genome is integrated, whereas it is mainly in episomal form in low-grade lesions (Klaes, Woerner et al., 1999, Hopman, Smedts et al. 2004, Wentzensen, Vinokurova et al. 2004, Vinokurova, Wentzensen et al. 2008, Hu, Zhu et al. 2015). While it is recognised that integration of HR-HPV is detected in high-grade precancerous lesions and malignant tumours, an increasing number of studies support the idea that HPV integration may take place at a much earlier stage of cervical

carcinogenesis, in lower grade lesions (Kulmala, Syrjanen et al. 2006, Cricca, Morselli-Labate et al. 2007, Huang, Chao et al. 2008, Gradissimo Oliveira, Delgado et al. 2013, Vega-Pena, Illades-Aguiar et al. 2013, Zubillaga-Guerrero, Illades-Aguiar et al. 2013, Marongiu, Godi et al. 2014, Hu, Zhu et al. 2015). Some of these studies have even demonstrated the presence of integrated HR-HPV DNA in women with normal cytology (Kulmala, Syrjanen et al. 2006, Gradissimo Oliveira, Delgado et al. 2013, Vega-Pena, Illades-Aguiar et al. 2013, Zubillaga-Guerrero, Illades-Aguiar et al. 2013, Marongiu, Godi et al. 2014). Some data also suggest that the rapid progression of early cervical lesions to high-grade lesions is closely associated with the integration of HPV 16 and in particular, a high integrated HPV16 viral load (Peitsaro, Johansson et al. 2002, Vega-Pena, Illades-Aguiar et al. 2013). Moreover, the HPV integration appears to be a risk factor for progression of precancerous CIN2-3 lesions to the cancer stage (Hopman, Smedts et al. 2004).

The detection of integration of high-risk HPV genomes in the host genome may therefore provide a useful marker to identify lesions at high risk of progression that will require treatment. It would then be possible to reduce the number of unnecessary colposcopies and the over-treatment of lesions that spontaneously regress.

Methods for detecting viruses are described by Anderson, et al., CA 2943626, filed March 19, 2015, entitled “HPV16 Antibodies as Diagnostic and Prognostic Biomarkers in Pre-invasive and Invasive Disease”; Lebofsky and Bensimon, *Briefings in Functional Genomics and Proteomics*, vo. 1 No. 4. 385-396 (January 2003), entitled “Single DNA molecule analysis: Applications of Molecular Combing”; Raybould, et al., *Journal of Virological Methods* 206 (2014) 51-54, entitled HPV integration Detection in CaSki and SiHa using detection of Integrated papillomavirus sequences and restriction-site PCR”; Meoiger, et al., EP 3106524 A1, filed October 14, 2013, entitled “PRDM14 and FAM19A4, Molecular Diagnostic Markers for HPV-induced Invasive Cancers and Their High-Grade Precursor Lesions”; Peitsaro, et al., *J.*

Clinical Microbiology, Mar. 2002, p. 886-891, entitled Integrated Human Papillomavirus Type 16 is Frequently Found in Cervical Cancer Precursors as Demonstrated by a Novel Quantitative Real-Time PCR Technique”; Jansen-Durr, et al., U.S. 2016/0237143 A1, filed January 6, 2016, entitled Anti-HPV E7Antibodies”; Luft, et al., Int. J. Cancer 92, 9-17 (2001), entitled “Detection of Integrated Papillomavirus Sequences by Ligation-Mediated PCR (DIPS-PCR) and Molecular Characterization in Cervical Cancer Cells”; Hu, et al., Nature Genetics, published online 12 January 2015; doi:10.1038/ng.3178, entitled “Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism”; and Lee, et al., U.S. 2016/0376662 A1, filed December 5, 2015, entitled “Improved Cervical Cancer Diagnosing Method and Diagnostic Kit for Same”.

Molecular combing procedures and Genomic Morse Code procedures are described by the cross-referenced applications above (WO2018/207022 filed May 11, 2018, US 2016/0047006 A1, filed March 4, 2015, U.S. Patent No. 7,985,542 B2, filed September 7, 2006, U.S. Patent No. 8,586,723 B2, filed September 5, 2007).

Due to its slow progression, cervical cancer can be prevented by screening and the treatment of the precancerous lesions that precede it. This screening is in European countries currently based on the cytological examination of cervical (Pap) smears. The national guidelines specify, for each type of anomaly, the cases in which colposcopy is indicated in order to take a biopsy sample to complete the diagnostic process. The integration of the high-risk HPV genome in the cell genome is considered to be a key event in the development of cervical cancer and as one of its most important risk factors. In most cervical carcinomas, the HPV genome is integrated, whereas it is mainly in episomal form in low-grade lesions. Other data also suggest that the rapid progression of early cervical lesions to high-grade lesions is closely associated with the integration of HPV. Detecting the integration of high-risk HPV

genomes in the cellular genome may therefore provide a useful marker for the identification of high-grade lesions or lesions at high risk of progression. This would make it possible to reduce the number of unnecessary colposcopies, avoid over-treatment of lesions that spontaneously regress and better target the lesions requiring treatment. In view of the morbidity associated with cervical cancer and the unmet need for a sensitive and specific way of diagnosing, monitoring and prognosing HPV-associated diseases, the inventors investigated application of Molecular Combing techniques. Results from these molecular combing studies have now been obtained and higher throughput procedures were investigated with a new objective of development of an early diagnostic or prognostic test that specifically measures a level of viral integration.

One of the major limitation of previous technologies was that they were not able to discriminate an episomal form and integrated form of a virus. This was a significant drawback as presence of virus nucleic acids in a free, episomal form and presence of a virus nucleic acid as DNA integrated into a host cell chromosome exert different biological effects, for example, a virus integrated into host genomic DNA can interrupt genes present in the host genomic DNA, leading to disruption of cellular processes controlled by the interrupted genes, such as processes leading to or associated with cancer.

In view of these technical problems and limitations of existing methods, the inventors sought to develop a new method to specifically detect and quantify viral DNA that has been integrated into the genomic DNA of a host cell or host cells. As shown herein, these methods can correlate the level of virus integration, such as HPV integration, with the severity and outcome of disrupted cellular functions, such as the severity of lesions associated with integration of HPV DNA into a host genome and predict or prognose the risk of particular outcomes of virus infection. The inventors also sought to develop a test that can detect at a very early step in viral integration, for example, to help diagnose and prognose an HPV

infection that may later be associated with or progress toward cervical cancer or toward a more malignant form of cervical cancer.

The inventors have now developed superior methods able to quantify HPV or HIV integration in a host genome, such as in human genomic DNA, as well as polynucleotide compositions that include biomarkers that cover a host gene sequence that has been disrupted by integration of a viral genome, such as the HPV or HIV genome and parts of the genome of an integrated virus

BRIEF DESCRIPTION OF THE INVENTION

The invention concerns the detection and the quantification of integrated nucleic acids of viruses and thus the detection and follow-up of reservoir cells harbouring such integrated viral genomes. One aspect of the invention is directed to a method for detecting a level of integrated viral DNA that includes removing episomal viral or vector nucleic acids from genomic DNA in a cell sample, and quantifying a number of integrations of viral DNA into the genomic DNA of the cells by a method of amplification of an integration region in the DNA sample; thereby detecting a level of integrated viral DNA in the genomic DNA from a cell sample, such as a biological sample containing HPV virus and DNA. This method provides a superior way of detecting the numbers of integrated viral DNA by reducing or eliminating episomal nucleic acid sequences, such as episomal HPV DNA. The method may further include detecting at least one biomarker that covers a junction between a disrupted host gene and at least part of an integrated HPV DNA. Novel biomarkers identified by the inventors include MAPK10 (Gene ID: 5602), PTPN13 (Gene ID: 5783), NUDT15 (Gene ID: 55270), MED4 (Gene ID: 29079), ITM2B (Gene ID: 9445), RB1 (Gene ID: 5925), LPAR6 (Gene ID: 10161), RAB11A (Gene ID: 8766), RPL13A (Gene ID: 23521), ZNF341 (Gene ID: 84905), OFD1 (Gene ID: 8481), DHRS3 (Gene ID: 9249), TBC1D22B (Gene ID: 55633), AFF3 (Gene

ID: 3899), CXCL6 (Gene ID: 6372), PF4V1 (Gene ID: 5197), IMMP2L (Gene ID: 83943), MMP12 (Gene ID: 4321), WDR20 (Gene ID: 91833), ALDHA1A (Gene ID: 216), TPRG1 (Gene ID: 285386), TUBD1 (Gene ID: 51174), MAST4 (Gene ID: 375449), LOC100132167, NFIX (Gene ID: 4784), CCAT1 (Gene ID: 100507056), GPR137B (Gene ID: 7107), RAB22A (Gene ID: 57403), C9orf3, MACROD2 (Gene ID: 140733), DACH1 (Gene ID: 1602), ATP10A (Gene ID: 57194), SPG11 (Gene ID: 80208), SORD (Gene ID: 6652), COL4A4 (Gene ID: 1286), GATSL1 (Gene ID: 729438), GATSL2 (Gene ID: 729438), MAP2 (Gene ID: 4133), EPN1 (Gene ID: 29924), ATXN3L (Gene ID: 92552), EGFL6 (Gene ID: 25975), and MAGI2 (Gene ID: 9863).

These biomarkers include polynucleotides having human sequences preferably ranging from 50 bp to 10 kb and polynucleotides having viral sequences, such as HPV or HIV sequences, ranging from 50 bp to 10 kb. These biomarkers are useful in methods of assessing risk, diagnosing and prognosing diseases, disorders and conditions associated with virus-infected host cells, for example, these biomarker sequences can be used for as probes for the detection of an integrated viral DNA in genomic DNA of a host cell. To be used as probes, these biomarkers are detectable and accordingly may be labelled such as labelled to be rendered fluorescent (fluorescent probes). These biomarkers may be formulated as a kit containing reagents and other supplies and/or equipment useful for detecting viral integration. Such a kit may also include control reagents so that controlled comparisons of a level of viral integration in an infected cell or cell sample may be made.

The invention encompasses technologies that allow an early and high throughput measurement of the HPV or HIV virus integrated into a host genome based on the results from a molecular combing study and biomarker discovery. These technologies are extendible to any other virus infection exhibiting free and integrated forms of viral DNA in a host cell.

BRIEF DESCRIPTION OF THE FIGURES

This application file contains at least one drawing executed in color. Copies of this patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1. Integration yes/no A-interim analysis: B- Final cross-sectional analysis: High Grade (HG) compared to the normal group ($p=0.0006$); Low Grade (LG) + HG compared to the normal group ($p = 0.0012$).

FIG. 2 shows the number of integrations /genome in the High Grade (HG) compared to the normal group ($p= 0.0001$). interim analysis.

FIG 3. Number of integration sites per genome in histology groups final cross-sectional analysis.

FIG. 4. Description of the analysis of samples by Molecular Combing

FIG. 5. Physical principle of molecular combing. Examples of YOYO-1-stained DNA molecules, combed at different densities, visualized under a fluorescence microscope at 40x magnification.

FIG. 6. Study design diagram.

FIG. 7. HPV 18 genome coverage by the 3 probes: one covers the region containing genes L1 and L2, the second viral genes E1 and E2, and the third viral oncogenes E6 and E7. The probes of the other thirteen genotypes (HPV 16, 31, 33, 45, 35, 39, 51, 52, 56, 58, 59, 66 et 68) are drawn on the same model. The probes corresponding to regions L1L2 and E1E2 are displayed in blue and the one corresponding to region E6E7 in cyan.

FIG. 8A. Example of signal corresponding to a single integration of an HPV 16 genome. FIGS. 8A-8D exemplify signals corresponding to integrated HPV genomes and reference signals.

FIG. 8B: 4 examples of signals corresponding to 4 integration sites, each containing multiple integrations of juxtaposed HPV 16 genomes. The cyan signals, corresponding to probe

E6E7, make it possible to visually count the number of integrated HPV genomes per integration site;

FIG. 8C: Three examples of signals corresponding to three integration sites, each containing multiple integrations of dispersed HPV 18 genomes, spaced by the host DNA.

FIG. 8D: Examples of reference signals generated by the probes covering a defined DNA locus of known size.

FIG. 9. Diagram of types of signals detected on the slides.

FIGS. 10A and 10B: Examples of integration of the HPV16 viral genome in the OFD1 gene.

DETAILED DESCRIPTION OF THE INVENTION

HPV refers to human papilloma virus. Human papillomavirus (HPV) is a group of viruses that are extremely common worldwide. There are more than 100 types of HPV, of which at least 14 are cancer-causing (also known as high risk type). Two HPV types (16 and 18) cause 70% of cervical cancers and precancerous cervical lesions. Vaccines against HPV 16 and 18 have been approved for use in many countries.

HPV DNA includes the polynucleotides of the HPV genome, individual genes of the HPV genome, as well as fragments of the HPV genome or genes, such as polynucleotide fragments that are recognized by probes to HPV polynucleotide sequences.

Control Subject or Patient. Those skilled in the art are aware of the value of controlled comparisons. A control subject or patient includes a subject who has not been exposed to HPV, has no symptoms or indicia of HPV infection, has no or a lower antibody titer to HPV, who exhibits a lower cellular response to HPV, or who exhibits fewer or no integrations of HPV DNA into genomic DNA, than that of a patient being evaluated. Positive control patients or subjects include those known to have integrated HPV DNA in their genomic DNA, who exhibit

symptoms of HPV infection, and/or who have HPV related cancers or conditions, such as those cancers or conditions described herein.

Molecular Combing has its conventional meaning as described in the applications, publications and other documents of the prior art as cited herein and incorporated by reference herein. Molecular Combing is carried out as a step of stretching nucleic acid, extracted from any source to be assessed as a sample to provide immobilized nucleic acids in linear and parallel strands (aligned nucleic acids). Molecular Combing is thus preferably performed with a controlled stretching factor (such as using a meniscus) formed on an appropriate surface (e.g., surface-treated glass slides). After stretching, it is possible to hybridize sequence-specific probes detectable for example by fluorescence microscopy (Lebofsky, Heilig et al. 2006). Thus, a particular nucleic acid sequence may be directly visualized on a single molecule level. The length of the fluorescent signals of the probes and/or their number, and/or their spacing on the slide provides a direct reading of the size and relative spacing of the probes. Some current techniques for the detection of HPV integration are described below. In particular molecular combing applied to samples may comprise the following steps : in a first step extracting very high molecular weight DNA from a biological specimen (blood, smear, etc.). Once extracted, the DNA molecules are "combed", i.e., attached by their ends to a silane-coated glass slide and uniformly stretched by a receding air-water interface (Bensimon 1994).

Once they are irreversibly fixed in this configuration to the glass substrate, DNA molecules are hybridised with a set of fluorescent probes specific for the DNA sequences of interest in order to obtain the specific fluorescent signature of this DNA region. In addition, a probe for a reference region may be used in order to normalize the result with respect to the number of cellular genomes combed on the coverslips. Only viral forms integrated in the cellular genome will be combed and analysed. Circular episomal forms are not analysed as they have no free DNA terminal to be combed.

After hybridisation, the slide is placed in a scanner in order to acquire images by epifluorescence microscopy corresponding to all the fields of view of the slide. Using specific software developed by Genomic Vision and commercialized for other applications under the trademark Fibervision 2016, it is possible to identify, among the thousands of fields of view, the regions of interest on the coverslip containing a fluorescent signal and to measure the size of these different signals. This last step is made possible by the presence of a constant stretching factor (such as a factor of 2 Kb/μm) which guarantees the determination of the physical distances within the region studied, by direct measurement of the probes and their spacing.

This approach can be used to study viral integration (i) with a high resolution, of approximately one kilobase (kb), including tandem integrations, (ii) directly without initial amplification of the genetic material and therefore with no selection bias associated with a choice of primer (iii) independently of viral transcription, and (iv) in a quantifiable and objective manner.

Routine techniques for the diagnosis of HR-HPV infection (detection of HPV DNA or E6E7 mRNA) and genotyping make it possible to demonstrate a persistent infection or to assess the presence of multi-infections and estimate the specific prevalence of types. However, they give an indirect picture of the oncogenic process, as none can describe the genomic integration of HR-HPV DNA in the cellular genome and none are able to quantify specifically the number of HPV virus integrated in host genome (HPV integrated forms or DNA). Techniques for detecting the integration of HR-HPV are currently only used for research purposes and all have limitations in this application. There is a real unmet need for an early diagnostic and/or prognostic test (kit) that would allow tracking and quantifying HPV integrated form at a very early step of patient management.

Integration of HR-HPV can be directly detected in cells by in situ hybridisation (ISH) or by fluorescence in situ hybridisation (FISH), frequently used in cytogenetics. Integrated and episomal forms can be distinguished by the diffuse (episomal) or punctuate (integrated) signal pattern. The interpretation of these patterns is partly subjective and therefore generates a variability of results between operators. Moreover, the resolution of these techniques is limited to 1-5 Mb (megabase) due to the steric hindrance of condensed DNA molecules and does not allow fine analysis of this integration (Lebofsky and Bensimon 2003). Southern blotting is also used to analyse the integration of HR-HPV. Although this method is relatively reliable, it is cumbersome and the use of radiolabelled probes has implications for the safety of operators. Integration can also be analysed by the detection of viral–cellular fusion transcripts, as is the case with the amplification of papillomavirus oncogene transcript (APOT) technique (Klaes, Woerner et al. 1999). However, this method requires extraction and amplification of mRNA which is much less stable than DNA. Moreover, it introduces bias, since it detects only transcriptionally active HPV DNA and it has been shown that in the CaSki cell line, which has a large number of genomes integrated in tandem, few seem to be actively transcribed (Van Tine, Knops et al. 2001, Ziegert, Wentzensen et al. 2003). Another technique is also based on the detection of the fusion of human and viral sequences, but at the level of DNA: DIPS (Detection of Integrated papillomavirus Sequences) (Luft, Klaes et al. 2001). However, this method is limited by its lack of sensitivity in the detection of integration when there are high concentrations of episomal forms and tandem integrations (Raybould, Fiander et al. 2014). The real-time PCR technique has these same limitations. This method determines the ratio between the number of copies of the viral E2 gene and the viral oncogene E6 (Peitsaro, Johansson et al. 2002). It is based on the fact that when the HPV genome is integrated, E2 is deleted in the vast majority of cases and E6 conserved. However, it was recently shown that cleavage during integration could occur at any site on the HPV genome, including in E6 and predominantly in

E1 (Hu, Zhu et al. 2015). All integrations resulting from a deletion outside E2 and all tandemly integrated forms which conserve intact HPV genomes are not therefore detected by this technique. Finally, the HIVID technique (High-throughput Viral Integration Detection) based on high-throughput (new-generation) sequencing (NGS) combined with computer analysis of reads, makes it possible to precisely locate HPV integration sites (Hu, 2015.1). However, this technique is cumbersome and complicated and cannot be used to determine the number of copies integrated at a locus.

All these techniques have been used to increase our understanding of the role of HPV integration in precancerous lesions and cervical cancers. The diversity of these techniques and their limitations explain the variability and sometimes divergence between the data obtained in this field as most of these techniques do not provide an unbiased, sensitive and high-resolution analysis of the integration loci. Other embodiments of the invention as based on removal of episomal DNA from a biological sample include the following.

Embodiment 1. A method for detecting a level of integrated viral DNA comprising:
removing episomal viral or vector nucleic acids from genomic DNA in a cell sample,
and
quantifying a number of integrations of viral DNA into the genomic DNA of the cells
by a method of amplification of an integration region in the DNA sample;
thereby detecting a level of integrated viral DNA in the genomic DNA from the cell
sample; and

optionally, performing a method of amplification on episomal nucleic acids removed
from the sample.

The method described in the embodiment above and other methods described herein
may be used to remove episomes with an efficiency of up to 95% based on the total number of
episomes in a sample, for example, it can remove 10, 20, 30, 40, 50, 60, 70, 75, 80, 85, 90, 95,

96, 97, 98, 99 or >99% of the episomes in a sample. In some embodiments one or more methods described herein may be used in combination to remove episomes. Preferably no more than 5% of the episomal nucleic acids remain in a sample of genomic DNA undergoing quantification of integrated viral DNA. This method also permits the quantification of both integrated viral DNA and episomal DNA or nucleic acids by quantifying each kind of DNA using PCR or other amplification methods. The methods encompassed by embodiment 1 may be applied or adapted for human as well as non-humans, such as livestock mammals susceptible to infection with an integrative or episomal virus.

In this and similar embodiments, the method may be used to data mine the genomic DNA of an individual or group of individuals to quantify an amount of dormant viral or other exogenous DNA incorporated into the genome for example endogenous retroviral insertions (ERV's) or retrovirus proviruses and distinguish dormant integrated DNA from episomes carrying the same or similar DNA or RNA. This aspect of the invention is one way a viral reservoir may be assessed and quantified.

Embodiment 2. A method according to embodiment 1 wherein the amplification method is quantitative polymerase chain reaction (qPCR), fluorescent *in situ* hybridization (FISH) or molecular combing.

Embodiment 3. The method of embodiment 1 or 2, wherein said removing comprises: permeabilizing cell membranes in the cell sample by exposing the cells to an extracting salt solution for a time and under conditions sufficient to make the cellular membranes permeable to episomal viral or vector nucleic acids from the cells and for the episomal viral or vector nucleic acids to leak out of the cells into a medium, separating the permeabilized cells from episomal or vector nucleic acids in the medium, and performing qPCR on the separated permeabilized cells; and

optionally, isolating and performing qPCR on the episomal or vector nucleic acids in the medium. Methods for making membranes permeable are known in the art, such as those described by Moen, WO 2006/096727 A2 which are hereby incorporated by reference.

Embodiment 4. The method of any one of embodiments 1 to 3, wherein said removing comprises:

embedding cells into an agarose plug having an agarose concentration ranging from about 0.5 wt% to 1.5 wt%,

infusing the plugs containing the embedded cells with a proteolytic enzyme for a time and under conditions sufficient to substantially digest cellular proteins,

washing the plugs for a time and under conditions sufficient to remove episomal nucleic acids,

extracting cellular genomic DNA caught in the plugs and

performing qPCR on the genomic DNA extracted from the washed agarose plugs.

performing qPCR on the separated permeabilized cells; and

optionally, isolating and performing qPCR on the nucleic acids washed out of the agarose plugs. Methods for embedding nucleic acids in agarose are known in the art and are incorporated by reference to Fritz and Musich, , Biotechniques. 1990 Nov;9(5):542, 544, 546-50.

Embodiment 5. The method of any one of embodiments 1 to 4, comprising partially depleting episomal HPV nucleic acids by biotinylating episomal nucleic acids and binding them to (strept)avidin, thus removing them from a mixture of genomic and episomal nucleic acids. In this method partial depletion of the episomal HPV is carried out thanks to biotinylated oligonucleotides targeting the E2 sequence of the HPV genome. Single copy HPV integration results in the disruption of the E1 and E2 open reading frames (Peitsaro et al., 2002, J. Clin. Microbiology ; Vernon et al., 1997, Int. J. Cancer ; Jeon and Lambert, 1995, Proc. Natl. Acad.

Sci USA). In head-to-tail tandem repeats integrations of the HPV genome, only the viral copy flanking the cellular DNA is disrupted in the E1 or E2 region when the internal copies keep intact E1 and E2 open reading frames. With the proposed protocol, the episomal HPV is discarded from the cellular genomic DNA thanks to its capture on streptavidin coated beads.

Embodiment 6. The method of any one of embodiments 1 to 5, comprising:

isolating nucleic acids from the cell sample using a plasmid purification column,

recovering genomic DNA from material eluting from the plasmid purification column,

and

performing qPCR on the recovered genomic DNA; and, optionally,

isolating and performing qPCR on the material bound to the plasmid purification column.

Separation of episomal or plasmid DNA using a separation column is known in the art and incorporated by reference to Moller et al., 2014, Proc. Natl. Acad. Sci USA; Moller, et al., *Proceedings of the National Academy of Sciences of the United States of America*. **112** (24), E3114-E3122, (2015) and Moller et al., *Genome-wide Purification of Extrachromosomal Circular DNA from Eukaryotic Cells*. *J. Vis. Exp.* (110), e54239, doi:10.3791/54239 (2016).

Plasmid separation columns are commercially available, for example, from Promega. Such columns and methods of their use are incorporated by reference to <https://www.promega.com/-/media/files/resources/paguide/letter/chap9.pdf?la=en> (last accessed November 30, 2018, incorporated by reference).

Other modes of separating genomic DNA from plasmid or episomal DNA are also incorporated by reference to the link above.

Embodiment 7. The method of any one of embodiments 1 to 6, wherein the cell sample contains viral DNA.

Embodiment 8. The method of any one of embodiments 1 to 7, wherein the cell sample contains HIV nucleic acids.

Embodiment 9. The method of any one of embodiments 1 to 7, wherein the cell sample contains human papilloma virus (HPV) nucleic acids.

Embodiment 10. The method of embodiment 9, further comprising detecting by PCR a ratio of DNA encoding HPV E2 ORF to DNA encoding HPV E6 ORF, especially by qPCR; wherein the ratio of DNA encoding E2 ORF to DNA encoding E6 ORF represents an amount of an episomal form in relation to an integrated form.

The measurement of E2/E6 ratio requires a quantitative method rather than an end-point PCR. Real-time (quantitative) PCR is employed and other quantitative methods may be used., such as molecular combing associated with Genomic Morse Code hybridization as described by and incorporated by reference to U.S. Patent No. 8,586,723 and U.S. Patent No. 9, 133,514, or by fluorescent *in situ* hybridization (FISH). In these methods, the quantification is performed through relative measure of specific probe fluorescence.

Embodiment 11. The method of embodiment 10, wherein the ratio of E2 ORF to E6 ORF is determined by real-time PCR using a set of primers and probes described by Table 1 for each HPV genotype, 16, 18, 33 and/or 58.

Embodiment 12. The method of any one of embodiments 9 to 11, further comprising determining HPV viral load comprising determining a ratio between E6/E7 and beta globin, MSH2 or at least one other human gene target.

Embodiment 13. The method of any one of embodiments 9, to 12 further comprising detecting at least one biomarker that covers a junction between a disrupted host gene and at least part of an integrated HPV DNA. A biomarker used in a method of the invention is a

hybrid or chimeric human-viral (e.g., human-HPV) nucleic acid sequence, typically a DNA or modified DNA sequence. Biomarkers comprise a human sequence ranging from about 50 bp to about 10 kb linked to a viral sequence ranging from about 0.5 to 10 kb. Exemplary length ranges for the human component range from about 50, 100, 200, 500, 1,000 (1 kb), 2, 000 (2 kb), 5,000 (5 kb) to 10,000 (10 kb) bp as well as any intervening length. Exemplary length ranges for the human component range from about 50, 100, 200, 500, 1,000 (1 kb), 2, 000 (2 kb), 5,000 (5 kb) to 10,000 (10 kb) bp as well as any intervening length. Exemplary length ranges for a viral component range from about 50, 100, 200, 500, 1,000 (1 kb), 2, 000 (2 kb), 5,000 (5 kb) to 10,000 (10 kb) bp as well as any intervening length. A biomarker may be produced or amplified by PCR or produced by other methods known in the art such as by chemical synthesis or by recombinant DNA techniques. In particular a biomarker suitable for use according to the invention is a hybrid or chimeric human-viral nucleic acid sequence where the viral sequence is from HPV, in particular from an E6 ORF of HPV, especially from and E6 ORF such as E6 ORF of HPV selected from the group of HPV16, HPV18, HPV33 and HPV58. A biomarker may be labelled in order to be detectable.

Embodiment 14. The method of any one of embodiments 9 to 13, further comprising detecting at least one biomarker selected from the group consisting of MAPK10 (Gene ID: 5602), PTPN13 (Gene ID: 5783), NUDT15 (Gene ID: 55270), MED4 (Gene ID: 29079), ITM2B (Gene ID: 9445), RB1 (Gene ID: 5925), LPAR6 (Gene ID: 10161), RAB11A (Gene ID: 8766), RPL13A (Gene ID: 23521), ZNF341 (Gene ID: 84905), OFD1 (Gene ID: 8481), DHRS3 (Gene ID: 9249), TBC1D22B (Gene ID: 55633), AFF3 (Gene ID: 3899), CXCL6 (Gene ID: 6372), PF4V1 (Gene ID: 5197), IMMP2L (Gene ID: 83943), MMP12 (Gene ID: 4321), WDR20 (Gene ID: 91833), ALDHA1A (Gene ID: 216), TPRG1 (Gene ID: 285386), TUBD1 (Gene ID: 51174), MAST4 (Gene ID: 375449), LOC100132167, NFIX (Gene ID: 4784), CCAT1 (Gene ID: 100507056), GPR137B (Gene ID: 7107), RAB22A (Gene ID:

57403), C9orf3, MACROD2 (Gene ID: 140733), DACH1 (Gene ID: 1602), ATP10A (Gene ID: 57194), SPG11 (Gene ID: 80208), SORD (Gene ID: 6652), COL4A4 (Gene ID: 1286), GATSL1 (Gene ID: 729438), GATSL2 (Gene ID: 729438), MAP2 (Gene ID: 4133), EPN1 (Gene ID: 29924), ATXN3L (Gene ID: 92552), EGFL6 (Gene ID: 25975), and MAGI2 (Gene ID: 9863).

These biomarkers are novel markers identified by the inventors based on published HPV integration sites; see Holmes, et al., 2016, Genomic Medicine, *Mechanistic signatures of HPV insertions in cervical carcinomas* (npj Genomic Medicine (2016) 1, 16004; doi:10.1038/npjgenmed.2016, incorporated by reference.

The inventors combined primers localized in disrupted genes due to virus integration and primers in integrated HPV. The chromosomal locus and breakpoints for all these flanking genes are indicated in Table 1 of the Holmes publication which is shown below.

Table 1. Clinical and genomic status of 72 cervical carcinoma cases

Table 1. (Continued)

Clinical Status				Genomic Capture/HPV NGS Status			Gene annotation	CGH array		
Case No.	Age	Clinical stage	Tumor size	HPV type	HPV status	Chromosomal break (s) (q)	Breakpoint 1	Breakpoint 2	Genes disrupted or nearby (within 500 kb)	FDA profiles
142	64	IBb	62	88	MI-SC	2q36.3 (2) 7q11.23 (3) 8q23.1 (3) 19q13.42 (1)	228,820,329 72,594,506 167,273,893 56,187,255	228,820,331 74,849,315 ** 229,942,001 79,584,528	COL4A4 GATSL1, GATSL2 ZFP912, QMR1 EPNT1	4 x AMP 2 x AMP 1.7 x AMP 2 x AMP 2 x AMP 2 x AMP 2 x AMP 2 x AMP 3 x AMP
201	52	IB	NA	16	MI-SC	1p22.2 (6) 7q21.11 (3)	13,387,991 78,694,568	13,390,002 78,694,568	ATXNL, EGFL6 MAGE2	2 x AMP 2 x AMP 3 x AMP

Biomarkers may be detected by a variety of methods including by molecular combing and qPCR. Those skilled in the art may select a suitable biomarker or set of biomarkers for use in a method of the invention based on the information in this table including age, clinical stage of disease, tumor size, HPV type, genomic capture/HPV NGS status, gene annotation and CGH array.

Embodiment 15. The method of embodiment 13 or 14, wherein forward and reverse oligonucleotide primers that specifically amplify an integration junction between host genomic DNA and integrated HPV DNA are used to produce the biomarker.

Embodiment 16. The method of embodiment 13, wherein the biomarker is human OFD1 (NG_008872.1) and is produced and the forward and reverse primers described by Table 2 are used to amplify an OFD1 gene/HPV16 and the HPV16/OFD1 gene junctions as biomarkers. Biomarkers produced accordingly are illustrated in Figure 10 for junctions of OFD1 gene and HPV16 genome sequences and are in particular biomarkers with the sequences of SEQ ID No.40 and SEQ ID No.44.

Embodiment 17. The method of any one of embodiments 13 to 16, at least one specific nucleic acids sequence complementary to an integration junction between host genomic DNA and integrated HPV DNA is used for the detection by hybridization of a biomarker.

Embodiment 18. The method of any one of embodiments 13 to 17, wherein said HPV DNA is selected from a DNA from the group consisting of HPV strains 16, 18, 21, 33, 45, 52 and 58.

Embodiment 19. A composition comprising one or more biomarkers.

Embodiment 20. The composition of embodiment 19, wherein said one or more biomarkers is selected from the group consisting of MAPK10, PTPN13, NUDT15, MED4, ITM2B, RB1, LPAR6, RAB11A, RPL13A, ZNF341, OFD1, DHRS3, TBC1D22B, AFF3, CXCL6, PF4V1, IMMP2L, MMP12, WDR20, ALDHA1A, TPRG1, TUBD1, MAST4, LOC100132167, NFIX, CCAT1, GPR137B, RAB22A, C9orf3, MACROD2, DACH1, ATP10A, SPG11, SORD, COL4A4, GATSL1, GATSL2, MAP2, EPN1, ATXN3L, EGFL6, and MAGI2. Embodiments related to Embodiment 20 include vectors and DNA constructs comprising the biomarkers described by Embodiment 20 as well as vectors containing a hybrid or chimeric sequence containing all or part of a complete viral genome, part of a disrupted human (or mammalian) gene sequence. These vectors or their components may be used as probes for detecting DNA integrated into nucleic acids (e.g., chromosomal or genomic DNA) as described in Embodiment 1 and the subsequent embodiments depending from Embodiment 1.

In some embodiments, a DNA, such as a primer or probe DNA described herein will be modified and distinguishable from a DNA molecule in its natural state. In some embodiments of the invention, PCR or other primer- or probe-based nucleic acid detection techniques may be performed with a modified probe or primer, such as a probe or primer that is labeled with a reporter fluorophore and/or a quencher molecule. Such reporters and quenchers are known in the art, see, for example, <http://www.bio-rad.com/en-us/applications-technologies/introduction-pcr-primer-probe-chemistries?ID=LUSOJW3Q3> (incorporated by reference, last accessed October 30, 2018).

Primer or probe modifications include, but are not limited to, substitution in an oligonucleotide of: 2-aminopurine for dA; 2,6-diaminopurine for dA; deoxyuridine (dU) for dT; 5-methyl dC for dC to increase Tm up to 0.5°C; hydroxymethyl dC for dC; or deoxyinosine or 5-nitroindole as a “universal base” for any dA, dC, dG or dT in an oligonucleotide primer. Additional modifications include the incorporation or substitution of 5-hydroxybutyln-2'-deoxyuridine which is a duplex stabilizing modified base; incorporation or substitution of 8-aza-7-deazaguanosine that eliminates naturally occurring, non-Watson-and-Crick secondary structures associated with guanine-rich sequences; substitution of a locked nucleic acid base which has a modification to its ribose backbone that locks the base into a C3'-endo position, for one or more nucleotides in a primer. Other modifications include incorporation of inverted dT at a 3'-end of an oligonucleotide to inhibit degradation by 3' exonucleases or extension by DNA polymerases; incorporation of inverted dideoxy-T at a 5' end of a sequence to prevent unwanted 5' ligations, or incorporation of dideoxy-C as a 3' chain terminator. Other modifications are incorporated by reference to <https://www.idtdna.com/site/Catalog/Modifications/Category/7> (last accessed November 7, 2018) or the 3', internal, or 5' modifications described by and incorporated by reference to <https://www.thermofisher.com/us/en/home/life-science/oligonucleotides-primers-probes-genes/custom-dna-oligos/oligo-configuration-options.html#5prime> (last accessed November 30, 2018). One, two, three or more of these modifications may be incorporated into an oligonucleotide primer, probe or other nucleic acid sequence disclosed herein. A nucleic acid may also be modified by addition of a modified base, such as those described above at a 5' or 3' end.

Embodiment 21. A kit comprising standardized and purified biomarkers that hybridize with host cell DNA and with integrated viral DNA sequences, and optionally, control reagents, one or more other reagents, supplies and/or equipment useful for detecting viral integration.

Embodiment 22. A method of preparation of genomic DNA containing suspected integrated viral DNA from the cell sample comprising:

embedding cells into an agarose plug having an agarose concentration ranging from about 0.5 wt% to 1.5 wt%,

infusing the plugs containing the embedded cells with a proteolytic enzyme for a time and under conditions sufficient to substantially digest cellular proteins,

washing the plugs for a time and under conditions sufficient to remove episomal nucleic acids,

extracting cellular genomic DNA caught in the plugs containing or not the integrated viral DNA.

Embodiment 23. A method for assessing a risk of having or developing a cervical cancer comprising: detecting or quantifying a number of integrations of HPV DNA into, or an integration pattern of HPV DNA in a sample of genomic DNA obtained from a patient or subject, thereby assessing the risk of having or developing cervical cancer, wherein said sample of genomic DNA is obtained by the method of embodiment 1 or dependent embodiments.

Embodiment 24. The method of embodiment 23, wherein a greater number of instances, or a greater amount of, integrated HPV DNA is indicative of a high risk of having or developing cancer or is indicative of a more aggressive or higher grade cancer compared to a patient or subject having fewer instances or lesser amounts of integrated HPV DNA.

Embodiment 25. The method of embodiment 23 or 24, wherein a different pattern of HPV DNA integrations into genomic host DNA, compared to those in a control subject or patient, is indicative of a high risk of having or developing cancer or is indicative of a more aggressive or higher grade cancer.

Embodiment 26. The method of embodiment 23, 24 or 25 that comprises assessing a risk of having a cervical cancer.

Embodiment 27. The method of any one of embodiments 23 to 25 that comprises assessing the risk of developing a cervical cancer.

Embodiment 28. The method of any one of embodiments 23 to 25, wherein said HPV DNA is from a high risk or pathogenic strain of HPV.

Embodiment 29. The method of any one of embodiments 23 to 28, wherein said HPV DNA is from a high risk or pathogenic strain of HPV selected from the group consisting of HPV strains 16, 18, 21, 33, 45, 52 and 58.

Embodiment 30. The method of any one of embodiments 23 to 28, wherein said HPV is from a low risk or non-pathogenic strain of HPV.

Embodiment 31. The method of any one of embodiments 23 to 28, wherein said HPV is from a low risk or non-pathogenic strain of HPV that is less pathogenic than any one of HPV strains 16, 18, 21, 33, 45, 52 and 58.

Embodiment 32. The method of any one of embodiments 23 to 31, further comprising detecting or quantifying a number of integrations of HPV DNA by comparison to a patient or subject not infected with HPV, or not infected with a pathogenic strain of HPV, having no lesions or other symptoms of HPV infection, or having substantially no antibody titer or cellular immunity to HPV or to a particular HPV strain.

Embodiment 33. The method of any one of embodiments 23 to 32, further comprising detecting or quantifying a number of integrations of HPV DNA by comparison to those in an earlier biological sample obtained from the same patient or subject.

Embodiment 34. The method of any one of embodiments 23 to 33, wherein said detecting or quantifying a number of integrations is performed using molecular combing of the genomic host DNA using probes that bind to HPV DNA sequences.

Embodiment 35. The method of any one of embodiments 23 to 34, wherein said detecting or quantifying a number of integrations is performed using molecular combing of the

genomic host DNA using probes to HPV 16, 18, 31, 33, 35, 39, 51, 52, 56, 58, 59, 66 and 68.

Embodiment 36. The method of any one of embodiments 23 to 35, wherein said detecting or quantifying a number of integrations is performed using molecular combing of the genomic host DNA using probes that bind to or cover HPV DNA L1 and L2, E1 and E2, and/or E6 and E7 sequences, wherein said probes may be labelled with the same different colored fluorescent tags.

Embodiment 37. The method of any one of embodiments 23 to 36, wherein the patient or subject has a cervical dysplasia or has a positive PAP test.

Embodiment 38. The method of any one of embodiments 23 to 37, wherein the patient or subject has been infected with human immunodeficiency virus (HIV), is immunosuppressed, has been exposed to diethylstilbestrol before birth, or is or has been treated for a precancerous cervical lesion or cervical cancer.

Embodiment 39. The method of any one of embodiments 23 to 38, wherein the patient has or is at risk of having anal, vaginal, vulvar, penile or oropharyngeal cancer.

Embodiment 40. The method of any one of embodiments 23 to 39, wherein the number or pattern of HPV integrations is quantified by, or correlated, with at least one of the following:

the number of HPV integration sites in host genomic DNA or the average number of such integrations,

the size in kb of HPV DNA integrations into host genomic DNA,

the number of HPV genomes integrated at each integration site,

the presence of absence of integrated HPV DNA,

the number of HPV integration sites per cellular genome,

the average number of HPV integration sites in host cells,

the mean number of HPV genomes integrated per integration site (or the mean size of integration sites),

maximum number of HPV genomes integrated per integration site (or the maximum size of integration sites),

minimum number of HPV genomes integrated per integration site (or minimum size of integration sites), or

number of HPV genomes integrated per cellular genome.

Embodiment 41. The method of any one of embodiments 23 to 40, wherein the number or pattern of HPV integrations is correlated with at least one parameter of lesion status including:

normal histology (including all abnormalities without intraepithelial lesions or signs of viral infection such as metaplasia, cervicitis, decidual lesions or adenosis),

low grade (LG) lesion, corresponding to former CIN1,

high grade (HG) lesion, corresponding to former CIN2, 3 and CIS (carcinoma in situ) or AIS (adenocarcinoma in situ);

normal cervix,

Grade 1 atypical transformation (AT),

Grade 2 atypical transformation (AT),

TAG2 a if there are no major signs,

TAG2 b if there are major signs,

TAG2 c when the appearance is suggestive of invasive cancer; and/or

atypical transformation (minor or major).

Embodiment 42. The method of any one of embodiments 23 to 41, wherein the number or pattern of HPV integrations is correlated with at least one parameter of cytological classification including:

negative for intraepithelial lesion or malignancy,
abnormal squamous cells,
typical squamous cells (ASC),
of undetermined significance (ASC-US),
cannot exclude high-grade squamous intraepithelial lesion (ASC-H),
low-Grade Squamous Intraepithelial Lesion (LSIL),
high-Grade Squamous Intraepithelial Lesion (HSIL),
squamous cell carcinoma,
abnormal glandular cells,
atypical glandular cells (AGC): endocervical (not otherwise specified (NOS) or commented),
endometrial or not otherwise specified
atypical glandular cells, favor neoplastic: endocervical or not otherwise specified
endocervical adenocarcinoma in situ (AIS), and/or
adenocarcinoma.

Other embodiments of the invention include extension of the methods of Embodiment 1 and the other embodiments described herein to other viruses that can produce free viral episomal DNA or which infect cells through both free episomal and genome-integrated forms. Such viruses include Herpes viruses (e.g. Epstein-Barr virus [EBV], Human Herpesvirus 6 [HHV-6]) and Human Immunodeficiency virus (HIV); see Morissette and Flamand, 2010, J. Virol.; Hamid et al., 2017, AIDS Res. Ther). Thus, the methods disclosed here may be generally used for detection and the quantification of integrated viruses and detection reservoir

cells harbouring such integrated viral genomes during clinical follow-ups or horizontal diagnoses. Moreover, the invention is applicable for diagnostic and/or prognostic of all diseases due to integrated virus (HPV or others). This includes particularly in HPV infections the cancers of head and neck, vaginal, vulval and anal cancers as well as other types of cancer, neoplasms, and proliferative diseases. In some embodiments, the methods described herein may be adapted to detect episomal or integrated DNA from genetic medical procedures.

Some aspects of the methods described herein concern the identification of biomarkers of the severity of an HPV infection in subjects or patients and the transfer of these results into an innovative kit for HPV diagnostic and/or prognostic use. The biomarker is characterized by high number of integrations of one or more High Risk (HR) HPV DNA in genome of cervical cells of patients or subjects comprising multiple complete genomes or fragment thereof containing at least 10% of the HPV genome corresponding to the size for example of the region E6 E7 DNA. It also relates to methods and tools for the detection of integration in the genome of subjects or patients, of HPV classified as HR such as HPV 16, 18, 31, 33, 45 35, 39, 51, 52, 56, 58, 59, 66 and 68. In addition, it concerns a method of assessing the risk of having or developing a cervical cancer comprising detecting or quantifying a number of integrations of HPV DNA into, or an integration pattern of HPV DNA in, genomic host DNA obtained from a patient or subject, thereby assessing the risk of having or developing cervical cancer. Specific, but not limited, embodiments include the following:

A method for assessing a risk of having or developing a cervical cancer including detecting or quantifying a number of integrations of HPV DNA into, or an integration pattern of HPV DNA in, genomic host DNA obtained from a patient or subject, thereby assessing the risk of having or developing cervical cancer. In some embodiments of this method a greater number of instances or a greater amount of, integrated HPV DNA is indicative of a high risk of having or developing cancer or is indicative of a more aggressive or higher grade cancer compared to a

patient or subject having fewer instances or lesser amounts of integrated HPV DNA. In other embodiments of this method a different pattern of HPV DNA integrations into genomic host DNA, compared to those in a control subject or patient, is indicative of a high risk of having or developing cancer or is indicative of a more aggressive or of a higher grade cancer. Such risks include the risk of having, developing, or relapsing with, cervical cancer. HPV DNA from different sources may be assessed using the methods disclosed herein, including those from pathogenic strains of HPV and from those closely associated with cancer. HPV strains include strains 16, 18, 21, 31, 33, 45, 52 and 58. In other embodiments, the method may be performed with HPV DNA from a lower risk strain, such as HJPV 6 or 11 or other strains merely associated with genital warts or mild cervical abnormalities.

The methods disclosed herein may also include detecting or quantifying a number of integrations of HPV DNA by comparison to a patient prior to HPV infection or clinical signs of HPV, a subject not infected with HPV, a subject not infected with a pathogenic strain of HPV, a subject having no lesions or other symptoms of HPV infection, or a subject having substantially no antibody titer or cellular immunity to HPV or to a particular HPV strain. For example, the methods disclosed herein may include detecting or quantifying a number of integrations of HPV DNA in a patient compared to the number of integrations in earlier biological sample obtained from the same patient or subject or with regard to a HPV-negative or HPV-positive control subject.

Detecting or quantifying a number of integrations in the methods disclosed herein may be performed using molecular combing of the genomic host DNA using probes that bind to HPV DNA sequences, for example, it may be performed using molecular combing of the genomic host DNA using probes to HPV 16, 18, 31, 33, 45, 35, 39, 51, 52, 56, 58, 59, 66 and/or 68. In some embodiments, the detecting or quantifying a number of integrations is performed using molecular combing of the genomic host DNA using probes that bind to or cover HPV

DNA L1 and L2, E1 and E2, and/or E6 and E7 sequences, wherein said probes may be labelled with the same or different colored fluorescent tags.

Patients assessed by the methods disclosed herein may have a cervical dysplasia or a positive PAP test. Some patients may have been previously infected with human immunodeficiency virus (HIV), may be immunosuppressed, may have been exposed to diethylstilbestrol before birth, or may have been treated for a precancerous cervical lesion or cervical cancer. In other embodiments, the methods described herein may be used to assess a risk of having anal, vaginal, vulvar, penile or oropharyngeal cancer.

In one embodiments of the method of the invention a number or pattern of HPV integrations is quantified by, or correlated, with at least one of the following: the number of HPV integration sites in host genomic DNA or the average number of such integrations; the size in kb of HPV DNA integrations into host genomic DNA; the number of HPV genomes integrated at each integration site; the presence or absence of integrated HPV DNA; the number of HPV integration sites per cellular genome; the average number of HPV integration sites in host cells; the mean number of HPV genomes integrated per integration site (or the mean size of integration sites); maximum number of HPV genomes integrated per integration site (or the maximum size of integration sites); minimum number of HPV genomes integrated per integration site (or minimum size of integration sites), or number of HPV genomes integrated per cellular genome.

In other embodiments of the methods disclosed herein, the number or pattern of HPV integrations is correlated with at least one parameter of lesion status including: normal histology (including all abnormalities without intraepithelial lesions or signs of viral infection such as metaplasia, cervicitis, decidual lesions or adenosis); low grade (LG) lesion, corresponding to former CIN1; high grade (HG) lesion, corresponding to former CIN2, 3 and CIS (carcinoma in situ) or AIS (adenocarcinoma *in situ*); normal cervix; Grade 1 atypical

transformation (AT); Grade 2 atypical transformation (AT); TAG2 a if there are no major signs; TAG2 b if there are major signs; TAG2 c when the appearance is suggestive of invasive cancer; and/or atypical transformation (minor or major).

In other embodiments of the methods disclosed herein, the number or pattern of HPV integrations is correlated with at least one parameter of cytological classification selected from negative for intraepithelial lesion or malignancy; presence or absence of abnormal squamous cells; presence or absence of typical squamous cells (ASC) of undetermined significance (ASC-US); an inability to exclude high-grade squamous intraepithelial lesion (ASC-H); presence or absence of low-grade Squamous Intraepithelial Lesion (LSIL); presence or absence of high-grade Squamous Intraepithelial Lesion (HSIL); presence or absence of squamous cell carcinoma; presence or absence of abnormal glandular cells; atypical glandular cells (AGC): endocervical not otherwise specified or commented); endometrial or not otherwise specified; atypical glandular cells, favor neoplastic: endocervical or not otherwise specified, presence or absence of endocervical adenocarcinoma *in situ* (AIS) and/or adenocarcinoma that is not endocervical, endometrial, extrauterine or not otherwise specified.

In some embodiments of the method of the invention, the number or pattern of HPV integrations is correlated with clinical outcome of cervical lesions in patients including progression, stability or regression. In others, the number or pattern of HPV integrations is correlated with viral clearance, HPV vaccination status, amelioration of symptoms of HPV infection or cure or evaluation of the performance of pharmaceutical treatment or personalized treatment.

Another aspect of the invention is a visualized DNA pattern obtained after hybridization of HR-HPV DNA labelled probes with a genomic DNA of a subject or a patient suspected to contain one or multiple integrated HR-HPV genomes or specific fragments thereof said DNA pattern comprising HR-HPV DNA and genomic DNA from normal or cancerous cells. This

DNA pattern may constitute HR-HPV DNA which is chosen among the genotypes 16, 18, 31, 33, 45, 35, 39, 51, 52, 56, 58, 59, 66 or 68. It may also constitute HR-HPV DNA that contains all or part of the E6 E7 DNA regions. A DNA pattern may function as a biomarker of HPV infection and integration.

EXAMPLE 1

Association between integration of high-risk HPV genomes detected by molecular combing and the severity and/or clinical outcome of cervical lesions

A clinical study was set-up to study the association between the integration of high-risk HPV genomes detected by molecular combing and the severity of cervical lesions in patients with an indication for colposcopy after an abnormal Pap smear. The cross-sectional part analysis was done on 410 patients HPV-HR. The parameters of integration determined in the analysis and the parameters of the lesion describing the lesion status are described below.

Integration Pattern. The integration pattern may be defined by the following variables with values that are directly determined from the above data: (i) presence or absence of integration or (ii) a number of HPV integration sites per cellular genome (iii) number of E6/E7 sequences integrated per cellular genome.

Lesion status. The results of biopsies under colposcopy and conisation specimens are given according to the new WHO classification published in 2014 include (i) normal histology (including all abnormalities without intraepithelial lesions or signs of viral infection such as metaplasia, cervicitis, decidual lesions or adenosis.), (ii) Low grade (LG) lesion, corresponding to former CIN1, (iii) High grade (HG) lesion, corresponding to former CIN2, 3 and CIS (carcinoma *in situ*), and (iv) AIS (adenocarcinoma *in situ*).

If there is a discrepancy between the histological data from biopsies performed under cervical colposcopy and the histology data from conisation specimens, the worst-case

histological result will be taken into account. The group "Colposcopy +, no histology": describes those where neither biopsy, nor conisation have been performed.

Results of the cross-sectional part analysis. The first data show that the % of patients with HPV integration and the number of integrations per genome is inferior in the normal group (abnormal cytology but normal biopsy) compared to the other groups. For example, according to the molecular combing technology, the percentage of patients with the DNA HR-HPV integration in the high grade group is 97,2% whereas in the normal group the percentage of subjects with the DNA HR-HPV is 82,5%

As shown by FIG. 1, the proportion test, a chi-square test with a Yates's correction for continuity is used ; see Yates, F (1934). "Contingency table involving small numbers and the χ^2 test". Supplement to the Journal of the Royal Statistical Society 1(2): 217–235. JSTOR 2983604, incorporated by reference.

As shown by FIG. 2, a Mann–Whitney–Wilcoxon test was used to compare the variable in two independent samples that were selected from populations having the same distribution; see David F. Bauer, J. Am. Statistical Assoc. 67(339): 687-690 (1972, incorporated by reference. Confidence sets are constructed using rank statistics.

Contribution of the Molecular Combing (MC) Technique to the study of HR-HPV integration. The objective of the scientific program is to study the rare and complex genetic event that is viral integration, using a sensitive, unbiased and high-resolution technique. None of the current methods mentioned above fully meets these criteria. FIG. 3 describes analysis and workflow of samples by Molecular Combing. The first step consists in extracting very high molecular weight DNA from a biological specimen (blood, smear, etc.). Once extracted, the DNA molecules are "combed", i.e., attached by their ends to a silane-coated glass slide and uniformly stretched by a receding air-water interface (Bensimon 1994). Images of combed DNA fibers are shown by FIG. 4.

Once they are irreversibly fixed in this configuration to the glass substrate, DNA molecules are hybridised with a set of fluorescent probes specific for the DNA sequences of interest in order to obtain the specific fluorescent signature of this DNA region. During this study, the probes used will be specific for HPV 16, 18, 31, 33, 45, 35, 39, 51, 52, 56, 58, 59, 66 and 68 or other HPV genomes, and a reference region in order to normalize the result with respect to the number of cellular genomes combed on the coverslips. Only HPV forms integrated in the cellular genome will be combed and analysed. Circular episomal forms are not analysed as they have no free DNA terminal to be combed.

After hybridisation, the slide is placed in a scanner in order to acquire images by epifluorescence microscopy corresponding to all the fields of view of the slide. Using specific software developed by Genomic Vision and commercialized for other applications under the trademark Fibervision 2016, it is possible to identify, among the thousands of fields of view, the regions of interest on the coverslip containing a fluorescent signal and to measure the size of these different signals. This last step is made possible by the presence of a constant stretching factor (2 Kb/ μ m) which guarantees the determination of the physical distances within the region studied, by direct measurement of the probes and their spacing.

This approach can be used to study viral integration (i) with a high resolution, of approximately one kilobase (kb), including tandem integrations, (ii) directly without initial amplification of the genetic material and therefore with no selection bias associated with a choice of primer (iii) independently of viral transcription, and (iv) in a quantifiable and objective manner. None of the current and conventional analytic techniques mentioned above have all of these characteristics allowing complete, reliable and fine analysis of the integration of HR-HPV genomes in the cell genome.

Based on the results from the molecular combing study, a diagnostic and prognostic kit can be developed using that technology. However, other tests can also be derived from the

results of the clinical study that will be faster, high throughput and easily implementable in laboratories. Those tests could be also developed for any other virus that infects cells through both free episomal and genome-integrated forms. The invention includes viruses such as Herpesviruses (e.g. Epstein-Barr virus [EBV], Human Herpesvirus 6 [HHV-6]) and Human Immunodeficiency virus (HIV)(Morissette and Flamand, 2010, J. Virol.; Hamid et al., 2017, AIDS Res. Ther).

As shown herein a method according to the invention initiates a new generation of diagnostic and prognostic compositions, kits and diagnostic and prognostic methods. These methods which typically involve qPCR to measure a quantity of HPV (or other viral) DNA copies in a sample but provide a superior result by reducing error introduced by the presence of episomal viral DNA or nucleic acids which are not integrated into genomic DNA of a host. The inventors describe several method for reducing or eliminating episomal DNA contamination in a chromosomal or genomic DNA sample.

EXAMPLE 1

Sample preparation: Pap-smear samples were kept in Thinprep® (-20°C) or Surepath® (4°C) liquid cell preservation media. Cell collection was carried out at room temperature. In the case of vials with brush inside, 1 ml of the medium was pipetted and dispensed on the brush in order to get as many cells as possible. The operation was repeated 5 times. The cell suspension was then transferred in a 15 ml tube and centrifuged at 6000g for 10 min at 20°C. The cell pellet was washed twice with 1 ml of 1xPBS in order to remove all the traces of methanol or isopropanol present in the preservative solutions of the Thinprep® or Surepath® samples. The supernatant was removed as much as possible and the number of cells estimated by comparing the volume of the pellet to those of calibrated pellets.

Depletion of episomal HPV:

(I) Permeabilization of cervical cells membranes, removal of episomal genomes and DNA extraction: Cells were permeabilized in 1 ml of 10 mM PIPES (pH 6.8) , 300 mM sucrose, 100 mM NaCl, 3 mM MgCl₂, 1 mM EGTA, 0.5-1% Triton-X100 at 4°C for 20 min. The extraction of the episomal HPV genomes was carried out with 1 ml of a cold 50-500 mM (NH₄)₂ SO₄ solution. After removal of the supernatant, the cell pellet was washed three times with a cold solution of 1XPBS.

DNA was extracted according to the Miller high-salt protocol described in Kurvinen et al. (Miller et al., 1988, Nucleic Acids Res.; Kurvinen et al., 2000, Eur. J. Cancer). Cells were lysed 1 ml of 10 mM Tris (pH 8.3)-400 mM NaCl-1% sodium dodecyl sulfate-2 mM EDTA-proteinase K at 300 µg/ml overnight at 37°C. The protein precipitation was carried out by adding 300 µl of saturated NaCl. After centrifugation, the supernatant was removed and DNA was precipitated with ice-cold absolute ethanol. The DNA pellet was dissolved in sterile water.

(II) -Inclusion of cells into agarose plugs and removal of episomal HPV: The embedding of cervical cancers cells into agarose plugs was essentially performed as described in the protocol of the FiberPrep® extraction kit (Genomic Vision's kit, Paris France)). For each agarose plug the cell pellet was resuspended in 45 µl of 1x PBS. Homogenization of the cells was carried out by pipetting 10 times up and down with a 200 µl tip. The cell suspension was hold at 50°C for 10 seconds on a water-bath with a microtube supporter. An equal volume (45 µl) of low melting agarose (2 to 1.2% w/v) melted at 68°C for 10 min and kept at 50°C was then added to the cell suspension. After homogenization by pipetting several times, the 90 µl of the cell suspension in melted agarose were quickly dispensed into the well of a DNA plug mold. The latter was set at 4°C for 30 min for correct jellification of the agarose plug. The final concentration of the agarose plug -between 1 and 0.6%- and the quality of the low melting agarose were chosen for both optimum trapping of the genomic DNA and elimination of the episomal HPV. For cells lysis, each plug was incubated 16-18h at 50°C in the ESP buffer (0.5

M EDTA pH 8.0, 10% (v/w) sarcosyl, 15U Proteinase K). After extensive washing in TE 10:1 buffer, the plug was melted at 68°C and the agarose digested for 16-18h at 42°C with 1.5 U of β -agarase. This DNA solution was used for real-time PCR or ddPCR experiments.

(III) Depletion of episomal HPV with E2 biotinylated oligonucleotides: Biotinylated oligonucleotides covering the coding and the non-coding sequences of the E2 genes were designed. 500 ng of DNA were used for each patient for subsequent episomal HPV depletion. Conditions for a specific thermal denaturation of the episomal HPV DNA while preserving double-stranded genomic DNA were used. After incubation of the partially denatured DNA with the E2 biotinylated oligonucleotides, E2 oligos-episomal HPV complexes were captured on streptavidin beads. Unbound DNA was used for real-time PCR and ddPCR experiments.

(IV) Depletion of episomal HPV with commercial plasmids purification kit: The experiments were carried out essentially as described by Moller et al. (Moller et al., 2014, Proc. Natl. Acad. Sci USA; Moller et al., 2016, Journal of Vizualized Experiments). Briefly, after cells lysis the solution of DNA (genomic + episomal) was loaded on the resin of a plasmid purification column. Following binding of the episomal HPV on the resin, the eluate was used for real-time PCR or ddPCR experiments.

Absolute quantification of integrated HPV by real-time PCR: A real-time PCR (TaqMan) method was used for the measurement of the absolute values of the E2 and E6 ORFs. Below are described the primers and probes used for the quantification of integrated HPV16, HPV18 and HPV58. Similar primers and probes could be designed for HPV31, 35, 39, 45, 51, 52, 56, 59, 66 and 68. The primers and probes were designed for specific amplification of the E2 hinge regions which are known to be disrupted most frequently during the process of viral integration (Peitsaro et al., 2002, J. Clin. Microbiology). The integrated copy numbers were calculated by subtraction of the number of E2 (episomal) copies from the total number of E6 (episomal and integrated copies). The absolute number of viral integrated copies per cell was

obtained by normalization with a single-copy human gene. This method can be used directly on cervical cell samples or after cleaning or removal of episomal form.

Real-time PCR. The Real-time PCR experiments were carried out with the ABI Prism 7700 Sequence Detection System and the taqMan Universal PCR Master Mix (PE Applied Biosystems, Perkin Elmer). The amplification conditions were 2 min at 50°C, 10 min at 95°C, and a two-step cycle of 95°C for 15 s and 60°C for 60 s for a total of 40 cycles. The primers and probes are presented in table 1. The sizes of the E2 amplimers were 83bp, 136bp, 90bp and 111bp for HPV16, HPV18, HPV33 and HPV58, respectively. The sizes of the E6 amplimers were 126bp, 144bp, 124bp and 101bp, for HPV16, HPV18, HPV33 and HPV58, respectively. The E2 probes were labeled with a VIC fluorescent dye at the 5' and the MGB/NFQ Quencher at the 3' end. The E6 probes were labeled with a FAM fluorescent dye at the 5' and a TAMRA Quencher at the 3' end. The TaqMan Copy Number Reference Assay RNase P was used as the standard reference assay for copy number analysis. The 87 bp amplicon maps within the single exon RPPH1 gene. Fifty nanograms of target DNA from biopsies were added to the reaction mixture. Three standard curves were obtained by amplification of dilutions series of 50 million to 500 copies of clones of HPV16, HPV18, HPV33 and HPV58.

The results were recorded as copy numbers in 50 ng of cellular. The integrated E6 was calculated by subtracting the copy numbers of E2 (episomal) from the total copy numbers of E6 (episomal and integrated). The number of integrated E6 per genome was calculated by dividing the numbers of integrated E6 by the copy numbers of the reference RNase P gene [copy number (E6 episomal + integrated)-copy number E2 (episomal) X2]/copy number (RNase P gene). Ratios of E2 to E6 of less than 1 indicate the presence of both integrated and episomal forms. The ratio of E2 to integrated E6 represents the amount of the episomal form in relation to the integrated form.

Table 1

Name	Forward primer	Reverse primer	Probe
HPV 16-E2	5' AAC GAA GTA TCC TCT CCT GAA ATT ATT AG 3' (SEQ ID No.2) ¹	5' CCA AGG CGA CGG CTT TG 3' (SEQ ID No.3)	VIC-5' CAC CCC GCC GCG ACC CAT A 3' – MGB/NFQ (SEQ ID No.4)
HPV 16-E6	5' ACC GGT CGA TGT ATG TCT TGT TG 3' (SEQ ID No.5)	5' GAT CAG TTG TCT CTG GTT GCA AAT C 3' (SEQ ID No.6)	FAM-5' TGC ATG GAG ATA CAC CTA CAT TGC ATG AAT ATA 3' – TAMRA (SEQ ID No.8) ²
HPV 18-E2	5' GGT GGT GCC AGC CTA TAA CAT T 3' (SEQ ID No.9)	5' CCA TAG TTC CTC GCA TGT GTC TT 3' (SEQ ID No.10)	VIC-5' AAA AGT AAA GCA CAT AAA GCT ATT GAA CTG CAA ATG GC 3' – MGB/NFQ (SEQ ID No.12) ³
HPV 18-E6	5' AAT ACT ATG GCG CGC TTT GAG 3' (SEQ ID No.13)	5' TTC AAA TAC CTC TGT AAG TTC CAA TAC TG 3' (SEQ ID No.15) ⁴	FAM-5' TAC AAG CTA CCT GAT CTG TGC ACG GAA CTG 3' – TAMRA (SEQ ID No.17) ⁵
HPV 33-E2	5' GAT AAC CGA CCA CCA CAA GC 3' (SEQ ID No.18)	5' TGC ACA GAA CAG CTT TGT AAG G 3' (SEQ ID No.19)	VIC-5' ACC AC4A GAC ACC GCC CAG CC 3' – MGB/NFQ (SEQ ID No.20)
HPV 33-E6	5' TGT CAA AGA CCT TTG TGT CCT 3' (SEQ ID No.21)	5' TTT CTC TAC GTC GGG ACC TC 3' (SEQ ID No.22)	FAM-5' CAG CGC CCT GCC CAA CGA CC 3' – TAMRA (SEQ ID No.23)
HPV 58-E2	5' GAG GCC ACC AAC AAC GAA AG 3' (SEQ ID No.24)	5' GTC CAC GGC GCA GTC TGT ATA 3' (SEQ ID No.25)	VIC-5' AAG CGA CGA CGA CTC GAT TTA CCA GAC TC 3' – MGB/NFQ (SEQ ID No.27) ⁶
HPV 58-E6	5' TGA CAG CTC AGA CGA GGA TGA A 3' (SEQ ID No.28)	5' CAC AAG TGT AAC AAC AAG TTA CAA TGT AGT 3' (SEQ ID No.30) ⁷	FAM-5' ACA AGA ACA ACC GGC CAC AGC TAA TT 3' – TAMRA (SEQ ID No.32) ⁸

1: Or its short version as SEQ ID No.1 (AAC GAA GTA TCC TCT CCT GAA ATT ATT)

2: Or its short version as SEQ ID No.7 (TGC ATG GAG ATA CAC CTA CAT TGC)

3: Or its short version as SEQ ID No. 11 (AAA AGT AAA GCA CAT AAA GCT ATT)

4: Or its short version as SEQ ID No.14 (TTC AAA TAC CTC TGT AAG TTC CAA TAC)

5: Or its short version as SEQ ID No.16 (TAC AAG CTA CCT GAT CTG TGC ACG)

6: Or its short version as SEQ ID No.26 (AAG CGA CGA CGA CTC GAT TTA CCA)

7: Or its short version as SEQ ID No.29 (CAC AAG TGT AAC AAC AAG TTA CAA TGT)

8: Or its short version as SEQ ID No.31 (ACA AGA ACA ACC GGC CAC AGC TAA)

Digital droplet PCR (ddPCR) quantification of integrated HPV ((to be used directly on cervix cells samples or after cleaning of episomal form): 100 ng of sample DNA were cleaved using restriction enzymes. The selection of these restriction enzymes was done using ddPCR Calculations Tools (BIO-RAD). The Mastermix for ddPCR included 1X ddPCR Supermix for Probes (no dUTP, BIO-RAD), 0.9 μ M primer and 0.25 μ M probe (Applied Biosystems) with 5 μ l of the cleaved sample DNA. Experiments were carried out as described in Lillsund Larson and Helenius (Lillsunde Larson and Helenius, 2017, Cell. Oncol.). Primers and probes were designed for HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68.

Q-PCR using primers anchored on HPV sequence and on host genome sequences:

During the process of integration in the host genome the HPV sequences are susceptible to disrupt endogenous genes. In the present invention we show that nucleic acids sequences covering the junctions between the disrupted host genes and all or part of the integrated HPV genomes constitute biomarkers that correlate with progression to neoplasia in HPV induced

cervical cancer. The nucleic acids sequences ranging from 1 base to 2.5 kb from either sides of the integration sites and so localized on both the host and the HPV genomes are considered as biomarkers. These biomarkers can be used to diagnose or assist in the diagnosis of HPV-induced cancer. They can also be used to increase the positive predictive value of current screening modalities. Exemplary biomarkers include, MAPK10, PTPN13, NUDT15, MED4, ITM2B, RB1, LPAR6, RAB11A, RPL13A, ZNF341, OFD1, DHRS3, TBC1D22B, AFF3, CXCL6, PF4V1, IMMP2L, MMP12, WDR20, ALDHA1A, TPRG1, TUBD1, MAST4, LOC100132167, NFIX, CCAT1, GPR137B, RAB22A, C9orf3, MACROD2, DACH1, ATP10A, SPG11, SORD, COL4A4, GATSL1, GATSL2, MAP2, EPN1, ATXN3L, EGFL6, MAGI2 (Holmes et al., 2016, Genomic Medecine). Forward and reverse oligonucleotide primers chosen both in the disrupted gene and the integrated HPV genome i.e that specifically amplify the integration junctions, are used for the biomarkers production. Alternatively, specific nucleic acids sequences complementary to the integration junctions are used for the detection by hybridization of the biomarkers mentioned above. An example of nucleic acids sequences involved during the integration of the HPV16 viral genome in the OFD1 human gene (ref seq NG_008872.1) is presented in Figure 10. Table 2 presents the sequences of the forward and reverse primers used to amplify the OFD1 gene/HPV16 and the HPV16/OFD1 gene junctions used as biomarkers. This method is used directly on cervix cells samples or after cleaning of episomal forms). Table 2:

Junction	Forward primer	Reverse primer	Amplicon size (bp)
Junction1 OFD1-HPV16	5'-CCAGATACCAATGTGTGGC-3' ID No.33	5'-GTGCCAAAAAGCATGCAACC-3' ID No.34	410
Junction2 HPV16-OFD1	5'-TTACTGCACAGGAAGCAAAACA-3' ID No.35	5'-ATGTCTTCCGGGGAACTGG-3' ID No.36	476

Multiplex Ligation-dependent Probe Amplification: An MLPA-assay was used to monitor HPV16/18 viral load and integration. A similar system is used to measure HPV integration level for the 14 high-risk HPV virus in cervix cells of patients. This method uses primers designed specifically on HPV viruses. Because the E6 and E7 genes are nearly always present in all HPV-related lesions, regardless of physical status, probes against these genes were designed for typing of the 14 HPV. For the detection of the integrated status, two E2 probe sets were developed for each HPV, which target the sequence most frequently deleted on integration into the human genome. After a first step of preamplification, MLPA reaction is performed and PCR products are analyzed by electrophoresis. Viral load is determined by determining the ratio between E6/E7 and 7 human targets (beta globin, MSH2 for example). The E2/E6 ratio is also plotted against the viral load to determine the viral integration cut-off value. This method can be used directly on cervical cell samples or after cleaning or removal of episomal form.

EXAMPLE 2

Clinical Study Protocol

This protocol involves the association between Integration of High-Risk HPV Genomes Detected by Molecular Combing or by FISH or similar techniques and the Severity and/or Clinical Outcome of Cervical Lesions. The study is indicated for patients with an abnormal cervical uterine smear undergoing diagnostic colposcopy. It is an exploratory study and involves DNA combing. The study is performed in compliance with ICH GCP.

Background and Rationale: The occurrence of cervical cancer is associated with persistent infection by one or more high-oncogenic risk human papillomaviruses (HPV). The most common genotypes associated with cervical cancer are HPV genotypes 16, 18, 31, 33 and 45, which are responsible for more than 80% of these cancers.

Because of its slow progression, cervical cancer can be prevented by screening and the treatment of the precancerous lesions that precede it. This screening is currently based on the cytological examination of cervical (Pap) smears. In the case of abnormal findings, the colposcopy is indicated in order to take a biopsy sample to complete the diagnostic process.

The integration of the high-risk HPV genome in the cell genome is considered to be a key event in the development of cervical cancer and as one of its most important risk factors. In most cervical carcinomas, the HPV genome is integrated, whereas it is mainly in episomal form in low-grade lesions. Other data also suggest that the rapid progression of early cervical lesions to high-grade lesions is closely associated with the integration of HPV. Detecting the integration of high-risk HPV genomes in the cellular genome may therefore provide a useful marker for the identification of high-grade lesions or lesions at high risk of progression. This would make it possible to reduce the number of unnecessary coloscopies, avoid over-treatment of lesions that spontaneously regress and better target the lesions requiring treatment.

One objective of the inventors was to investigate the association between the integration of high-risk HPV genomes detected by molecular combing and the severity of cervical lesions in patients with an indication for colposcopy. Other objectives include investigating the association between integration of high-risk HPV genomes detected by molecular combing and viral clearance as well as the association between the integration of high-risk HPV genomes detected by molecular combing and the clinical outcome of cervical lesions. It also involves investigating the integration rate detected by molecular combing according to the type of high-risk HPV.

Methodology. The study is designed as multicenter cohort study with prospective inclusion. The patient population studied is all women all women aged 25 to 65, consulting a Department of Obstetrics and Gynaecology participating in the research for colposcopy indicated after an abnormal Pap test (ASC-US, ASC-H, atypical glandular cells, LSIL, HSIL).

Participation in the research is proposed to all women eligible to take part in this study during a gynecology visit for colposcopy indicated after an abnormal Pap test. Patients agreeing to participate in the research (signed informed consent form) will be included in the study.

Cross-sectional part of the study (baseline visit). A cervical smear is collected and HPV genotyping is performed on this sample. Patients with a negative result for high-risk HPV genotypes will leave the study. For patients with a positive HPV result for any high-risk HPV genotype(s), the study of the integration of the HPV genome by molecular combing is performed using this sample. Photographs of the cervix are taken with a video colposcope. Appropriate management (regular monitoring or treatment of lesions) will be proposed to the patient by the gynecologist using the colposcopy results and, if appropriate, the histological analysis of biopsies performed by colposcopy.

Longitudinal part of the study. Patients for whom follow-up is proposed will participate in the longitudinal part of the study. Follow-up visits at 6, 18 and 30 months are made and involve collection of a cervical smear sample for cytological analysis. Cervical smear samples are also collected at 12, 24 and 36 months for HPV genotyping and for patients with a positive HPV result for any high-risk HPV genotype(s), the study of the integration of the HPV genome by molecular combing will also be performed using the cervical smear sample. Colposcopy, cervical images are taken with a video colposcope. Biopsies are performed when they are considered necessary by the gynecologist. For patients whose care consists of treating the lesion, participation in the study is discontinued. For patients who have become negative for high-risk HPV and the lesion regressed (confirmed in 2 consecutive visits), the participation in the study is discontinued. Pregnancy is also a reason for discontinuation during the longitudinal part of the study.

Statistical Analysis. Test for factors associated with the severity of cervical lesions, including various integration parameters of high-risk HPV genomes by univariate (Student's

test, Wilcoxon, Chi2 or Fisher exact tests) and multivariate analysis (logistic regression). Test for factors associated with viral clearance, including various integration parameters of high-risk HPV genomes by univariate (Student's t test, Wilcoxon, Chi2 or Fisher exact tests) and multivariate analysis (logistic regression). Test for factors associated with the progression of cervical lesions (regression or progression), and in particular different parameters of integration of high-risk HPV genomes by univariate (Student's test, Wilcoxon, Chi2 or Fisher exact tests) and multivariate analysis (logistic regression). One interim analysis is done with accumulating cross-sectional data upon having about 100 subjects enrolled in the smaller group. The main goal of the interim analysis is a sample size re-estimation. There are no formal stopping rules. A possible inflation of type I error rate will not be adjusted for because of the exploratory design of the study.

Exclusion/Inclusion Criteria. Inclusion criteria include women 25 to 65 years of age visiting the site to undergo colposcopy in the context of an abnormal cervical uterine smear (ASC-US, ASC-H, glandular anomalies, LSIL, HSIL) performed at least one month and at most 6 months before agreeing to participate in the study as well as written consent. Exclusion criteria include those vaccinated against HPV treated for a cervical disorder with normal cytology follow-up for less than 2 years, with a known positive HIV test, with a chronic disease generating immunosuppression, with immunosuppression treatment in progress, with general corticoid treatment for 2 weeks or longer in the last 6 months, pregnant, and those with participation in a clinical trial with investigational drugs within the last 3 months before the enrolment or during the present trial period.

Evaluation. Participants are evaluated for Integration of HPV. This includes integration (presence/absence), number of HPV integration sites per cellular genome, mean number of HPV genomes integrated per integration site (or the mean size of integration sites), maximum number of HPV genomes integrated per integration site (or the maximum size of

integration sites), minimum number of HPV genomes integrated per integration site (or minimum size of integration sites), number of HPV genomes integrated per cellular genome, lesion status, viral clearance, clinical outcome, and cure.

The primary objective of this investigation is to study the association between the integration of high-risk HPV genomes detected by molecular combing and the severity of cervical lesions in patients with an indication for colposcopy after an abnormal Pap smear. Other objectives include assessment of patients who underwent colposcopy for an abnormal Pap test and for whom simple monitoring is indicated: the association between the integration of HR-HPV genomes detected by molecular combing and viral clearance; the association between the integration of HR-HPV genomes detected by molecular combing and the clinical outcome of the cervical lesions; and assessment of the integration rate detected by molecular combing for each type of HR-HPV.

Integration of HPV genomes studied by molecular combing. For each patient, the parameters assessed during the analysis by molecular combing are: the estimated number of combed cell genomes on the coverslip. This number of combed cellular genomes varies from one coverslip to another, depending on the amount of DNA extracted and the combing density. It is therefore used for the standardisation of results. The number of HPV integration sites on the coverslip. The size in kilobase (kb) of HPV genome integrations. The number of HPV genomes integrated at each integration site. The integration pattern is defined by the following variables with values that are directly determined from the above data: integration (presence/absence), number of HPV integration sites per cellular genome, mean number of HPV genomes integrated per integration site (or the mean size of integration sites), maximum number of HPV genomes integrated per integration site (or the maximum size of integration sites), minimum number of HPV genomes integrated per integration site (or minimum size of integration sites), number of HPV genomes integrated per cellular genome. These values will

be calculated: on all detected HPV signals, then only on HPV signals above 10 kb. This signal selection overcomes the problem of the potential contamination by viral episomal forms that may be linearized during handling, thereby allowing these forms to be combed by their free ends.

Lesion status. The results of biopsies under colposcopy and conization specimens are given, whenever possible, according to the new WHO classification published in 2014: normal histology (including all abnormalities without intraepithelial lesions or signs of viral infection such as metaplasia, cervicitis, decidual lesions or adenosis.), low grade (LG) lesion, corresponding to former CIN1, high grade (HG) lesion, corresponding to former CIN2, 3 and CIS (carcinoma in situ), AIS (adenocarcinoma in situ). Nevertheless, it is possible to continue using the WHO 2003 classification (normal cervix, condyloma, CIN1, CIN2, CIN3, AIS). If there is a discrepancy between the histological data from biopsies performed under cervical colposcopy and the histology data from conization specimens, the worst-case histological result will be taken into account.

French terminology: For this study, the French and international terminologies will be used for colposcopic classification: normal cervix: includes ectropions, metaplasia and Nabothian cysts. Grade 1 atypical transformation (AT): corresponds to a centripetal area of re-epithelialisation (like normal metaplasia) but with dystrophic epithelium not producing glycogen. This area will appear as: non-congestive on examination without preparation, slightly acetowhite with sharp borders not containing any crypt openings and iodine negative with sharp borders after application of Lugol's iodine solution. Grade 2 atypical transformation (AT) defined on examination without preparation by the presence of an area of congestion, then an intense acetowhite area, with blurred margins, with the presence of crypt openings and iodine negative appearance with blurred margins after application of Lugol's iodine solution. This lesion complex presents a centrifugal progression both towards the ectocervix and endocervix

showing its dysplastic nature. This grade 2 AT has several stages of increasing severity according to the number of signs of severity on the images (vascular erosions, ulcerations, vegetation, necrosis etc.). TAG2 a if there are no major signs, TAG2 b if there are major signs, and TAG2 c when the appearance is suggestive of invasive cancer. For each colposcopic appearance: normal, TAG1 or TAG2, the level of the squamocolumnar junction (SCJ) will be specified: visible or not visible. When the SCJ is not visible, colposcopy is considered non-contributory.

International terminology: This is mainly centred on the examination without preparation and after application of acetic acid as Lugol's iodine solution is not an integral and systematic part of colposcopic examination (especially for English-speaking practitioners). Normal cervix: including all aspects of normal cervix in the French terminology. Atypical Transformation: corresponds to the appearance of an acetowhite area in the transformation zone (i.e. between the original SCJ and the new SCJ: repair zone). This atypical transformation (AT) may be: minor: few acetowhite areas, without any sign of seriousness corresponding to mainly CIN1 lesions but also including grade 1 atypical transformation considered here as a minor appearance of acetowhitening major: more marked acetowhitening, with areas with additional signs that may correspond to mainly CIN2 + lesions. For each colposcopy appearance: the level of the squamocolumnar junction (SCJ) will be specified: visible or not visible. When the SCJ area is not visible, colposcopy is considered non-contributory.

Cytological classification. Cytological analysis of cervical smears is reported using the Bethesda system terminology (2001). This provides information about: the type of specimen: conventional cervical smear, monolayer, the quality of the specimen: satisfactory or not for analysis, general classification: presence/absence of abnormal epithelial cells.

Interpretation/Result: Negative for intraepithelial lesion or malignancy, Abnormal squamous cells, Atypical squamous cells (ASC), Of undetermined significance (ASC-US),

Cannot exclude high-grade squamous intraepithelial lesion (ASC-H), Low-Grade Squamous Intraepithelial Lesion (LSIL), High-Grade Squamous Intraepithelial Lesion (HSIL), Squamous cell carcinoma, Abnormal glandular cells, Atypical glandular cells (AGC): endocervical (not otherwise specified (NOS) or commented), endometrial or not otherwise specified, Atypical glandular cells, favor neoplastic: endocervical or not otherwise specified, Endocervical adenocarcinoma in situ (AIS). Adenocarcinoma: endocervical, endometrial, extrauterine or not otherwise specified.

Clinical outcome of cervical lesions in patients with simple follow-up. The clinical outcome of cervical lesions will be evaluated solely for patients participating in the longitudinal part of the study. The outcome of histological lesions discovered during an abnormal smear for which surveillance was decided will be assessed according to coloscopic, cytological and histological criteria. The three possibilities are progression, stability or regression. In case of a discrepancy, the worst-case endpoint will be taken into account.

Viral clearance. Viral clearance is evaluated solely for patients participating in the longitudinal part of the study. Viral clearance will be evaluated in 2 ways: HR-HPV clearance: All high-risk HPV detected at each visit will be considered globally. Thus, HR-HPV viral clearance will be defined by the absence of any HR-HPV during a follow-up visit and by specific type clearance: Each HR-HPV will be considered independently. Clearance of a specific viral type will be defined by a negative test during a monitoring visit for a HPV subtype present at baseline. The HR-HPV subtypes not present at baseline which could be detected during the follow-up will not be taken into account.

Cure. Cure is defined by complete regression of cervical lesions at 2 consecutive follow-up visits and clearance of HR-HPV.

Investigational Plan. FIG. 5 provides a study design diagram. The study is

open-label, single arm, multi-centre exploratory study with two parts: cross-sectional part (one study visit) and a longitudinal part (follow-up) – up to 6 visits during 36 months

The estimated number of patients was 993 enrolled /655 evaluable patients. The sample size was re-calculated based on an estimation of the accumulated data parameters upon having about 100 subjects enrolled in the smaller severity group (whichever it is).

The duration of a patient's study participation can vary according to patient care: if the decision is follow-up with regular visits: the patient's participation will last at least 6 months and no more than 3 years, according to the decision to treat or recovery, or if the decision is to treat: the patient's participation will be one-time (no follow-up) and will consist only of the initial colposcopy visit.

To meet the main objective, two groups of patients I evaluated: One group of patients with abnormal cervical uterine smears, positive HR-HPV genotyping and a normal histology result or a low-grade histological lesion. One group of patients with abnormal cervical uterine smears, positive HR-HPV genotyping and a high-grade histological lesion.

Interim Analysis. One interim analysis will be done with accumulating cross-sectional data upon having about 100 subjects enrolled in the smaller group (whichever it is). The main goal of the interim analysis is a sample size re-estimation, however all main variables will be analysed too. There is no independent data monitoring committee to assess the interim outcome. The interim statistical report is available to the Sponsor directly; no data assessment meeting is planned. There are no formal stopping rules. Nevertheless, based on the interim results the sponsor will decide whether to continue or discontinue, seek extra data, and/or make modification of the study design. Unless this happens, however, the principal investigators and central administrative staff will remain ignorant of the interim results of the accumulated data. For the statistical considerations of the interim analysis see the statistical section of this study protocol.

Cross-sectional part analysis. Patients enrolment was stopped after inclusion of 688 subjects. 410 of them, classified as HPV high risk were considered for the statistical analysis. ¶

The statistical consideration is described in the statistical section of this study protocol and the statistical report is available to the Sponsor directly.

Cervical uterine sample management at M0, M12, M24 and M36. The cervical uterine sample will be collected with the ThinPrep (Hologic) device. The total volume of the cell suspension obtained will be 20 mL. The samples will be transported to the accredited central laboratory for genotyping by special courier at -20°C. The service provider will take 4mL of the sample for HPV genotyping, the remaining volume of the sample (16mL) will be frozen at -20°C. According to the results of the HPV genotyping, the laboratory destroys the samples that are negative for HR-HPV (sending a certificate of destruction to Sponsor) stored the HR-HPV positive samples at -20°C, the samples will be transported in group to Genomic Vision at -20°C upon agreement. Genomic Vision performs the analysis by Molecular Combing on the HR-HPV positive samples. Genomic Vision stores the rest of the samples for which Molecular Combing analysis has been performed for a period of 10 years after the end of the study (biological collection). Sponsor is responsible for the destruction of these samples by requesting a certificate of destruction.

Cervical uterine sample management at M6, M18, M30. Cervical-uterine sample collection will be performed with the ThinPrep (Hologic) device and will be shipped as usual to the site's cytology laboratory, which will perform the cytological analysis of the sample.

Molecular combing to search for HPV integration. The search for HPV genome integration involves 5 steps:

Step No. 1: DNA Extraction. The cells from the biological sample to be analysed are placed in a block of agarose called a “plug,” in which several enzyme treatment stages lead to

sample lysis and the elimination of all proteins attached to the DNA (histones, transcription factors, etc.). The agarose is then digested to obtain naked DNA in solution.

Step No. 2: Combing. A glass slide coated with a fine layer of silane is dipped into the DNA solution. The double-stranded DNA is fixed irreversibly on the silanized slide at one and/or the other of the two ends by hydrophobic interactions. The glass slide is then removed from the solution vertically at a constant speed of 300 $\mu\text{m/s}$. The force stretches the DNA is applied exclusively at the meniscus and leads to a uniform stretching of the DNA molecules regardless of their length. The DNA strands are thus aligned parallel to each other and form a mat on the slide, which includes, according to its density, around 100 complete genomes.

Step No. 3: Hybridisation. The stretched DNA molecules are irreversibly fixed on the glass slide after combing; however, they remain accessible to additional DNA sequences of probes covering HPV genomes 16, 18, 31, 33, 45, 35, 39, 51, 52, 56, 58, 59, 66 and 68. These probes are obtained by random priming allowing the incorporation of modified nucleotides in the sequences that are detected with the help of specific antibodies of each modified nucleotide coupled with fluorochromes, thus generating a fluorescent signal.

The HPV genomes studied are covered by 3 probes (Figure 3): one specific to the viral genome region containing genes L1 and L2, the second specific to the region containing viral genes E1 and E2, and the third to the region containing viral oncogenes E6 and E7. The probes corresponding to regions L1L2 and E1E2 will be displayed in blue and the one corresponding to region E6E7 in cyan (green + blue). In addition, probes are generated that cover a reference locus of the host DNA, in order to quantify the number of host genomes combed and standardize HPV integrations by patient cell. These reference signals are displayed in red. See FIG. 6.

Step No. 4: Acquisition of fluorescent signals. Image acquisition is performed by epifluorescence microscopy. The slide is lit by wavelengths corresponding to the various

fluorochromes by means of filtered light (excitation spectrum) and the re-emitted fluorescent light is captured in the appropriate wavelengths (emission spectrum). In order to acquire the images, Genomic Vision has a scanner that is able to divide the slide into several thousand visual fields and capture the images for each visual field. The images thus generated are analysed by software that allows the detection of the regions of interest containing a fluorescent signal of interest and their measurement; see FIG. 7.

Step No. 5: Analysis. The HPV 16, 18, 31, 33, 45 35, 39, 51, 52, 56, 58, 59, 66 and 68 genomes are covered by probes: the probes covering L1L2 and E1E2 are displayed in blue and the one covering E6E7 in cyan (blue + green). The reference locus is displayed in red, see FIG. 8 which diagrams types of signals detected on the slides.

For each patient, the parameters which are noted during the Molecular Combing analysis are as follows:

Estimated number of host genomes combed on the slide: It is calculated by adding the sizes (in kb) of all the reference signals (red) visible on the analysed slide and dividing that sum by the theoretical size (in kb) of the locus, then by 2 (because there are 2 alleles per genome). In the case of smears that suggest a cancerous lesion, which have considerable genetic instability, we will start from the premise that this region has not been modified. Indeed, since the locus is especially poor in repeated sequences, there is a lower risk of being modified. This number of combed host genomes varies from one slide to another, according to the amount of DNA extracted and the density of the combing. It therefore serves to standardize the results.

Number of HPV integration sites on the slide. It corresponds to the number of HPV signals (blue) detected on the slide. In the example in the figure above, we can count 5.

Size in kilobases (kb) of the HPV genome integrations: This corresponds to the size of the HPV signals (blue) measured in kb.

Number of integrated HPV genomes at each integration site. This corresponds to the number of cyan signals visible for each blue HPV signal. For example, in the figure above, the framed HPV signal has 4 integrated HPV genomes at that site.

The integration profile will be defined by the following variables, whose values are deduced directly from the data stated above: The number of HPV integration sites per patient genome, the average number of integrated HPV genomes per integration site (or the average size of the integrations), the maximum number of integrated HPV genomes per integration site (or the maximum size of the integrations), the minimum number of integrated HPV genomes per integration site (or the minimum size of the integration sites), the number of integrated HPV genomes per patient genome.

These values are calculated for all the HPV signals (blue) detected and then only for the HPV signals (blue) greater than 10kb. This selection of signals allow the elimination of potential contamination by episomal viral forms that are linearized during handling, allowing these forms to be combed by their free ends.

Centralized reading of biopsies and cones. With regard to the histology samples taken (biopsy in the case of a visible lesion and cone samples if the lesion is treated), the histological analysis with a centralised reading of the histology samples is performed by an accredited laboratory. For this purpose, the samples (blocks of tissues or slides) is sent by the research site to this laboratory for assessment, which is also perform immunohistochemical staining tests for p16 and Ki67 transformation markers. The samples are archived by the laboratory for 10 years (blocks of tissues) and 5 years (slides).

Centralized reading of cervical photographs. The photographs of the cervix are taken at 3 successive times during the colposcopic examination: Exam without preparation with an optional photograph taken with a green filter to analyse the epithelium and the vessels; Exam after applying 3% acetic acid after waiting from 30 seconds to 1 minute to allow the

identification of the squamocolumnar junction and the search for acidophilus; Exam after applying Lugol's solution. A colposcopic diagram (optional) based on the findings after applying acetic acid allows visualization of the squamocolumnar junction, the extent of the acidophilus and, especially, the exact location and number of the biopsy or biopsies. A centralised reading of the photographs of the cervix are performed by 2 experienced gynaecologists who are independent of the study. The colposcopic situation is determined for each photograph.

A sample research plan is described below.

	Baseline Visit	Follow-up Visit				
M0		M6	M12	M18	M24	M30
M36						
Consent	X					
Urine pregnancy test	X	X	X	X	X	X
Collection of medical history and lifestyle info	X					
Cervical uterine smear	X	X	X	X	X	X
Cytological analysis		X	X			X
HPV genotyping	X		X		X	X
Molecular combing ¹	X	X	X	X	X	X
Colposcopy	X		X		X	X
Photograph of the cervix	X		X		X	X
Cervical biopsies + centralised histology analysis ²		X	X	X	X	X

¹ in high-risk HPV positive patients

² as recommended

Terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. For example, as used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items and may be abbreviated as “/”.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word “comprise”, and variations such as “comprises” and “comprising” means various components can be co-jointly employed in the methods and articles (e.g., compositions and apparatuses including device and methods). For example, the term “comprising” will be understood to imply the inclusion of any stated elements or steps but not the exclusion of any other elements or steps.

As used herein in the specification and claims, including as used in the examples and unless otherwise expressly specified, all numbers may be read as if prefaced by the word “substantially”, “about” or “approximately,” even if the term does not expressly appear. The phrase “about” or “approximately” may be used when describing magnitude and/or position to indicate that the value and/or position described is within a reasonable expected range of values and/or positions. For example, a numeric value may have a value that is +/- 0.1% of the stated value (or range of values), +/- 1% of the stated value (or range of values), +/- 2% of the stated value (or range of values), +/- 5% of the stated value (or range of values), +/- 10% of the stated

value (or range of values), +/- 15% of the stated value (or range of values), +/- 20% of the stated value (or range of values), *etc.* Any numerical range recited herein is intended to include all sub-ranges subsumed therein.

REFERENCES

- Choo, K. B., C. C. Pan and S. H. Han (1987). "Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames." *Virology* 161(1): 259-261.
- Cricca, M., A. M. Morselli-Labate, S. Venturoli, S. Ambretti, G. A. Gentilomi, G. Gallinella, S. Costa, M. Musiani and M. Zerbini (2007). "Viral DNA load, physical status and E2/E6 ratio as markers to grade HPV16 positive women for high-grade cervical lesions." *Gynecol Oncol* 106(3): 549-557.
- Crosbie, E. J., M. H. Einstein, S. Franceschi and H. C. Kitchener (2013). "Human papillomavirus and cervical cancer." *Lancet* 382(9895): 889-899.
- Dyson, N., P. M. Howley, K. Munger and E. Harlow (1989). "The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product." *Science* 243(4893): 934-937.
- Ferber, M. J., E. C. Thorland, A. A. Brink, A. K. Rapp, L. A. Phillips, R. McGovern, B. S. Gostout, T. H. Cheung, T. K. Chung, W. Y. Fu and D. I. Smith (2003). "Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma." *Oncogene* 22(46): 7233-7242.
- Fritz RB¹, Musich PR. "Unexpected loss of genomic DNA from agarose gel plugs". (1990) *Biotechniques*. Nov;9(5):542, 544, 546-50.
- Gradissimo Oliveira, A., C. Delgado, N. Verdasca and A. Pista (2013). "Prognostic value of human papillomavirus types 16 and 18 DNA physical status in cervical intraepithelial neoplasia." *Clin Microbiol Infect* 19(10): E447-450.
- Hamid F.S., Kim, J. and Shin, C-G. (2017). "Distribution and fate of HIV-1 unintegrated DNA species: a comprehensive update". *AIDS Research and Therapy* 14(1):9

Herrick, J., C. Conti, S. Teissier, F. Thierry, J. Couturier, X. Sastre-Garau, M. Favre, G. Orth and A. Bensimon (2005). "Genomic organization of amplified MYC genes suggests distinct mechanisms of amplification in tumorigenesis." *Cancer Res* 65(4): 1174-1179.

Holmes A, Lameiras S, Jeannot E, Marie Y, Castera L, Sastre-Garau X, Nicolas A. (2016). « Mechanistic signatures of HPV insertions in cervical carcinomas ». *NPJ Genom Med*. Mar 16;1:16004.

Hopman, A. H., F. Smedts, W. Dignef, M. Ummelen, G. Sonke, M. Mravunac, G. P. Vooijs, E. J. Speel and F. C. Ramaekers (2004). "Transition of high-grade cervical intraepithelial neoplasia to micro-invasive carcinoma is characterized by integration of HPV 16/18 and numerical chromosome abnormalities." *J Pathol* 202(1): 23-33.

Hu, Z., D. Zhu, W. Wang, W. Li, W. Jia, X. Zeng, W. Ding, L. Yu, X. Wang, L. Wang, H. Shen, C. Zhang, H. Liu, X. Liu, Y. Zhao, X. Fang, S. Li, W. Chen, T. Tang, A. Fu, Z. Wang, G. Chen, Q. Gao, S. Li, L. Xi, C. Wang, S. Liao, X. Ma, P. Wu, K. Li, S. Wang, J. Zhou, J. Wang, X. Xu, H. Wang and D. Ma (2015). "Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism." *Nat Genet* 47(2): 158- 163.

Huang, L. W., S. L. Chao and B. H. Lee (2008). "Integration of human papillomavirus type-16 and type-18 is a very early event in cervical carcinogenesis." *J Clin Pathol* 61(5): 627-631.

Jeon, S., B. L. Allen-Hoffmann and P. F. Lambert (1995). "Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells." *J Virol* 69(5): 2989-2997.

Jeon, S. and P. F. Lambert (1995). "Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis." *Proc Natl Acad Sci U S A* 92(5): 1654-1658. Klaes, R., S. M. Woerner, R. Ridder, N. Wentzensen, M. Duerst, A. Schneider, B. Lotz, P. Melsheimer and M. von Knebel Doeberitz (1999). "Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes." *Cancer Res* 59(24): 6132-6136.

Kulmala, S. M., S. M. Syrjanen, U. B. Gyllensten, I. P. Shabalova, N. Petrovichev, P. Tosi, K. J. Syrjanen and B. C. Johansson (2006). "Early integration of high copy HPV16

detectable in women with normal and low grade cervical cytology and histology." *J Clin Pathol* 59(5): 513-517.

Kurvinen K, Yliskoski M, Saarikoski S, Syrjänen K, Syrjänen S. (2000) « Variants of the long control region of human papillomavirus type 16 ». *Eur J Cancer*. Jul;36(11):1402-10.

Lebofsky, R. and A. Bensimon (2003). "Single DNA molecule analysis: applications of molecular combing." *Brief Funct Genomic Proteomic* 1(4): 385-396.

Lillsunde Larsson G. and Helenius G. (2017) „Digital droplet PCR (ddPCR) for the detection and quantification of HPV 16, 18, 33 and 45 » *Cell Oncol.* (2017) 40:521–527Luft, F., R. Klaes, M. Nees, M. Durst, V. Heilmann, P. Melsheimer and M. von Knebel Doeberitz (2001). "Detection of integrated papillomavirus sequences by ligationmediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells." *Int J Cancer* 92(1): 9-17.

Marongiu, L., A. Godi, J. V. Parry and S. Beddows (2014). "Human Papillomavirus 16, 18, 31 and 45 viral load, integration and methylation status stratified by cervical disease stage." *BMC Cancer* 14: 384.

Miller SA, Dykes DD, Polesky HF. (1988). « A simple salting out procedure for extracting DNA from human nucleated cells". *Nucleic Acids Res.* Feb 11;16(3):1215

Møller HD, Bojsen RK, Tachibana C, Parsons L, Botstein D, Regenberg B. (2016) « Genome-wide Purification of Extrachromosomal Circular DNA from Eukaryotic Cells ». *J Vis Exp.* Apr 4;(110)

Møller HD, Parsons L, Jørgensen TS, Botstein D, Regenberg B. (2015) « Extrachromosomal circular DNA is common in yeast ». *Proc Natl Acad Sci U S A.* Jun 16;112(24):E3114-22.

Morissette, G. and Flamand, L. (2010). « Herpesviruses and Chromosomal Integration ». *Journal of Virology* 84(23):12100-9

Peitsaro, P., B. Johansson and S. Syrjanen (2002). "Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique." *J Clin Microbiol* 40(3): 886-891.

Peter, M., C. Rosty, J. Couturier, F. Radvanyi, H. Teshima and X. Sastre-Garau (2006). "MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors." *Oncogene* 25(44): 5985-5993.

Pett, M. and N. Coleman (2007). "Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?" *J Pathol* 212(4): 356-367.

Raybould, R., A. Fiander, G. W. Wilkinson and S. Hibbitts (2014). "HPV integration detection in CaSki and SiHa using detection of integrated papillomavirus sequences and restriction-site PCR." *J Virol Methods* 206: 51-54.

Sabol I, Salakova M, Smahelova J, Pawlita M, Schmitt M, Gasperov NM, Grce M, Tachezy R. (2008). "Evaluation of different techniques for identification of human papillomavirus types of low prevalence." *J Clin Microbiol*. 46(5):1606-13.

Scheffner, M., B. A. Werness, J. M. Huibregtse, A. J. Levine and P. M. Howley (1990). "The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53." *Cell* 63(6): 1129-1136.

Theelen W, Speel EJ, Herfs M, Reijans M, Simons G, Meulemans EV, Baldewijns, MM, Ramaekers FC, Somja J, Delvenne P, Hopman AH. (2010) « Increase in viral load, viral integration, and gain of telomerase genes during uterine cervical carcinogenesis can be simultaneously assessed by the HPV 16/18 MLPA-assay ». *Am J Pathol*. Oct;177(4):2022-33

Van Tine, B. A., J. Knops, T. R. Broker, L. T. Chow and P. T. Moen, Jr. (2001). "In situ analysis of the transcriptional activity of integrated viral DNA using tyramideFISH." *Dev Biol (Basel)* 106: 381-385.

Vega-Pena, A., B. Illades-Aguiar, E. Flores-Alfaro, E. Lopez-Bayghen, M. A. LeyvaVazquez, E. Castaneda-Saucedo and C. Alarcon-Romero Ldel (2013). "Risk of progression of early cervical lesions is associated with integration and persistence of HPV-16 and expression of E6, Ki-67, and telomerase." *J Cytol* 30(4): 226-232.

Vernon SD, Unger ER, Miller DL, Lee DR, Reeves WC. (1997) « Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer ». *Int J Cancer*. Feb 20;74(1):50-6.

Vinokurova, S., N. Wentzensen, I. Kraus, R. Klaes, C. Driesch, P. Melsheimer, F. Kisseljov, M. Durst, A. Schneider and M. von Knebel Doeberitz (2008). "Type-dependent integration frequency of human papillomavirus genomes in cervical lesions." *Cancer Res* 68(1): 307-313.

von Knebel Doeberitz, M., T. Bauknecht, D. Bartsch and H. zur Hausen (1991). "Influence of chromosomal integration on glucocorticoid-regulated transcription of growth-

stimulating papillomavirus genes E6 and E7 in cervical carcinoma cells." *Proc Natl Acad Sci U S A* 88(4): 1411-1415.

Wentzensen, N., R. Ridder, R. Klaes, S. Vinokurova, U. Schaefer and M. Doeberitz (2002). "Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions." *Oncogene* 21(3): 419-426.

Wentzensen, N., S. Vinokurova and M. von Knebel Doeberitz (2004). "Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract." *Cancer Res* 64(11): 3878-3884.

Ziegert, C., N. Wentzensen, S. Vinokurova, F. Kisseljov, J. Einenkel, M. Hoeckel and M. von Knebel Doeberitz (2003). "A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques." *Oncogene* 22(25): 3977-3984.

Zubillaga-Guerrero, M. I., B. Illades-Aguiar, M. A. Leyva-Vazquez, E. Flores-Alfaro, E. Castaneda-Saucedo, J. F. Munoz-Valle and L. C. Alarcon-Romero (2013). "The integration of HR-HPV increases the expression of cyclins A and E in cytologies with and without low-grade lesions." *J Cytol* 30(1): 1-7. zur Hausen, H. (2002). "Papillomaviruses and cancer: from basic studies to clinical application." *Nat Rev Cancer* 2(5): 342-350.

All publications and patent applications mentioned in this specification are herein incorporated by reference in their entirety to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference, especially referenced is disclosure appearing in the same sentence, paragraph, page or section of the specification in which the incorporation by reference appears.

The citation of references herein does not constitute an admission that those references are prior art or have any relevance to the patentability of the technology disclosed herein. Any discussion of the content of references cited is intended merely to provide a general summary of assertions made by the authors of the references, and does not constitute an admission as to the accuracy of the content of such references.

CLAIMS

Claim 1. A method for detecting a level of integrated viral DNA or dormant DNA, such as proviral DNA, comprising:

removing episomal viral or vector nucleic acids from genomic DNA in a cell sample, and quantifying a number of integrations of viral DNA into the genomic DNA of the cells by a method of amplification of an integration region in the DNA sample; thereby detecting a level of integrated viral DNA in the genomic DNA from the cell sample; and

optionally, performing a method of amplification on episomal nucleic acids removed from the sample.

Claim 2. A method according to claim 1 wherein the amplification method is quantitative polymerase chain reaction (qPCR), fluorescent *in situ* hybridization (FISH) or molecular combing.

Claim 3. The method of claim 1 or 2, wherein said step of removing comprises: permeabilizing cell membranes in the cell sample by exposing the cells to an extracting salt solution for a time and under conditions sufficient to make the cellular membranes permeable to episomal viral or vector nucleic acids from the cells and for the episomal viral or vector nucleic acids to leak out of the cells into a medium, separating the permeabilized cells from episomal viral or vector nucleic acids in the medium, and

performing qPCR on the separated permeabilized cells; and optionally, isolating and performing qPCR on the episomal viral or vector nucleic acids in the medium.

Claim 4. The method of any one of claims 1 to 3, wherein said removing comprises:

embedding cells into an agarose plug having an agarose concentration ranging from about 0.5 wt% to 1.5 wt%,

infusing the plugs containing the embedded cells with a proteolytic enzyme for a time and under conditions sufficient to substantially digest cellular proteins,

washing the plugs for a time and under conditions sufficient to remove episomal nucleic acids,

extracting cellular genomic DNA caught in the plugs and

performing qPCR on the genomic DNA extracted from the washed agarose plugs.

performing qPCR on the separated permeabilized cells; and

optionally, isolating and performing qPCR on the nucleic acids washed out of the agarose plugs.

Claim 5. The method of any one of claims 1 to 4, comprising partially depleting episomal HPV nucleic acids by biotinylating episomal nucleic acids and binding them (strept)avidin, thus removing them from a mixture of genomic and episomal nucleic acids.

Claim 6. The method of any one of claims 1 to 5, comprising:

isolating nucleic acids from the cell sample using a plasmid purification column,

recovering genomic DNA from material eluting from the plasmid purification column,

and

performing qPCR on the recovered genomic DNA; and, optionally,

isolating and performing qPCR on the material bound to the plasmid purification column.

Claim 7. The method of any one of claims 1 to 6, wherein the cell sample contains viral DNA.

Claim 8. The method of any one of claims 1 to 7, wherein the cell sample contains HIV nucleic acids.

Claim 9. The method of any one of claims 1 to 7, wherein the cell sample contains human papilloma virus (HPV) nucleic acids.

Claim 10. The method of claim 9, further comprising detecting by PCR a ratio of DNA encoding HPV E2 ORF to DNA encoding HPV E6 ORF, especially by qPCR ; wherein the ratio of DNA encoding E2 ORF to DNA encoding E6 ORF represents an amount of an episomal form in relation to an integrated form.

Claim 11. The method of claim 10, wherein the ratio of E2 ORF to E6 ORF is determined by real-time PCR using a set of primers and probes described by Table 1.

Claim 12. The method of any one of claims 9 to 11, further comprising determining HPV viral load comprising determining a ratio between E6/E7 and beta globin, MSH2 or at least one other human gene target.

Claim 13. The method of any one of claims 9 to 12, further comprising detecting at least one biomarker that covers a junction between a disrupted host gene and at least part of an integrated HPV DNA.

Claim 14. The method of claim 9, further comprising detecting at least one biomarker selected from the group consisting of MAPK10 (Gene ID: 5602), PTPN13 (Gene ID: 5783), NUDT15 (Gene ID: 55270), MED4 (Gene ID: 29079), ITM2B (Gene ID: 9445), RB1 (Gene ID:

5925), LPAR6 (Gene ID: 10161), RAB11A (Gene ID: 8766), RPL13A (Gene ID: 23521), ZNF341 (Gene ID: 84905), OFD1 (Gene ID: 8481), DHRS3 (Gene ID: 9249), TBC1D22B (Gene ID: 55633), AFF3 (Gene ID: 3899), CXCL6 (Gene ID: 6372), PF4V1 (Gene ID: 5197), IMMP2L (Gene ID: 83943), MMP12 (Gene ID: 4321), WDR20 (Gene ID: 91833), ALDHA1A (Gene ID: 216), TPRG1 (Gene ID: 285386), TUBD1 (Gene ID: 51174), MAST4 (Gene ID: 375449), LOC100132167, NFIX (Gene ID: 4784), CCAT1 (Gene ID: 100507056), GPR137B (Gene ID: 7107), RAB22A (Gene ID: 57403), C9orf3, MACROD2 (Gene ID: 140733), DACH1 (Gene ID: 1602), ATP10A (Gene ID: 57194), SPG11 (Gene ID: 80208), SORD (Gene ID: 6652), COL4A4 (Gene ID: 1286), GATSL1 (Gene ID: 729438), GATSL2 (Gene ID: 729438), MAP2 (Gene ID: 4133), EPN1 (Gene ID: 29924), ATXN3L (Gene ID: 92552), EGFL6 (Gene ID: 25975), and MAGI2 (Gene ID: 9863).

Claim 15. The method of claim 13, wherein forward and reverse oligonucleotide primers that specifically amplify an integration junction between host genomic DNA and integrated HPV DNA are used to produce the biomarker.

Claim 16. The method of claim 13, wherein the biomarker is human OFD1 (NG_008872.1) and is produced and the forward and reverse primers described by Table 2 are used to amplify an OFD1 gene/HPV16 and the HPV16/OFD1 gene junctions as biomarkers.

Claim 17. The method of claim 13, wherein at least one specific nucleic acids sequence complementary to an integration junction between host genomic DNA and integrated HPV DNA is used for the detection by hybridization of a biomarker.

Claim 18. The method of claim 13, wherein said HPV DNA is selected from a DNA from the group consisting of HPV strains 16, 18, 21, 33, 45, 52 and 58.

Claim 19. A composition comprising one or more biomarkers as defined in any one of claims 15 or 16.

Claim 20. The composition of claim 19, wherein said one or more biomarkers is selected from the group consisting of MAPK10, PTPN13, NUDT15, MED4, ITM2B, RB1, LPAR6, RAB11A, RPL13A, ZNF341, OFD1, DHRS3, TBC1D22B, AFF3, CXCL6, PF4V1, IMMP2L, MMP12, WDR20, ALDHA1A, TPRG1, TUBD1, MAST4, LOC100132167, NFIX, CCAT1, GPR137B, RAB22A, C9orf3, MACROD2, DACH1, ATP10A, SPG11, SORD, COL4A4, GATSL1, GATSL2, MAP2, EPN1, ATXN3L, EGFL6, and MAGI2.

Claim 21. A kit comprising standardized and purified biomarkers that hybridize with host cell DNA and with integrated viral DNA sequences, and optionally, control reagents, one or more other reagents, supplies and/or equipment useful for detecting viral integration, in particular a kit comprising one or more biomarkers of claim 20.

Claim 22. A method of preparation of genomic DNA containing suspected integrated viral DNA from the cell sample comprising:

embedding cells into an agarose plug having an agarose concentration ranging from about 0.5 wt% to 1.5 wt%,

infusing the plugs containing the embedded cells with a proteolytic enzyme for a time and under conditions sufficient to substantially digest cellular proteins,

washing the plugs for a time and under conditions sufficient to remove episomal nucleic acids,

extracting cellular genomic DNA caught in the plugs containing or not the integrated viral DNA.

Claim 23. A method for assessing a risk of having or developing a cervical cancer comprising: detecting or quantifying a number of integrations of HPV DNA into, or an integration pattern of HPV DNA in a sample of genomic DNA obtained from a patient or subject, thereby assessing the risk of having or developing cervical cancer, wherein said sample of genomic DNA is obtained by the method of claim 1.

Claim 24. The method of claim 23, wherein a greater number of instances, or a greater amount of, integrated HPV DNA is indicative of a high risk of having or developing cancer or is indicative of a more aggressive or higher grade cancer compared to a patient or subject having fewer instances or lesser amounts of integrated HPV DNA.

Claim 25. The method of claim 23, wherein a different pattern of HPV DNA integrations into genomic host DNA, compared to those in a control subject or patient, is indicative of a high risk of having or developing cancer or is indicative of a more aggressive or higher grade cancer.

Claim 26. The method of claim 23 that comprises assessing a risk of having a cervical cancer.

Claim 27. The method of claim 23 that comprises assessing the risk of developing a cervical cancer.

Claim 28. The method of claim 23, wherein said HPV DNA is from a high risk or pathogenic strain of HPV.

Claim 29. The method of claim 23, wherein said HPV DNA is from a high risk or pathogenic strain of HPV selected from the group consisting of HPV strains 16, 18, 21, 33, 45, 52 and 58.

Claim 30. The method of claim 23, wherein said HPV is from a low risk or non-pathogenic strain of HPV.

Claim 31. The method of claim 23, wherein said HPV is from a low risk or non-pathogenic strain of HPV that is less pathogenic than any one of HPV strains 16, 18, 21, 33, 45, 52 and 58.

Claim 32. The method of any one of claims 23 to 28, further comprising detecting or quantifying a number of integrations of HPV DNA by comparison to a patient or subject not infected with HPV, or not infected with a pathogenic strain of HPV, having no lesions or other symptoms of HPV infection, or having substantially no antibody titer or cellular immunity to HPV or to a particular HPV strain.

Claim 33. The method of any one of claims 23 to 32, further comprising detecting or quantifying a number of integrations of HPV DNA by comparison to those in an earlier biological sample obtained from the same patient or subject.

Claim 34. The method of any one of claims 23 to 33, wherein said detecting or quantifying a number of integrations is performed using molecular combing of the genomic host DNA using probes that bind to HPV DNA sequences.

Claim 35. The method of any one of claims 23 to 34, wherein said detecting or quantifying a number of integrations is performed using molecular combing of the genomic host DNA using probes to HPV 16, 18, 31, 33, 45 35, 39, 51, 52, 56, 58, 59, 66 and 68.

Claim 36. The method of claim 23, wherein said detecting or quantifying a number of integrations is performed using molecular combing of the genomic host DNA using probes that bind to or cover HPV DNA L1 and L2, E1 and E2, and/or E6 and E7 sequences, wherein said probes may be labelled with the same different colored fluorescent tags.

Claim 37. The method of claim 23, wherein the patient or subject has a cervical dysplasia or has a positive PAP test.

Claim 38. The method of claim 23, wherein the patient or subject has been infected with human immunodeficiency virus (HIV), is immunosuppressed, has been exposed to diethylstilbestrol before birth, or is or has been treated for a precancerous cervical lesion or cervical cancer.

Claim 39. The method of claim 23, wherein the patient has or is at risk of having anal, vaginal, vulvar, penile or oropharyngeal cancer.

Claim 40. The method of claim 23, wherein the number or pattern of HPV integrations is quantified by, or correlated, with at least one of the following:

the number of HPV integration sites in host genomic DNA or the average number of such integrations,

the size in kb of HPV DNA integrations into host genomic DNA,

the number of HPV genomes integrated at each integration site,

the presence or absence of integrated HPV DNA,

the number of HPV integration sites per cellular genome,

the average number of HPV integration sites in host cells,

the mean number of HPV genomes integrated per integration site (or the mean size of integration sites),

maximum number of HPV genomes integrated per integration site (or the maximum size of integration sites),

minimum number of HPV genomes integrated per integration site (or minimum size of integration sites), or

number of HPV genomes integrated per cellular genome.

Claim 41. The method of claim 23, wherein the number or pattern of HPV integrations is correlated with at least one parameter of lesion status including:

normal histology (including all abnormalities without intraepithelial lesions or signs of viral infection such as metaplasia, cervicitis, decidual lesions or adenosis),
low grade (LG) lesion, corresponding to former CIN1,
high grade (HG) lesion, corresponding to former CIN2, 3 and CIS (carcinoma in situ) or AIS (adenocarcinoma in situ);
normal cervix,
Grade 1 atypical transformation (AT),
Grade 2 atypical transformation (AT),
TAG2 a if there are no major signs,
TAG2 b if there are major signs,
TAG2 c when the appearance is suggestive of invasive cancer; and/or
atypical transformation (minor or major).

Claim 42. The method of claim 23, wherein the number or pattern of HPV integrations is correlated with at least one parameter of cytological classification including:

negative for intraepithelial lesion or malignancy,
abnormal squamous cells,
typical squamous cells (ASC),
of undetermined significance (ASC-US),
cannot exclude high-grade squamous intraepithelial lesion (ASC-H),
low-Grade Squamous Intraepithelial Lesion (LSIL),
high-Grade Squamous Intraepithelial Lesion (HSIL),

squamous cell carcinoma,
abnormal glandular cells,
atypical glandular cells (AGC): endocervical (not otherwise specified (NOS) or commented),
endometrial or not otherwise specified
atypical glandular cells, favor neoplastic: endocervical or not otherwise specified
endocervical adenocarcinoma in situ (AIS), and/or
adenocarcinoma.

FIG. 1A

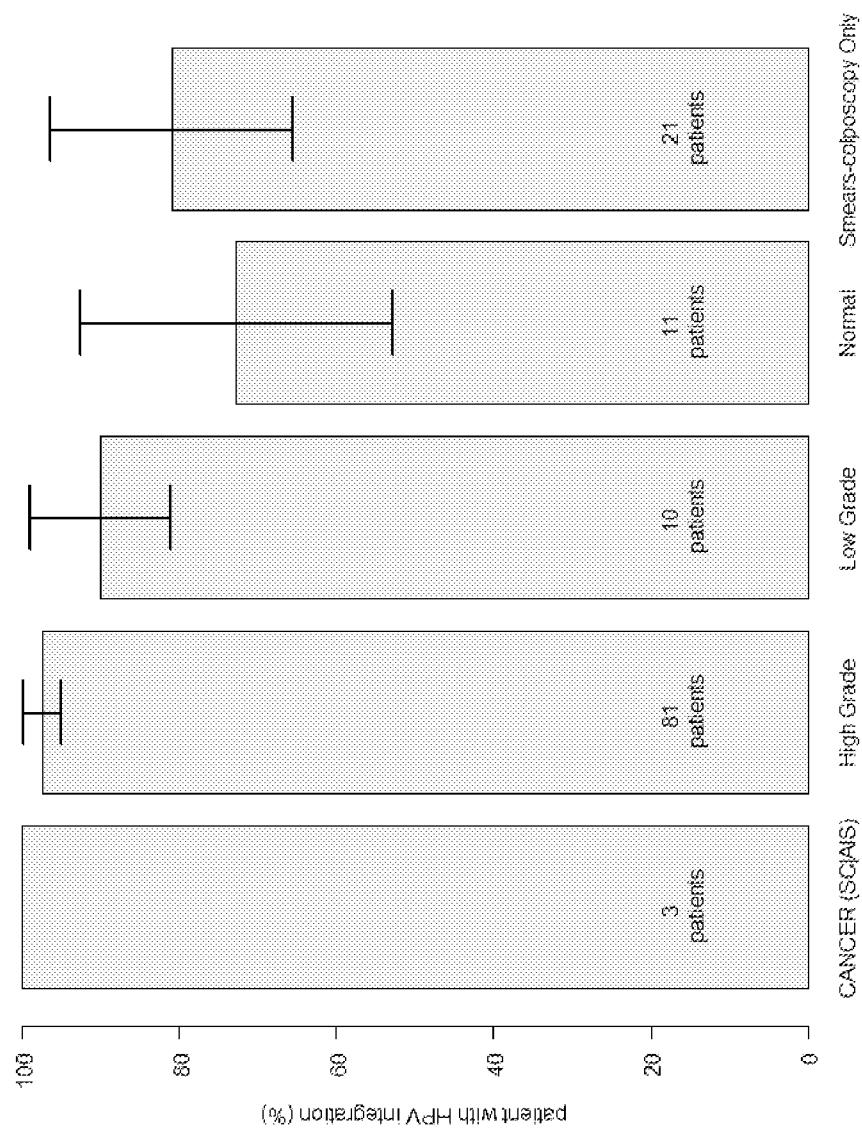


FIG. 1B



FIG. 2

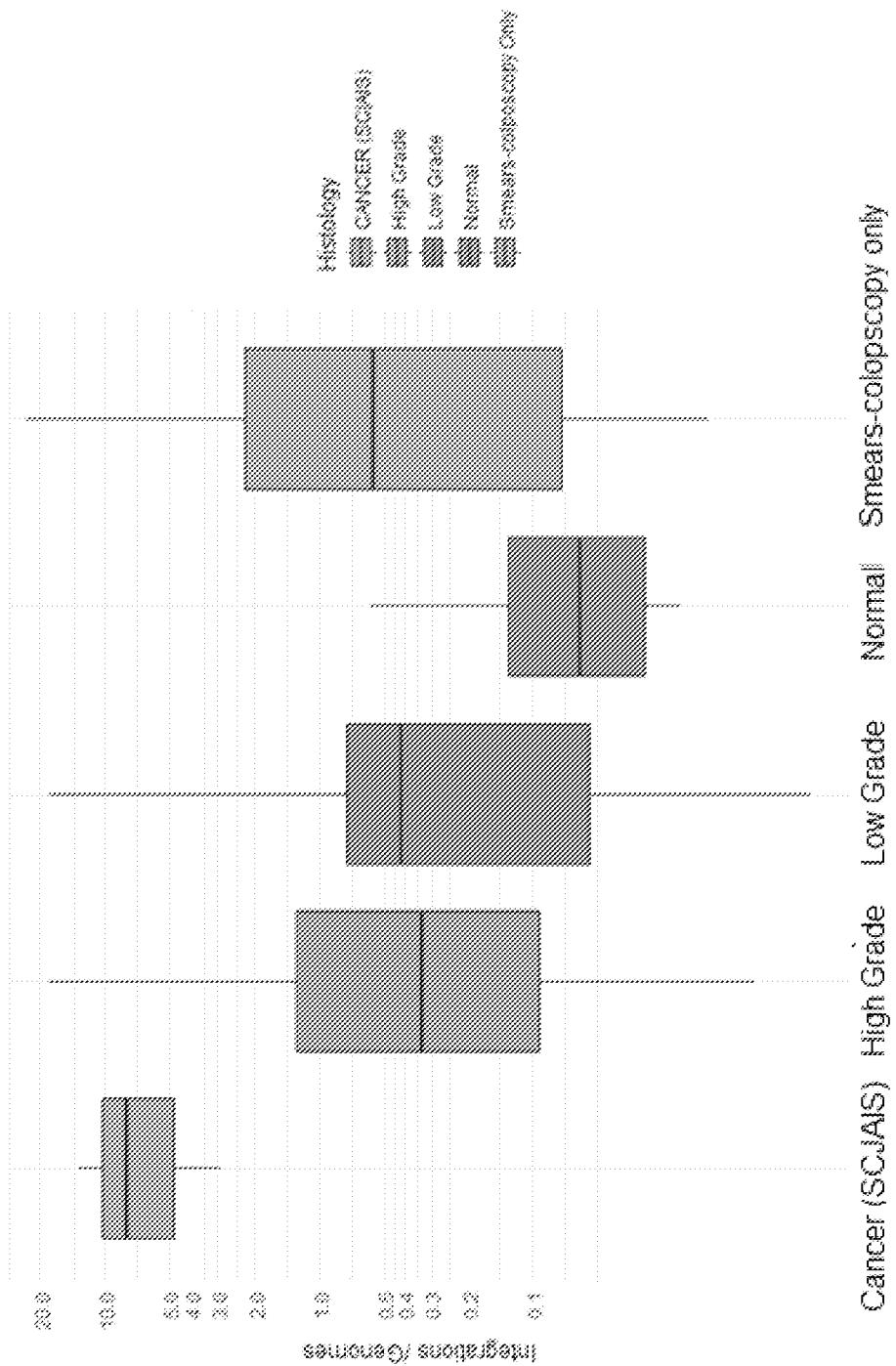
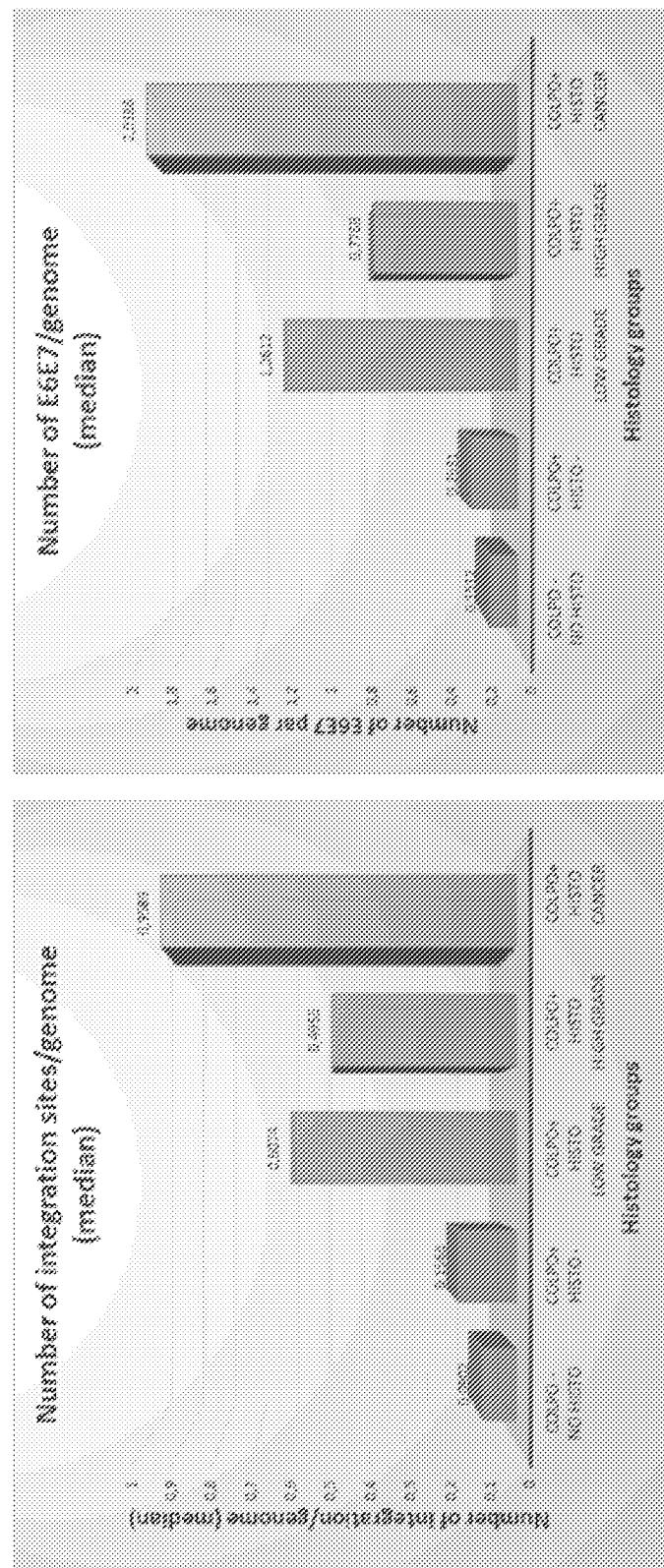
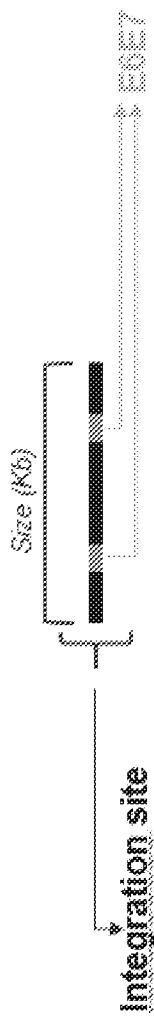
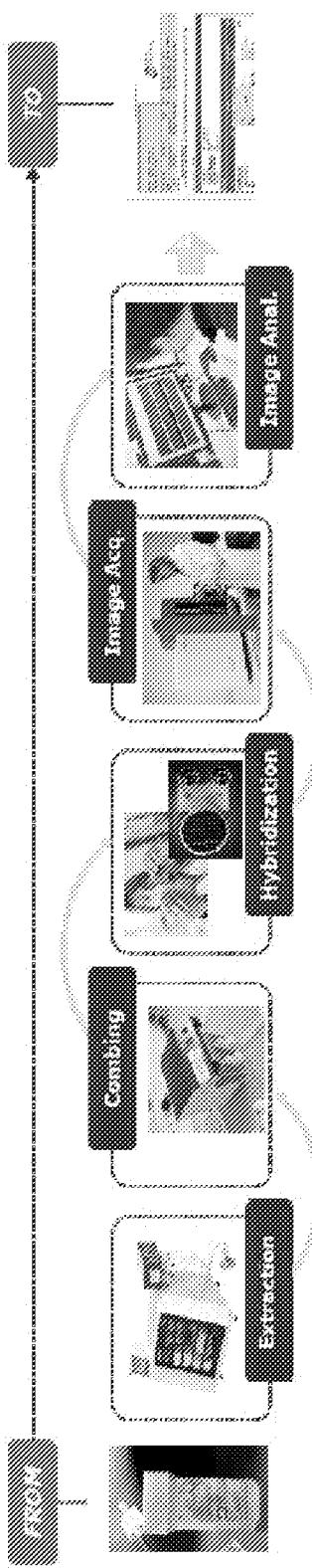


FIG. 3A



➤ Median value of integration per genome for LG+H/G+Cancer is 3 times higher than for normal.

FIG. 4



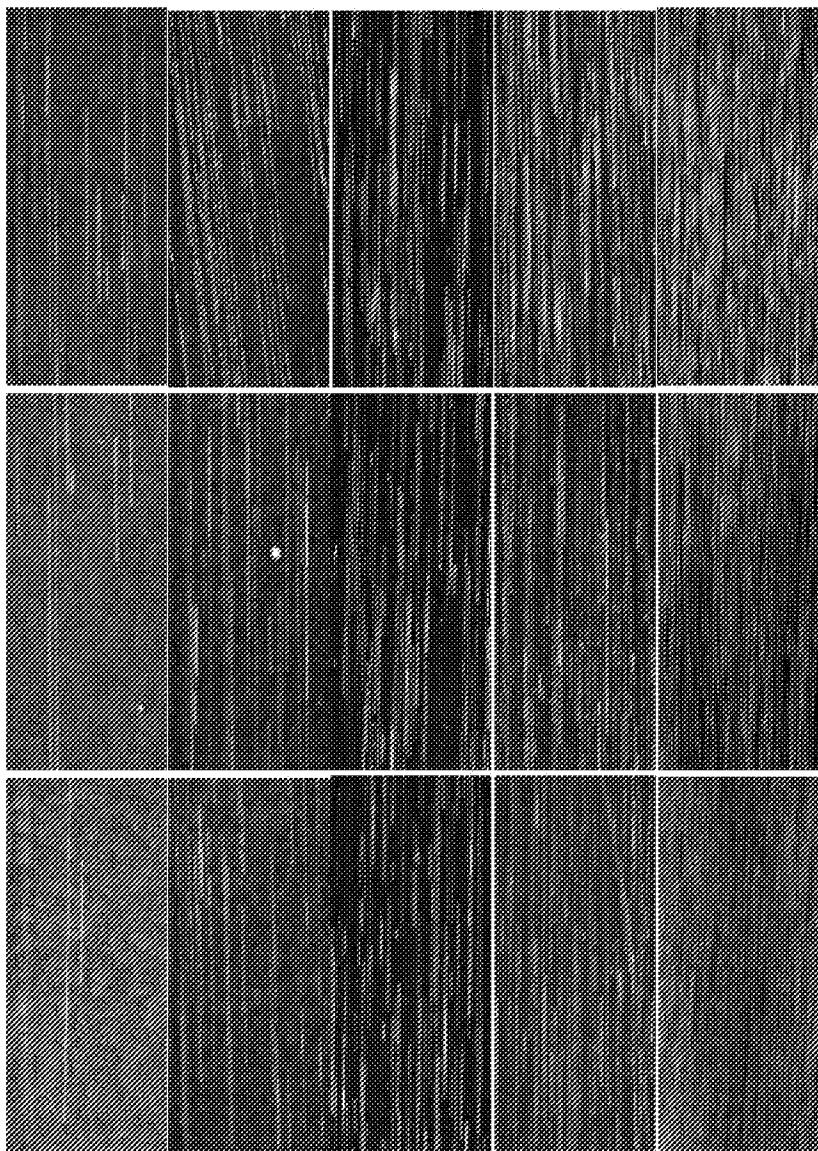


FIG. 5

FIG. 6

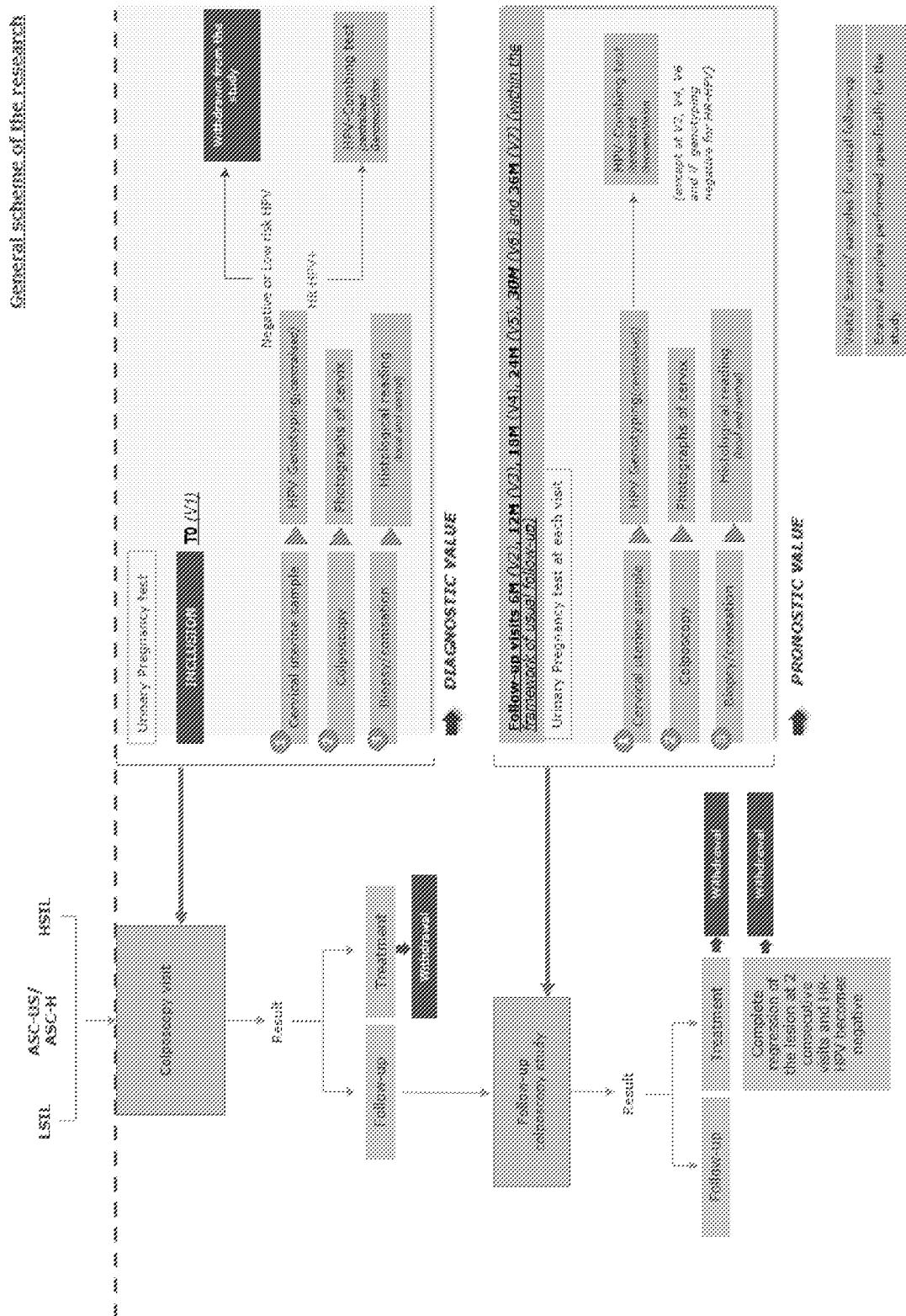


FIG. 7

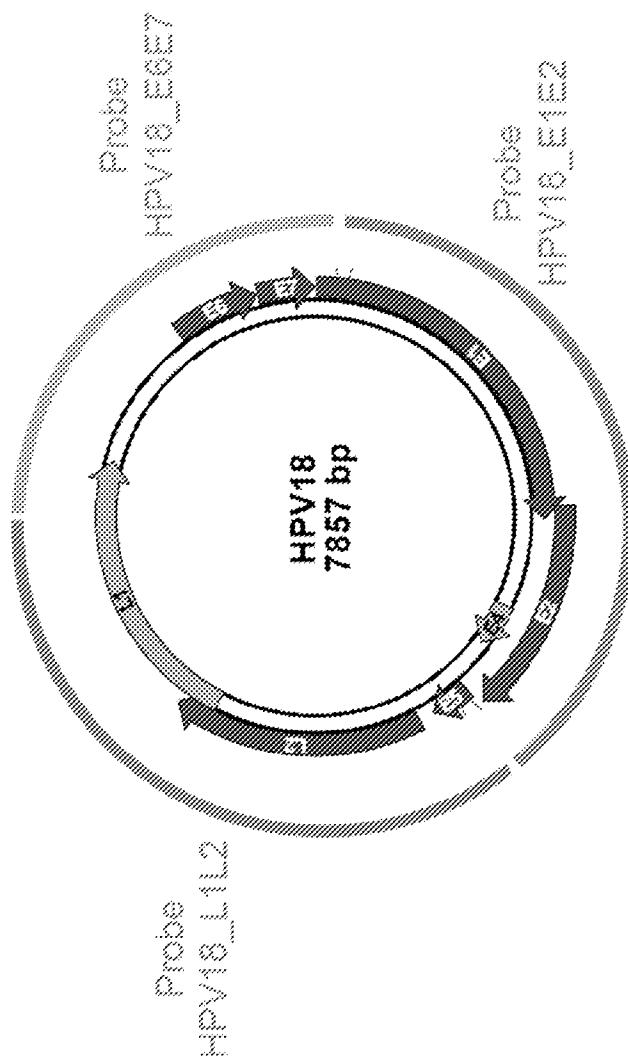
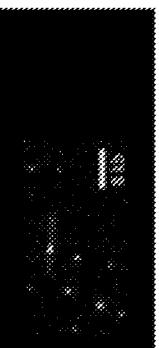


FIG. 8A

FIG 8 A

A/ Single integration
Integration of a single HPV genome at
the integration site



B/ Heterozygous tandem integration
Integration of several HPV genomes at the
integration site

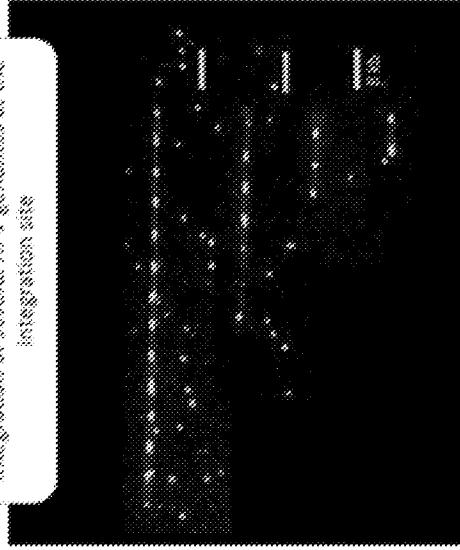


FIG. 8C

FIG 8 D

C/ Dispersed tandem integration
Integration of several HPV genomes spread
by the host DNA at an integration site

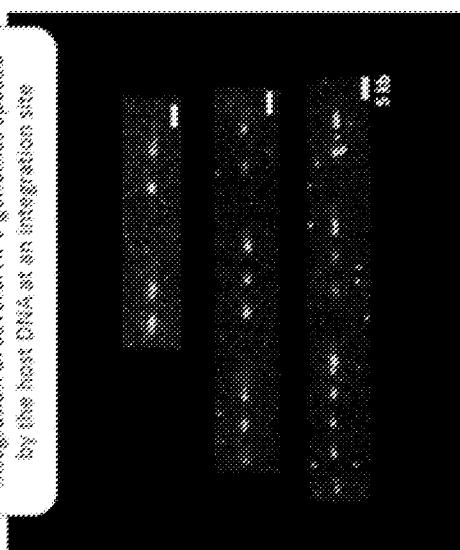


FIG. 8B

FIG 8 B

B/ Heterozygous tandem integration
Integration of several HPV genomes at the
integration site

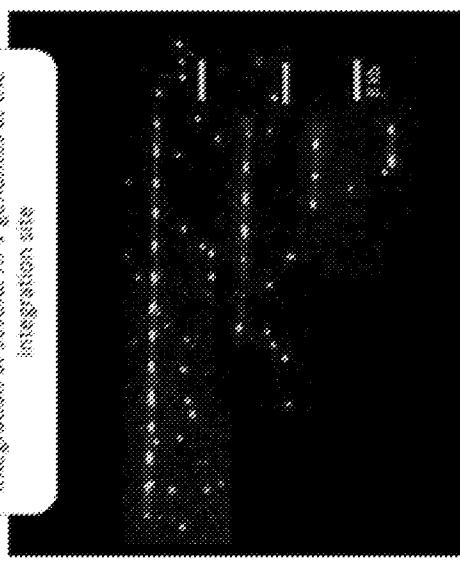


FIG. 8C

FIG 8 C

C/ Dispersed tandem integration
Integration of several HPV genomes spread
by the host DNA at an integration site

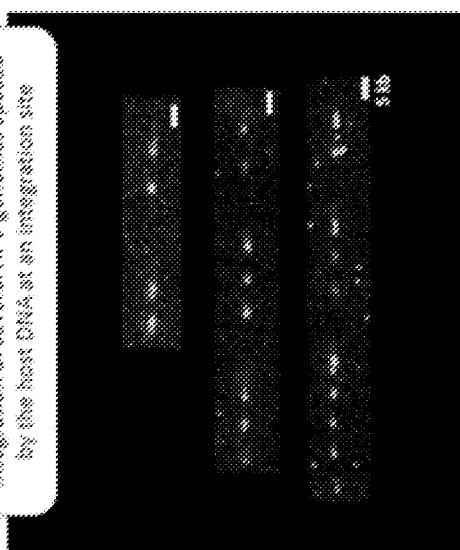


FIG. 9

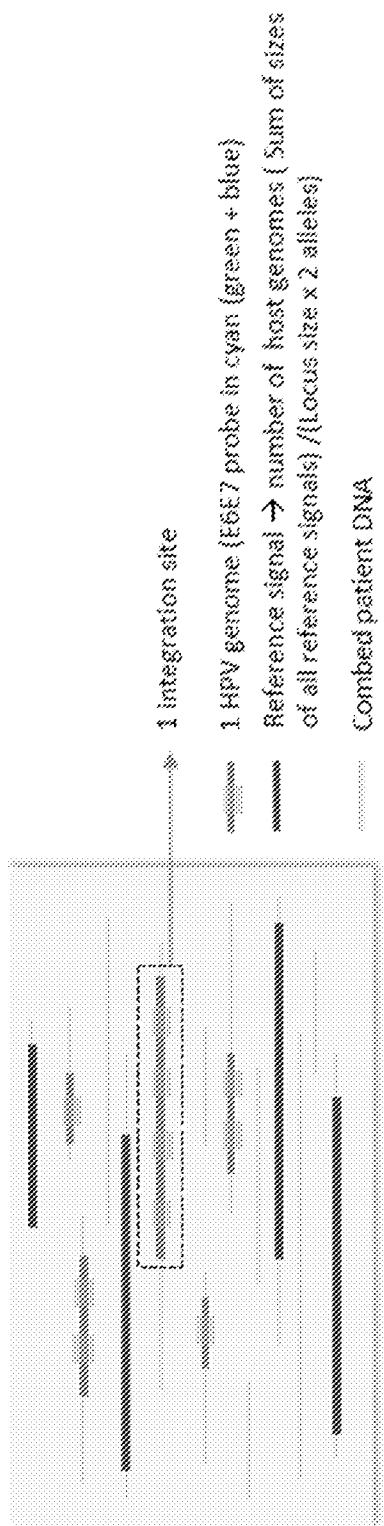


FIG. 10A**10A-Junction 1, OFD1 gene (GenBank refseq NG_008872.1) /HPV16 genome (GenBank refseq K02718.1) – SEQ ID No.37**

```

AGATGTTTACATTCAAGGTATTTTGGTTTGTGAGACGGAGTCTCGCTCTGTCACCTAGGCTGGA
GTGCAGTGGTGGCAGATCTGGCTCACTACAAGCTCCACCTCCACGGTTACGCCATTCTCCACCTCAGCC
TCCCGCAGTAGCTGGGACTACAGGCACCCACCAACACGCCGGCTAATTGGTATTGGTATTAGAGGGG
GGGTCTCAGCGTGTAGCCAGGATGATCTCGATCTCGACCTCGTGAATCCGCCCGCCTCGGCCTCCAA
AGTGCCTGGGATTAACGGGCTGAGCCACACGCCGACCCATTCAATGATTTCAGGAAATTGCTGAA
CAAGAAGTGAAGTTGGGAAATTCAAGTAGATCTGGAGAAACCAAGGGTAGATAGTTGATG
TTGAGCTGGGCTGGGCTGGGAGTACAGTGGCACAACTCAGCTCACTGCAACCTCTGGTCAAGTGAT
GTTGCCAGGGCTGGGAGTACAGTGGCACAACTCAGCTCACTGCAACCTCTGGTCAAGTGATG
TCTTGTGCTCAAGCCTCAGCCCTCCAAGTGACTGAGCTGGGACTACATGTTGATGGTAAACACGTGTGTA
AGCTGGGCTGGGCTGGGAGTACAGTGGCACAACTCAGCTCACTGCAACCTCTGGTCAAGTGATG
AATTTGCTGCTTATTCATTTGCTACATCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
TTTGTGAGGCCAGGGCCATTGTTGACTCTAACGGGATTCAGGCTGGGCTGGGCTGGGCTGGGCTGGGCT
TGTGTTTAAATAGTCTAATGCAACTAATGTTGAACTAATGTTGAACTAATGTTGAACTAATGTTGAA
CAATGCCCTGGTTGCTGGCAGCTGGCCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT
ACTACTGAAATCCTACATGTTGCTCAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA
GTGGGCAATAATTTGGCTGTTTAACTAACTTAACTTAACTTAAACTTAAACTTAAACTTAAACTTAA
GGCAACTAAATGGTACCCCTACATGGAACTGTTAGCTCATACATGGTAAAGGTAAAGGTAAAGGTAA

```

The left side of the integration junction (in white) corresponds to the OFD1 gene

Sequence (SEQ ID No.38) and the right side of the junction (highlighted in grey) corresponds to the HPV16 genome (SEQ ID No.39). In dark grey are shown the forward and reverse primers used for the biomarker amplification (the amplified sequence is SEQ ID No.40).

FIG. 10B

110B-Junction 2, HPV16 genome (GenBank refseq K02718.1) / OFD1 gene (ref seq NG_008872.1) SEQ ID No.41

ATGGCTGATCCCTGGTACCAATGGGAAGAGGGTACGGGATGTAATGGATGGTTTATGTAGGGCTG
TATGGAAAAAACACAGGGATGCTATATCAGATGACGAGAACGAAATGACAGTGATACAGGTGAAGA
TTTCGTTGATTTATAGTAATGATAATGATTAACTACAGGGCAGAACAGAGACAGCAGCACAATGCGTG
TTTACAGGAAATGAGATGCAACATAGAGATGCACTAAACGAAAGTATTGGGTAGTC
CACTTAGTGTATAATTAGTGGATGTTGACAAATAATTAGTCTAGATAATGTTATGAA
ACAAGTGAAGCTGCAAAAGCAGATTATGGAAAGAACAGGGATATGGCAATACTGAAAGTGGAA
ACTCAGGCGAGATGTTACAGGTAGAACGGGGCCATGAGACTGAACCATCTAGTTCCCTGATCTGACC
TTGAGTTGTAGCCAAACTAAAGGCAAGGGTCAAAAGAGCTCAGCAAGAGGGCGAAACGCTGGAAAAGGC
TTTCAGAAGTТАCCATGGAGGTCAATAAAACCTGTCCAAAGGCCACTAGCAGCAAAAGGCCACCA
TCTCTGCACTTGGTCAAGCTTCAAAACATTACTTCAAGGAAATGACAGTGTGCTGTGGAAAGCACTGAC
ACAGAGTTGTCTTGAGCAGCTCAAGTGGGCACACTTGAAGAAAGGAATGACAGTGTGCTGTGGAAAGCACTGAC
TGGCAGTGCAGCCCTCGAGGTCCCGGGGGCACTCTCCAGACGCCTCTCTCCACACCCCTTCCAAA
GGCAAAAGGCTCAGGAAATGTTCTGGAAAGGTAAAGCCACACCCGGCACAAGGGTTGGGGGTGCT

The left side of the junction (highlighted in grey) corresponds to the HPV16 genome (SEQ ID No.42)

and the right side of the junction (in white) corresponds to the OFD1 gene (SEQ HD No.43)

sequence. In dark grey are shown the forward and reverse primers used for the biomarker amplification (the amplified sequence is SEQ ID No.44).

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2019/001313

A. CLASSIFICATION OF SUBJECT MATTER
 INV. C12Q1/70 C12Q1/6806
 ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2006/096727 A2 (CELLAY LLC [US]; MOEN JR PHILLIP T [US]) 14 September 2006 (2006-09-14) cited in the application the whole document para. 6, 15, 18, 59, 105 -----	1-3,7-9
Y	the whole document para. 6, 15, 18, 59, 105 -----	2-6, 10-42
X	US 2015/024946 A1 (ZHANG YONG [CN] ET AL) 22 January 2015 (2015-01-22) the whole document para. 7, 8, 10 -----	19,21
Y	the whole document para. 7, 8, 10 -----	1-18,20, 22-42
X	CN 104 830 869 B (ZHUHAI YAMA BIOENGINEERING CO LTD) 1 June 2018 (2018-06-01) the whole document Table 3, para. 93 -----	19-21
Y	the whole document Table 3, para. 93 -----	1-18, 22-42
		-/-

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

4 March 2020

Date of mailing of the international search report

13/03/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Sauer, Tincuta

INTERNATIONAL SEARCH REPORT

International application No PCT/IB2019/001313

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>N DELAROQUE ET AL: "Persistent virus integration into the genome of its algal host, <i>Ectocarpus siliculosus</i> (Phaeophyceae).", <i>JOURNAL OF GENERAL VIROLOGY</i>., vol. 80, no. 6, 1 June 1999 (1999-06-01), pages 1367-1370, XP055673458, GB ISSN: 0022-1317, DOI: 10.1099/0022-1317-80-6-1367 the whole document abstract, p. 1367, col. 2 -----</p>	1-42
Y	<p>Gemonic Vision: "FiberPrep DNA Extraction Kit", , 1 September 2017 (2017-09-01), pages 1-18, XP055672748, Retrieved from the Internet: URL:http://www.genomicvision.com/wp-content/uploads/User-Manual_FiberPrep-EXT-001_version-3_September-2017.pdf [retrieved on 2020-03-02] the whole document p. 5, item 5; p. 8, item 9 -----</p>	1-42
Y	<p>WO 2018/207022 A1 (GENOMIC VISION [FR]; LEMEE FANNY [FR] ET AL.) 15 November 2018 (2018-11-15) cited in the application the whole document p. 15, claims 1-20 -----</p>	1-42

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2019/001313

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 2006096727	A2	14-09-2006	NONE	
US 2015024946	A1	22-01-2015	CN 103320522 A US 2015024946 A1	25-09-2013 22-01-2015
CN 104830869	B	01-06-2018	NONE	
WO 2018207022	A1	15-11-2018	EP 3622090 A1 US 2019112673 A1 WO 2018207022 A1	18-03-2020 18-04-2019 15-11-2018