



HAL
open science

Repurposing the Hybrid Capture 2 (HC2) screening test for whole-genome sequencing of human papillomaviruses

A. Debernardi, W. Jarassier, C. Soret, C. Mougin, Samuel Alizon, Ignacio G. Bravo, D. Guenat, J. L. Prétet, Q. Lepiller

► To cite this version:

A. Debernardi, W. Jarassier, C. Soret, C. Mougin, Samuel Alizon, et al.. Repurposing the Hybrid Capture 2 (HC2) screening test for whole-genome sequencing of human papillomaviruses. *Archives of Virology*, 2021, 166 (12), pp.3421-3425. 10.1007/s00705-021-05259-9 . hal-03425708

HAL Id: hal-03425708

<https://hal.science/hal-03425708>

Submitted on 4 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TITLE: Repurposing the Hybrid Capture 2 (HC2) screening test for HPV whole-genome sequencing

A. Debernardi¹, W. Jarassier², C. Soret^{1,2}, C. Mougin^{1,2}, S. Alizon³, I.G. Bravo^{3,4}, D. Guenat^{1,2}, J.L. Prétet^{1,2}, Q. Lepiller^{1,2,5}

¹EA3181, UBFC, LabEx LipSTIC ANR-11-LABX-0021, Besançon, France

²Centre National de Référence Papillomavirus, CHRU de Besançon, Besançon, France

³French National Center for Scientific Research (CNRS), Laboratory MIVEGEC (CNRS, IRD, Univ Montpellier), Montpellier, France

⁴Center for Research on the Ecology and Evolution of Diseases (CREES), Montpellier, France

⁵Department of virology, CHRU de Besançon, Besançon, France

Running title: High-throughput HPV sequencing upon HC2 capture

#Correspondence: Dr. Quentin Lepiller, Laboratoire de Virologie, CHU Besançon, 3, Boulevard Fleming, 25 030, Besançon France, E-Mail: q1lepiller@chu-besancon.fr

Phone number: +33(0)370632513

ORCID : 0000-0003-2892-4216

DECLARATIONS

FUNDING: *Part of this work was funded by the Région Franche-Comté. AD is a recipient of a predoctoral scholarship from the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.*

COMPETING INTERESTS: *None declared*

AVAILABILITY OF DATA AND MATERIAL: *All data are available upon request*

CODE AVAILABILITY: *Not applicable*

ETHICS APPROVAL: *All samples were stored into a biobank for which a declaration of preparation and storage of human samples for research use has been sent to the « Ministère de l'Enseignement Supérieur et de la Recherche » (n°DC-2014-2086).*

CONSENT TO PARTICIPATE: *Patients were informed about the storage and processing of their specimen. Opposition to the use of the specimen was collected.*

ABSTRACT:

Simple and standardized approaches for Human Papillomavirus (HPV) genome analysis by Next Generation Sequencing are needed. The aim of the study was to develop a protocol for high-risk (hr)-HPV deep sequencing, directly based on the commercial and widely used Hybrid Capture 2 (Qiagen) test, without any additional probe design. This protocol was applied to 15 HPV-positive and 2 HPV-negative cervical samples or cell lines, and validated at the genotype level by comparing the sequencing results to a commercial genotyping technique. The performance of our protocol, presented in this proof-of-principle study, supports its use for accurate characterizations of hrHPV genetic variations.

KEYWORDS: High-risk HPV, Next Generation Sequencing

INTRODUCTION

Human papillomaviruses (HPVs) are a leading cause of cancers worldwide at anogenital and oropharyngeal sites [1, 2]. HPVs are classified into genera, species, lineages, sub-lineages and variants. Bzhalava *et al.* estimated that at least 400 different HPV types exist [3]. Among the various HPV types infecting the squamous epithelium of the lower anogenital tracts of both males and females, some are rarely, if ever, found in cervical cancer and called low-risk HPVs (lrHPVs) while others are considered high-risk (hr) types because they have been detected in cancers. Many studies have confirmed that 13 hrHPVs (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68) are well-established as causing cervical precancerous and cancerous lesions and HPV16 is by far the most potent carcinogenic genotype [4]. Within specific genotypes, such as HPV16, different HPV variants sharing more than 98% identity may be identified [5] and can differ in their carcinogenicity [6]. Since some variants have been associated to a higher risk of developing cancers, their detection and characterization by next generation sequencing (NGS) techniques are becoming an important public health issue. From a practical point of view, there is a need to develop simple, robust and standardized NGS techniques for hrHPVs. Among commercial and in-house techniques available to easily detect hrHPVs, the Digene Hybrid Capture 2 High-Risk HPV DNA Test (HC2) (QIAGEN, Hilden, Germany) has been widely used and is still regarded as a reference to validate the performance of other HPV commercial kits [7]. This test does not perform any PCR-based amplification of the viral genome and includes a step of hybridization between denatured viral DNA and short-RNA probes designed to cover the complete genome of all hrHPVs. Here, we took advantage of this robust hybridization step to integrate it into a protocol of HPV library capture, followed by NGS. This repurposing of the HC2 assay does not require any additional probe design.

MAIN CONTENT

Fourteen cervical samples were obtained from the Department of Gynaecology and Obstetrics, Besançon University Hospital, France, in routine practice and characterized by conventional cytology and HPV testing (see below). All samples were included in the COCOPAH collection (n°DC-2014-2086). Samples specifically included in this study were selected for their HPV status to cover both HC2-positive and negative samples and several types of hrHPV. Cervical samples were obtained by scraping the cervix using a cytobrush immediately collected in 1 ml of a specific transport medium (Digene Specimen Transport Medium, Qiagen), transported to the laboratory within 24 h, and stored at 4°C until further processing for hrHPV testing by HC2.

C-33 cells (HPV-negative cervical cancer cells), Ca Ski cells (cervical cancer cells containing integrated HPV16 genomes), and W12 cells (cervical epithelial cells containing episomal copies of HPV16 genome) were cultured in Eagle's Minimum Essential Medium (EMEM), Roswell Park Memorial Institute (RPMI) medium, and F-Medium, respectively, all supplemented with 10% of Foetal Bovine Serum (FBS) and 1% of streptomycin/penicillin as previously described [8, 9], and used as either negative or positive controls, respectively.

HC2 assay was used for hrHPV screening according to the manufacturer's instructions. Briefly, cervical samples were denatured in a solution of sodium hydroxide. Then, denatured HPV DNA was hybridized to RNA probes designed to cover the complete genome of 13 hrHPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68). Hybridization products were secondarily captured on a solid phase coated with anti-hybrids antibodies. Hybrids were then reacted with alkaline phosphatase conjugated secondary anti-hybrids antibodies and detected with a chemiluminescent substrate. Results were expressed in Relative Light Units (RLU) and standardized with respect to the response for 1 pg/mL HPV16 DNA (Positive Control, PC). Samples with a RLU/PC ≥ 1 were considered positive.

All samples were also analysed by duplex quantitative real-time PCR, targeting HPV16 and HPV18 E6 genes, on LightCycler 480 (Roche), as previously described [10]. Results were expressed in HPV copies per ml (copies/ml). Furthermore, HPV genotypes were determined for all samples using the INNO-LiPA® HPV Genotyping Extra II kit (Fujirebio, Les Ulis, France), according to the manufacturer's instructions. This technique allows for the detection of 13 hrHPVs (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68), 6 probable hrHPVs (phrHPV 26, 53, 66, 70, 73, 82), and 13 lrHPVs (HPV 6, 11, 40, 42, 43, 44, 54, 61, 62, 67, 81, 83, 89).

Regarding the NGS protocol, DNA was extracted from the different cell lines and from cervical samples using the QIAmp DNA mini kit (QIAGEN), according to the manufacturer's instructions. For cervical samples, since DNA was already denatured in a basic solution for HC2 testing, a step of neutralization with potassium acetate (3M) and acetic acid (5M) was required prior to DNA extraction. Libraries were prepared with the KAPA Hyper Prep Kit (Roche Diagnostics, Meylan, France), according to the manufacturer's instructions, prior to capture with the HC2 assay (Figure 1).

Since the HC2 test is designed to capture HPV DNA by hybridization, HPV-specific RNA

probes provided in the HC2 kit were mixed with 20 μ L of the denatured libraries. Hybridized libraries were then captured at the bottom of the anti-hybrids coated wells, washed, and then eluted by heating 20 minutes at 90°C in a Tris-EDTA elution buffer. Eluted libraries were secondarily amplified and cleaned-up on magnetic beads using the KAPA Hyper Prep Kit (Roche Diagnostics). Sequencing was performed on MiSeq (Illumina, San Diego, CA, USA) using the MiSeq Reagent kit, according to the manufacturer's instructions.

The bioinformatics analysis pipeline is described in Table S1. Briefly, Sequencing quality was controlled (FastQC) and reads were trimmed (Cutadapt). Libraries were screened (Fastq_screen) against a panel of 32 complete HPV genomes (corresponding to genotypes detected by the INNO-LiPA[®] test) and the hg19 human genome (references in Table S2). A sample was suspected to contain HPV DNA if a minimum of 0.01% of reads specifically mapped to a reference HPV genome. Different quality threshold metrics (count of reads mapping viral genome, MAPQ score, coverage and insert size) were used to select exploitable HPV-associated sequences. For each sample, the mapping depth was evaluated on specific regions of the potential HPV genomes. The presence of multiple HPV genotypes in a sample was indicated by the number of mapped reads against one or several HPV references, and the proportion of viral genome covered.

Our NGS protocol with prior HPV DNA capture by the HC2 test (Figure 1) was applied to 14 cervical samples (4 HC2⁻/HPV16⁻, 6 HC2⁺/HPV16⁺, and 4 HC2⁺/HPV16⁻), 2 HPV16-positive cell lines (CaSki and W12), and to the HPV-negative C-33 cells. Raw sequencing data are summarized in Table S3. For each sample, from 3.18×10^5 to 2.48×10^6 raw reads were generated. After trimming, between 1.82×10^4 and 1.33×10^6 reads mapped to the human genome whereas 3.89×10^2 to 2.78×10^5 reads mapped to HPV genomes. The percentage of HPV-related reads that mapped to a specific HPV genotype varied from 0% (i.e., no specific genotype identified, as observed for samples S2-S4) to 99.83% (HPV16 in the case of W12 cells). When restricting the analysis to HPV16, the most prevalent and carcinogenic genotype, the fraction of the genome covered by sequencing reached at least 99.9% in 73% of cases (n=8/11) at 10X and 50X.

To evaluate the applicability and performance of our protocol, we compared the genotypes obtained from our sequencing analysis with the genotypes determined on the same sample by the commercial technique INNO-LiPA[®]. Results are shown in Table 1, including the results of the HC2 test and the HPV16/HPV18-specific PCR. The HC2 test exclusively detects hrHPVs, and, indeed, no NGS read mapped any HPV from samples with a negative HC2 test (S1 to S4). Three of these samples were nevertheless positive with the INNO-LiPA[®] assay to probably oncogenic HPV66 and HPV53 and to non-oncogenic HPV44 (respectively S1, S3 and S4).

All samples determined by PCR to contain HPV16 DNA (S5 to S8, S13, S14) were positive for HC2, as well as for INNO-LiPA[®] assay and for the NGS protocol. Moreover, HPV16 was appropriately detected by NGS from CaSki and W12 cells (containing integrated and episomal forms of HPV16 genome, respectively). Other hrHPVs detected by NGS on HC2 positive samples were consistent with INNO-LiPA[®] genotyping, namely HPV33, 35, and 56. Inconsistent results between INNO-LiPA[®] and NGS were found for HPV51 with

disagreements on both directions (S13-14), as well as with HPV82, detected only by NGS (S11). We further determined whether our protocol allows the identification of specific HPV16 variants. Since polymorphisms have been described inside the HPV16 E6 protein, including in positions 10 (R10G/I) and 83 (L83V) [11], we predicted the amino acid residues encoded by the HPV16 E6 sequences in these two positions for the 4 HPV16-positive samples S4-S8 and for the 2 HPV-16 positive cell lines CaSki and W12. The L83V mutation was observed in samples S5, S6, S7, and in the two cell lines, whereas the R10G mutation was only observed in the two cell lines (Table S4).

In this study, we present the proof-of-principle for a straightforward NGS approach for HPV analysis, by repurposing the power of the classical HC2 screening test for capturing oncogenic HPVs and applied it for standard library preparation and NGS without any additional probe design. We have validated this approach by comparing the HPV genotypes identified in clinical cervical samples with this new protocol against the results obtained using a commercial genotyping technique. The two techniques yielded very consistent results. In particular, HPV16 was perfectly identified by our sequencing protocol. In our study, the main discrepant results were observed for HPV51 which was not detected (S14), detected (S13) or even associated with HPV82 (S11). In this latter case, since HPV82 and HPV51 are phylogenetically related (*Alphapapillomavirus 5 species*) [5, 12], it is possible that a reduced sequencing quality may have impaired the differentiation between the 2 genotypes. In one sample (S9), HPV66 could be identified (with a poor coverage) by the sequencing technique, in association with a positive HC2 test result, although this genotype does not belong to the 13 hrHPVs detectable by the HC2 test. This phenomenon may be due to partial cross-reactivity of HC2 probes, which has been reported by the manufacturer and elsewhere [13, 14]. NGS may provide important information regarding HPV-induced carcinogenesis, both at the cellular and at the viral levels. In HPV-infected cells from clinical samples, NGS is increasingly used to identify genomic HPV integration sites, disrupted cellular genes, and genetic or epigenetic alterations of the host genome [15, 16]. At the viral level, NGS may help identifying specific HPV variant sublineages and genetic variations associated with an increased risk of developing cancers [17, 18]. For instance, by using a large collection of HVP-16 sequences obtained by NGS, Mirabello *et al.* have observed that the conservation of the viral E7 protein was critical for HPV16 carcinogenesis [18]. Moreover, a low prevalence of HPV16 APOBEC3-induced mutations has been associated to HPV persistence and progression to (pre-)cancerous lesions [16]. However, simple and standardised sequencing protocols are urgently needed for the screening of these hrHPV genetic variations. This study may have some limitations. First, since our approach is based on the hybridization of DNA libraries by a panel of RNA probes design to cover 13 hrHPVs, it cannot readily be extended to study other HPV genotypes. Furthermore, the validation of our sequencing protocol was limited to the comparison of HPV sequences at the genotype level and to the identification of previously described polymorphism in the HPV16 E6 sequence. Additional genetic variations, including a more exhaustive identification of HPV

intra-genotype variants, will be explored by further studies. Finally, our NGS protocol requires to purchase the commercial HC2 test (the cost of which may vary by country) but with no additional cost, since the use of the kit is left to the discretion of the purchaser. The protocol by itself is proposed in a non-commercial way.

In summary, we introduce a simple and reliable sequencing approach for 13 hrHPV genomes, which is directly appended to a commercial detection technique. The performance of this protocol at the HPV genotype level supports its use for accurate characterizations of hrHPV genetic variations.

Acknowledgements: We thank Professor Paul Lambert for kindly providing us with W12 cells.

REFERENCES

1. de Martel C, Plummer M, Vignat J, Franceschi S (2017) Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer* 141:664–670. <https://doi.org/10.1002/ijc.30716>
2. Walboomers JM, Jacobs MV, Manos MM, et al (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 189:12–19. [https://doi.org/10.1002/\(SICI\)1096-9896\(199909\)189:1<12::AID-PATH431>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F)
3. Bzhalava D, Eklund C, Dillner J (2015) International standardization and classification of human papillomavirus types. *Virology* 476:341–344. <https://doi.org/10.1016/j.virol.2014.12.028>
4. Guan P, Howell-Jones R, Li N, et al (2012) Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *Int J Cancer* 131:2349–2359. <https://doi.org/10.1002/ijc.27485>
5. Burk RD, Harari A, Chen Z (2013) Human papillomavirus genome variants. *Virology* 445:232–243. <https://doi.org/10.1016/j.virol.2013.07.018>
6. Mirabello L, Clarke MA, Nelson CW, et al (2018) The Intersection of HPV Epidemiology, Genomics and Mechanistic Studies of HPV-Mediated Carcinogenesis. *Viruses* 10:. <https://doi.org/10.3390/v10020080>
7. Meijer CJLM, Berkhof J, Castle PE, et al (2009) Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. *Int J Cancer* 124:516–520. <https://doi.org/10.1002/ijc.24010>

8. Paget-Bailly P, Meznad K, Bruyère D, et al (2019) Comparative RNA sequencing reveals that HPV16 E6 abrogates the effect of E6*I on ROS metabolism. *Sci Rep* 9:5938. <https://doi.org/10.1038/s41598-019-42393-6>
9. Vaisman CE, Del Moral-Hernandez O, Moreno-Campuzano S, et al (2018) C33-A cells transfected with E6*I or E6*II the short forms of HPV-16 E6, displayed opposite effects on cisplatin-induced apoptosis. *Virus Res* 247:94–101. <https://doi.org/10.1016/j.virusres.2018.02.009>
10. Jacquin E, Saunier M, Mauny F, et al (2013) Real-time duplex PCR for simultaneous HPV 16 and HPV 18 DNA quantitation. *Journal of Virological Methods* 193:498–502. <https://doi.org/10.1016/j.jviromet.2013.07.023>
11. Zehbe I, Voglino G, Delius H, et al (1998) Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms. *Lancet* 352:1441–1442. [https://doi.org/10.1016/S0140-6736\(05\)61263-9](https://doi.org/10.1016/S0140-6736(05)61263-9)
12. de Villiers E-M (2013) Cross-roads in the classification of papillomaviruses. *Virology* 445:2–10. <https://doi.org/10.1016/j.virol.2013.04.023>
13. Barzon L, Militello V, Pagni S, Palù G (2012) Comparison of INNO-LiPA Genotyping Extra and Hybrid Capture 2 assays for detection of carcinogenic human papillomavirus genotypes. *Journal of Clinical Virology* 55:256–261. <https://doi.org/10.1016/j.jcv.2012.07.013>
14. Poljak M, Marin IJ, Seme K, Vince A (2002) Hybrid Capture II HPV Test detects at least 15 human papillomavirus genotypes not included in its current high-risk probe cocktail. *Journal of Clinical Virology* 25:89–97. [https://doi.org/10.1016/S1386-6532\(02\)00187-7](https://doi.org/10.1016/S1386-6532(02)00187-7)
15. Tuna M, Amos CI (2017) Next generation sequencing and its applications in HPV-associated cancers. *Oncotarget* 8:8877–8889. <https://doi.org/10.18632/oncotarget.12830>
16. Zhu B, Xiao Y, Yeager M, et al (2020) Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance. *Nat Commun* 11:886. <https://doi.org/10.1038/s41467-020-14730-1>
17. Mirabello L, Yeager M, Cullen M, et al (2016) HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. *J Natl Cancer Inst* 108:. <https://doi.org/10.1093/jnci/djw100>
18. Mirabello L, Yeager M, Yu K, et al (2017) HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell* 170:1164–1174.e6. <https://doi.org/10.1016/j.cell.2017.08.001>

TABLES AND FIGURES

Figure 1: Library preparation and capture. Library preparation included the classical steps of DNA fragmentation, end-repair and A-tailing, adapter ligation, and amplification (steps 1-4). Denatured library was hybridized to hrHPV-specific RNA probes from the commercial HC2 test. Hybrids were captured by anti-hybrids antibodies and secondarily eluted, amplified and sequenced on MiSeq (steps 5-8).

Table 1: Results of HPV testing. Cervical samples (S1 to S14) were characterized for their HPV status by HC2 test, HPV16-specific PCR, and INNO-LiPA® genotyping test. Results were compared to HPV genotypes determined by our sequencing protocol (NGS). Sequencing was also performed on 2 HPV⁺ and 1 HPV⁻ cell lines. Genotypes identified by NGS with a poor genome coverage are indicated in parentheses, and hrHPV genotypes appear in bold.

Table 1

Technique	Cervical samples														Cell lines		
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	C-33	CaSki	W12
HC2 (RLU/PC)	0	0	0	0	1610	1017	1694	861	346	172	2393	1053	140	324			
HPV16 (Copies/mL)	<1	<1	<1	<1	1E+06	4.0E+04	2.7E+04	1.8E+04	<1	<1	<1	<1	2.1E+02	9.1E+03			
HPV18 (Copies/ml)	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1			
INNO-LiPA® (Genotype)	66	0	53	44	16	16	16	16	66	33	51	51+56	16+35	16+51+56 +44			
NGS (Genotype)	0	0	0	0	16	16	16	16 + (56)	(66)	33	51 + 82	(51) + 56	16+35+51	16+56	0	16	16

Table S1 : Details on the bioinformatics analysis pipeline.

Step	Software (version)	Comment	Distributor or reference
Sequencing quality control	FastQC (v0.11.8)	-	[1]
Reads trimming	Cutadapt (v2.0)	Parameters: -O 7 -q 30,30 -m 35 -e 0.1 Additional options : -a -A -g -G with a list of universal illumina adapters	[2]
Screening of the libraries	Fastq_screen (v0.13.0)	Standard parameters with the sequence aligner BWA MEM (v0.7.17)	[3]
Mapping score determination and sorting for samples matching with HPV genome	BWA MEM (v0.7.17) Samtools (v1.9)	-	[4]
Removal of duplicates	Picard (v2.18.27)	-	[5]
Control of mapping quality	Qualimap bamqc v2.2.1	-	[6]
Compilation of outputs	Multiqc v1.0	-	[7]
Filtration of exploitable HPV-associated sequences	Custom program in R (3.6.0) ProgressiveMauve (v2015.02.13)	Specific parameters for ProgressiveMauve: --seed-family --scoring-scheme=ancestral --collinear	[8,9]

[1] Andrews, S. (2010). FastQC: A Quality Control tool for High Throughput Sequence Data.

[2] Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10–12.

[3] Wingett, S.W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 7, 1338.

[4] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

[5] Broad Institute (2019). Picard Toolkit.

[6] García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28, 2678–2679.

[7] Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.

[8] Darling, A.E., Mau, B., and Perna, N.T. (2009). Progressive Mauve: Multiple alignment of genomes with gene flux and rearrangement. *ArXiv:0910.5780 [q-Bio]*.

[9] Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* 5, e11147.

Table S2 : HPV references used.

Genotype	Sequence Identifiers (GI)	Extraction data	Genotype	Sequence Identifiers (GI)	Extraction data
HPV6	60955	24/01/2019	HPV53	9627377	24/01/2019
HPV11	333026	24/01/2019	HPV54	9628437	24/01/2019
HPV16	333031	24/01/2019	HPV56	397053	24/01/2019
HPV18	60975	24/01/2019	HPV58	222386	24/01/2019
HPV26	9627305	24/01/2019	HPV59	557236	24/01/2019
HPV31	333048	24/01/2019	HPV61	9628574	24/01/2019
HPV33	333049	24/01/2019	HPV62	39932599	24/01/2019
HPV35	396997	24/01/2019	HPV66	1020290	24/01/2019
HPV39	333245	24/01/2019	HPV67	3228267	24/01/2019
HPV40	397014	24/01/2019	HPV68	71726685	24/01/2019
HPV42	333211	24/01/2019	HPV70	1173493	24/01/2019
HPV43	40804474	24/01/2019	HPV73	1491692	24/01/2019
HPV44	1020242	24/01/2019	HPV81	40804509	24/01/2019
HPV45	397022	24/01/2019	HPV82	6970427	24/01/2019
HPV51	333087	24/01/2019	HPV83	5059324	24/01/2019
HPV52	397038	24/01/2019	HPV89	22095322	24/01/2019

Table S3 : Sequencing data.

Sample	Sample Type	Raw reads	Trimmed reads	Reads mapped to human genome	Reads mapped to HPV genome	Reads mapped to the specific HPV genotype after filtering	% of HPV-associated reads mapping to the specific genotype	Weighted average Insert Size (bp)	Fraction of HPV genome covered by minimum	
									10 x	50x
HPV16										
S5	Human	2.48E+06	2.28E+06	1.33E+06	2.78E+05	2.78E+05	99,82%	360	100.0%	100.0%
S6	Human	9.65E+05	9.00E+05	5.51E+05	3.34E+04	3.31E+04	99,11%	451	100.0%	100.0%
S7	Human	4.08E+05	3.34E+05	1.82E+04	4.53E+02	3.30E+02	72,85%	426	90.4%	2.20%
S8	Human	8.95E+05	8.35E+05	5.37E+05	1.82E+04	1.75E+04	95,91%	431	100.0%	99.9%
S13	Human	8.71E+05	7.99E+05	5.36E+05	2.70E+03	4.81E+02	17,82%	228	58.3%	7.50%
S14	Human	7.98E+05	7.45E+05	4.52E+05	4.02E+03	3.21E+03	79,74%	401	100.0%	99.9%
CaSki 1ng	Cell line	5.94E+05	5.62E+05	3.60E+05	5.69E+03	5.26E+03	92,39%	332	100.0%	99.9%
CaSki 10ng (1)	Cell line	5.67E+05	5.40E+05	3.68E+05	8.32E+03	8.01E+03	96,19%	315	100.0%	99.9%
CaSki 10 ng (2)	Cell line	5.92E+05	5.63E+05	3.84E+05	9.19E+03	8.95E+03	97,35%	360	100.0%	99.9%
CaSki 50 ng	Cell line	3.18E+05	2.99E+05	2.07E+05	2.90E+03	2.79E+03	96,10%	286	97.0%	74.0%
W12 10 ng	Cell line	1.01E+06	8.54E+05	5.04E+05	1.12E+05	1.12E+05	99,83%	480	100.0%	100.0%
HPV33										
S10	Human	8.03E+05	7.64E+05	5.41E+05	1.04E+03	3.24E+02	31,06%	180	59.0%	0.4%
HPV35										
S13	Human	8.71E+05	7.99E+05	5.36E+05	2.70E+03	1.68E+03	62,10%	461	92.8%	53.2%
HPV51										
S11	Human	2.44E+06	2.18E+06	1.24E+06	2.95E+04	2.69E+04	91,11%	297	100.0%	100.0%
S12	Human	7.94E+05	7.37E+05	4.80E+05	3.84E+04	4.60E+01	0,12%	20	0.5%	0.2%
S13	Human	8.71E+05	7.99E+05	5.36E+05	2.70E+03	2.37E+02	8,78%	369	64.9%	0.2%
HPV56										
S8	Human	8.95E+05	8.35E+05	5.37E+05	1.82E+04	7.50E+01	0,41%	30	1.7%	0.0%
S12	Human	7.94E+05	7.37E+05	4.80E+05	3.84E+04	3.79E+04	98,70%	439	100.0%	100.0%
S14	Human	7.98E+05	7.45E+05	4.52E+05	4.02E+03	4.08E+02	10,14%	411	77.6%	18.9%
HPV66										
S1	Human	6.85E+05	6.49E+05	4.59E+05	3.89E+02	1.00E+00	0,26%			
S9	Human	9.58E+05	9.19E+05	6.63E+05	1.19E+03	9.70E+01	8,19%	172	10.7%	2.3%
HPV82										
S11	Human	2.44E+06	2.18E+06	1.24E+06	2.95E+04	2.51E+03	8,50%	218	82.2%	33.6%
No HPV										
S2	Human	7.29E+05	6.95E+05	5.06E+05	4.27E+02		0,00%			
S3	Human	9.96E+05	9.52E+05	6.94E+05	5.31E+02		0,00%			
S4	Human	5.66E+05	5.43E+05	3.87E+05	5.70E+02		0,00%			

Negative control										
C-33 10 ng (1)	Cell line	6.24E+05	5.91E+05	3.83E+05	7.36E+02		0,00%			
C-33 10 ng (2)	Cell line	6.86E+05	6.48E+05	4.71E+05	4.78E+02		0,00%			

Table S4: Prediction of amino acid residues in HPV16 E6 protein.

HPV16 positive samples	Prediction of the amino acid residues in HPV16 E6	
	Position 10	Position 83
S5	R (Arginine)	V (Valine)
S6	R (Arginine)	V (Valine)
S7	R (Arginine)	V (Valine)
S8	R (Arginine)	L (Leucine)
CaSki cells	G (Glycine)	V (Valine)
W12 cells	G (Glycine)	V (Valine)