BMJ Open

Natural history, dynamics, and ecology of human papillomaviruses in genital infections of young women: protocol of the PAPCLEAR cohort study

Carmen Lía Murall,[1] Massilva Rahmoun,[1] Christian Selinger,[1] Monique Baldellou,[2] Claire Bernat,[1] Marine Bonneau,[3] Vanina Boué,[1] Mathilde Buisson,[4] Guillaume Christophe,[2] Giuseppe D'Auria,[5,6] Florence De Taroni,[2] Vincent Foulongne,[7,8] Rémy Froissart,[1] Christelle Graf,[3] Sophie Grasset,[1,2] Soraya Groc,[1,8] Christophe Hirtz,[9] Audrey Jaussent,[10] Julie Lajoie,[11] Frédérique Lorcy,[12] Eric Picot,[2] Marie-Christine Picot,[10] Jacques Ravel,[13] Jacques Reynes,[14] Thérèse Rousset,[12] Aziza Seddiki,[4] Martine Teirlinck,[2] Vincent Tribout,[2] Édouard Tuaillon,[7,8] Tim Waterboer,[15] Nathalie Jacobs,[16] Ignacio G Bravo,[1] Michel Segondy,[7,8] Nathalie Boulle,[7,12] Samuel Alizon[1]

For numbered affiliations see end of article.

Correspondence to
Dr Samuel Alizon;
samuel.alizon@cnrs.fr

ABSTRACT

Introduction Human papillomaviruses (HPVs) are responsible for one-third of all cancers caused by infections. Most HPV studies focus on chronic infections and cancers, and we know little about the early stages of the infection. Our main objective is to better understand the course and natural history of cervical HPV infections in healthy, unvaccinated and vaccinated, young women, by characterising the dynamics of various infection-related populations (virus, epithelial cells, vaginal microbiota and immune effectors). Another objective is to analyse HPV diversity within hosts, and in the study population, in relation to co-factors (lifestyle characteristics, vaccination status, vaginal microbiota, human genetics).

Methods and analysis The PAPCLEAR study is a single center longitudinal study following 150 women, aged 18–25 years, for up to 2 years. Visits occur every 2 or 4 months (depending on HPV status) during which several variables are measured, such as behaviours (via questionnaires), vaginal pH, HPV presence and viral load (via qPCR), local concentrations of cytokines (via MesoScale Discovery technology) and immune cells (via flow cytometry). Additional analyses are outsourced, such as titration of circulating anti-HPV antibodies, vaginal microbiota sequencing (16S and ITS1 loci) and human genotyping. To increase the statistical power of the epidemiological arm of the study, an additional 150 women are screened cross-sectionally. Finally, to maximise the resolution of the time series, participants are asked to perform weekly self-samples at home. Statistical analyses will involve classical tools in epidemiology, genomics and virus kinetics, and will be performed or coordinated by the Centre National de la Recherche Scientifique (CNRS) in Montpellier.

Ethics and dissemination This study has been approved by the Comité de Protection des Personnes Sud Méditerranée I (reference number 2016-A00712-49); by the Comité Consultatif sur le Traitement de l'Information

Strengths and limitations of this study

► Short time interval between the visits (every 2 months for infected women) and additional self-sampling every week at home.
► The combination of virological (virus load), immunological (cytokine concentrations and immune cell percentages) and environmental (vaginal microbiota composition, pH) measurements at each visit.
► A limitation is that the density of the follow-up limits the number of participants (n=150) which can affect the power of epidemiological analyses.
► We complement the longitudinal study with a cross-sectional study of n=150 women to increase statistical power.

en matière de Recherche dans le domaine de la Santé (reference number 16.504); by the Commission Nationale Informatique et Libertés (reference number MMS/ABD/ AR1612278, decision number DR-2016–488) and by the Agence Nationale de Sécurité du Médicament et des Produits de Santé (reference 20160072000007). Results will be published in preprint servers, peer-reviewed journals and disseminated through conferences.

Trial registration number NCT02946346; Pre-results.

INTRODUCTION

Epidemiology of human papillomavirus genital infections in young adults and public health implications

Infections by human papillomaviruses (HPVs) are likely the most common sexually transmitted infection (STI) globally. It is often estimated that, worldwide, more than 80% of sexually active individuals will be

infected by an HPV type.[1 2] In France, a study performed in 2013 in the Paris area estimated the prevalence of HPV genital infections to be 25% in women below 25 years of age.[3] In the area of Montpellier (France), the prevalence of oncogenic HPVs (also referred to as 'high-risk' (HR) HPVs) in pregnant women aged 16–42 years was close to 20%.[4] These numbers are consistent with worldwide estimates according to which HPVs are most prevalent in women under 25 years of age, with an estimated overall prevalence of 24%.[5]

Fortunately, the vast majority of infections by HPVs are asymptomatic and benign. Even for HPV16, which is probably the most oncogenic human virus, only a minority of infections (less than 10%) become persistent,[6] and then a minority of these (12%) progress to cancer if untreated.[1 7] Indeed, it is estimated that approximately 70%–100% of infections by HPVs are cleared within 12–24 months, with strong differences between virus types.[6 8–10] Recent studies suggest that primo-infections could be shorter in young girls[11] but, in general, there are many unknowns about the biology of non-persisting infections.[12]

Our lack of knowledge partly comes from the fact that in vaccine trials, from which most of the data on infection duration originate, participants are followed every 6 months for several years.[6 8 10 13] This frequency is sufficient to estimate the time to clearance (or persistence) but it is not precise enough to understand the within-host dynamics, often referred to as 'kinetics',[14] of infections that last on average 6–24 months. Arbitrarily, after 24 months, an infection is often considered to be persistent.[15]

Some factors have been shown to correlate with persistence (eg, immunosuppression, smoking and co-infection with other STIs[16]) but we do not know how these affect viral kinetics. Also, some changes in the interactions between virus and the host immunity appear to be related to persistence and disease progression[17–20] but, again, we do not know the underlying interactions between the viruses, the host target cells, and the immune response in acute infections.[12] Finally, it has been argued that the vaginal microbiota may differ between HPV-infected and HPV-uninfected women[21] and that specific microbiota composition may interact with HPV detection.[22] However, it is difficult to disentangle the cause and the consequence. For instance, does the microbiota composition change after the establishment of an HPV infection, or do certain microbiota compositions increase susceptibility to HPV infection?

A better understanding of the within-host infection dynamics and of the determinants of clearance and persistence of viral infection is particularly important in the context of vaccination.[23–26] Indeed, the long-term efficacy of the anti-HPVs vaccines at the population level will largely depend on the within-host viral dynamics because, ultimately, most selective pressures on viral populations occur via the immune response.[27] Furthermore, a better understanding of acute HPV infections can shed new light on issues related to latency, fertility or immunotherapies.[12]

Prevention strategies and treatment

Treatment

Since most infections by HPVs are benign in young adults and clear within six to 24 months, the current standard of care is to avoid over-treatment, even in the presence of cervical lesions.[28] Clinical interventions (colposcopies, biopsies and surgery) are less often performed with young women (<25 years) and only for high-grade (precancerous) lesions (cervical intraepithelial neoplasia grade 2 (CIN-2) or more). Low-grade lesions (CIN-1) are not systematically treated but rather monitored yearly to detect any progression to high-grade lesions.

Genital warts caused by non-oncogenic HPVs (often referred to as 'low-risk' (LR) HPVs) can be removed by surgery or treated with bi- and trichloroacetic acid, cryotherapy or other treatments.[29]

HPV vaccination

There are currently three licensed vaccines: a bivalent vaccine (Cervarix) targeting HPV16 and HPV18 (together accounting for 70% of cervical cancers[1]), a quadrivalent vaccine (Gardasil) that additionally targets HPV6 and HPV11 (non-oncogenic, but highly prevalent and associated to benign proliferative lesions) and, since 2014, a nonavalent vaccine (Gardasil 9) that targets five more oncogenic types (HPV31, HPV33, HPV45, HPV52 and HPV58, which altogether account for 20% of cervical cancers[25]). These vaccines succeed in eliciting a protective immune response against new infections by the targeted viruses, and are used throughout the world, although with wide variation in coverage (for reviews, see, e.g, refs[30 31]).

Vaccination campaigns in France started in 2006 but with limited coverage: it reached 28.5% in 2008[32] and has been decreasing ever since.[33] The vaccine is recommended for girls from 11–14 years of age, currently with a vaccination scheme of two doses with a 6-month interval. A catch-up is organised for girls aged 15–19 years, with a three-dose vaccination scheme. Vaccination is reimbursed by the French Social Security but is not mandatory. It is also recommended for men who have sex with men as well as for immuno-compromised people.[33] Vaccination is now the primary prevention strategy against cervical cancers.

Screening

In France, the secondary prevention strategy against cervical cancer is routine individual cytology-based screening for precancerous and cancerous cervical lesions in women between 25 and 65 years old. Cytology can also be performed in younger women if they report risk factors for cervical cancer (multiple partners, chronic STIs or HIV infection[33]). Detection of oncogenic HPVs is proposed for triage in case of abnormal cytology (ie, high-grade or low-grade squamous intraepithelial lesion (HSIL and LSIL, respectively) or atypical squamous cells of undetermined significance (ASCUS)).

Primary objectives

The first primary objective of this cohort study is to decipher the kinetics and ecology of cervical HPV infections in healthy young women, that is, follow the population dynamics of the virus, the target epithelial cells, the vaginal microbiota, and the immune effectors.

The second primary objective is to characterise the diversity of genital HPVs in young women in the region of Montpellier in relationship with their lifestyle, vaccination status, vaginal microbiota and human genetics.

Secondary objectives

A secondary objective is to characterise the acquisition and clearance dynamics of cervical HPV infections as a function of viral diversity, host immunity, vaginal microbiota and human genetics.

A final objective is to investigate variations in genetic diversity of HPVs during cervical infections.

METHODS AND ANALYSIS
Participants

The study population is composed of young women at risk of HPV infection. The age class was chosen because it exhibits high HPV prevalence (24% worldwide[5] and approximately 25% in France[3]). Inclusion of younger women would have raised technical issues because of the requirement for parental consent.

Women are recruited through a social media page, and through posters and leaflets distributed at the universities in Montpellier and at the Montpellier STI screening centre (*Centre Gratuit d'Information de Dépistage et de Diagnostic*, CeGIDD). The composition of the population visiting the CeGIDD has already been documented in an earlier study.[34] In total, the centre is visited by approximately 3,000 women per year, the majority of which are under 25 years of age (80%). Approximately 40% of the attendants report three or more partners over the last 12 months and approximately 50% report using adequate behaviour for prevention against HIV.

Since all participants are healthy, they are referred to as participants rather than patients. As in any longitudinal study, ensuring participant commitment is challenging. To achieve this goal, we have set up a compensation of €40 per visit and an additional €10 in case of a complete follow-up. Furthermore, participants who have answered a sufficient number of questionnaires and brought back a sufficient number of self-samples get a €100 bonus at the end. Overall, a participant performing 12 return visits would gain a total compensation of €650.

Inclusion criteria

Participants are women from 18 to 25 years old living in the metropolitan area of Montpellier. They must be sexually active with at least one new partner over the last 12 months. These criteria are fixed to maximise the incidence of new HPV infections. Participants must be able to and willing to give written informed consent: they must sign an informed consent form, understand the requirements for the study and be affiliated to a French social security scheme (which is a state requirement).

Women cannot be included in the study if they have a history of HPV-associated pathology (genital warts or cervical lesions), if they are pregnant or intending to become pregnant in the coming year, infected by HIV, undergoing (or planning to undergo) intense medical treatment (biotherapy, chemotherapy, immunosuppression), planning on moving outside the Montpellier metropolitan area within the next 18 months, in a dependency or employment with the sponsor or the investigator, if they participated in a clinical trial involving administration of drugs within the last 4 weeks or if they belong to a vulnerable group (e.g., children, adults with physical or mental disabilities).

Design/setting

This study has a longitudinal component aimed at deciphering within-host dynamics and a cross-sectional component aimed at understanding the diversity of HPV infections in young adults in the area of Montpellier, France. The general structure of the study is shown in figure 1.

If a woman fits the main inclusion criteria, she can go through an inclusion visit (V_1) with a physician (gynaecologist or midwife) at the CeGIDD. During this visit, the study investigator presents the study and checks all inclusion criteria before asking the participant to read and sign the informed consent form. Participants then undergo a medical consultation during which a number of samples are collected (see below). They then fill out health and lifestyle questionnaires and are given cotton-flocked swabs for self-sampling at home the next visit, along with instructions on how to fill in weekly questionnaires through an online form (these are performed throughout the study).

An appointment is scheduled 4 weeks later for the results visit (V_2), where the cervical cytology results are communicated. Additional samples are collected and self-sample swabs for home collection are provided.

The next return visits (V_i, where $i>2$) are as follows:

► Participants with a positive HPV detection test (see below), that is, infected by an *Alphapapillomavirus*, at V_1 join the HPV positive (HPV+) arm of the study with return visits scheduled every 2 months.
► Participants with a negative HPV detection test at V_1 join the HPV negative (HPV-) arm with return visits scheduled every 4 months.
► HPV- participants infected by an *Alphapapillomavirus* move to the HPV+ arm.

Intervals between visits are based on earlier results showing that HPV infections last from 9 to 18 months on average depending on the HPV type[6 8–10] and that a total follow-up of 4 months yields results that are difficult to analyse.[22] The longer interval in the HPV- arm is based on the estimated incidence for HPV genital infections in young women which is greater than 30%.[35 36]
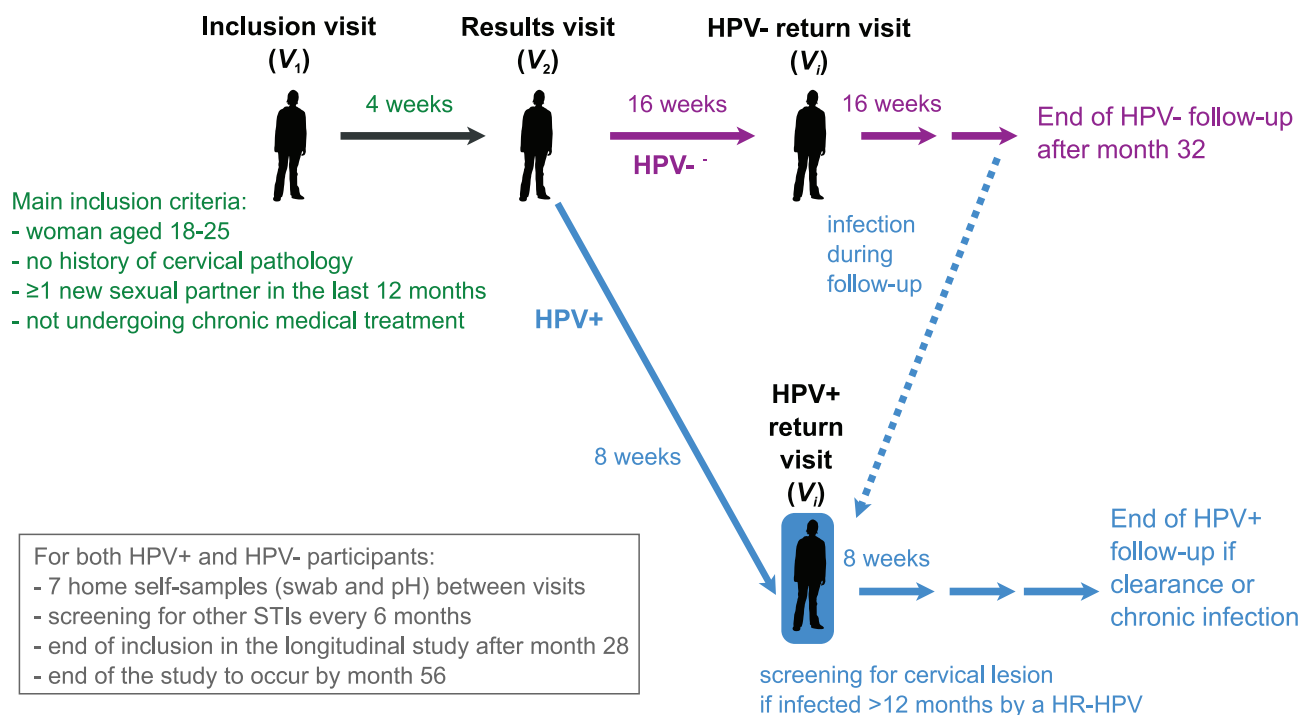
Figure 1 General structure of the PAPCLEAR study. For the longitudinal study, participants have an inclusion visit (V_1), a results visit (V_2) and then return visits (Vi with $i>2$). For the cross-sectional study, participants only have V_1 and V_2. HPV, human papillomavirus; STI, sexually transmitted infection.

Participants in the HPV- arm are followed until month 32 of the study.

Participants in the HPV+ arm are followed until they clear the infection or until they have been infected for 24 months (after which we consider that the infection is persistent). Clearance is defined as being negative at two visits in a row for the first HPV type detected in the follow-up.

In between these visits to the CeGIDD, participants are asked to perform regular (every week for HPV+ and every second week for HPV-) self-samples using vaginal swabs (every week for HPV+ and every second week for HPV-), to measure vaginal pH and to fill in a short questionnaire. Self-samples are stored in the participants' freezer and brought back at every visit.

The study will end with the last HPV+ participant having cleared the infection or been infected for 24 months.

Patients and public involvement
Participants did not play a role in the design of this study and the results of the study will be disseminated to participants who have left the study and to the general public via an email newsletter in French.

Visits
The summary of the visit schedule and of the samples collected at each visit is shown in table 1.

Inclusion visit (V_1)
This visit takes place at the CeGIDD and is scheduled by the Clinical Research Technician (CRT) via phone or email.

Women meet a study investigator, who explains the goals and requirements of the study and checks that the inclusion criteria are met. If so, after a general discussion, the informed consent forms are signed.

A female physician/midwife performs a general exam and then a gynaecological exam during which the following samples are taken:
► Vaginal pH cotton swab (EcoCare).
► Vaginal swab (Copan ESwab) in 1 mL Amies liquid for DNA extraction and microbiota analysis.
► Vaginal swab (Copan ESwab) in 1 mL of RNA preservation medium.
► Ophthalmic sponge (Weck-cel) to collect cervical secretions for cytokines analysis.
► Cervical smear in 20 mL of PreservCyt liquid (Thinprep test) for HPV and HSV assays, and cytology evaluation.

Following the gynaecological consultation, the participant meets with a nurse to measure body temperature, blood pressure and draw 20 mL of blood (a 5 mL tube for single nucleotide polymorphisms (SNP) genotyping, a 10 mL tube for immunophenotyping and a 5 mL tube for HPV antibody titration). For the longitudinal study, the nurse provides the participant with three self-sampling kits, three pH strips, a freezer box to bring back to the next visit and instructions on how to perform the home sampling and store the samples in her personal freezer until the next visit.

If the participant has not been tested for STIs in the last 3 months, the nurse draws an additional blood tube of 5 mL to test for STIs (HIV, hepatitis C virus, hepatitis

Table 1 Summary of the visit schedules and samples take

Participants	Inclusion (V_1) All	Results (V_2) All	Return (V_i, with $i>2$) HPV+	Return (V_i, with $i>2$) HPV-
Time	Day 0	+4 weeks	+8 weeks	+16 weeks
Eligibility	¤			
Consent	¤			
Gynaecological consult	¤	¤	¤	¤
Vaginal pH cotton swab	¤	¤	¤	¤
two vaginal swab samples (Copan ESwab)	¤	¤	¤	¤
One ophthalmological sponge sample	¤		¤	¤
One cervical smear in PreservCyt (cytology)	¤		+	
One cervical smear in PBS (Phosphate Buffered Saline)		¤	+	¤
Blood sampling (HPV antibodies)	¤		¤	
Blood sampling (sequencing)	¤			
Blood sampling (immunophenotyping)	¤	◁		
Other STI detection	★	★	★	★
Questionnaire #1 (inclusion)	¤			
Questionnaire #2 (visit)		¤	¤	¤
Questionnaire #3 (home)	¤	¤	¤	¤
Returning self-sampling samples		¤	¤	¤
Serious adverse event collection		¤	¤	¤

The cross-sectional study only includes the first two columns (V_1 and V_2). ¤ Samples taken at visits. + Participants infected by a HR-HPV for 12 month will have one PBS smear replaced by a Thinprep smear to perform a cytology and check for lesions. ◁ This sample is only taken at the first HPV + visit of a formerly HPV- participant. ★ STI detection will be performed at inclusion unless the participant has been tested within the last 3 months and during the study every 6 months if a new partner has been reported or request. HPV, human papillomavirus; STI, sexually transmitted infection.

B virus) and collects vaginal self-samples for chlamydiae and gonorrhoea detection. Syphilis testing is prescribed to participants who meet the STI clinic's guidelines.

Finally, the participant meets with the CRT to fill in questionnaires #1 (inclusion visit) and #3 (home). The CRT answers any remaining questions, explains how to fill the home questionnaires (#3) and sets an appointment for the Results visit.

Results visit (V_2)
During this visit, the participants are given the result of cervical lesion screening using the liquid cytology (normal, ASCUS, LSIL or HSIL). Participants with a high-grade lesion (HSIL) exit the study and are referred to the gynaecology service of the Centre Hospitalier Universitaire (CHU) of Montpellier.

During this visit, the physician/midwife collects additional samples: two vaginal swabs for DNA and RNA analysis, and a cervical smear in 5 mL of PBS (to confirm HPV status and perform flow cytometry analyses).

The participant fills in questionnaires #2 (for return visits) and #3 (home). An appointment for the next visit is set and swabs for home self-sampling are given.

Return visits (V_i)
These visits only occur in the longitudinal study.

HPV- arm
Participants uninfected by HPV visit the clinic every 4 months until month 26. During these visits, the same samples as in the inclusion visit (V_1) are collected by the physician/midwife except for the cervical smear, which is put in PBS instead of Thinprep.

The nurse only draws blood if a screening test for STIs other than HPV is required. The participant then fills in questionnaires #2 and #3 and an appointment is set for the next visit in 16 weeks.

If an HPV infection is detected in the cervical smear collected during this visit, the participant moves to the HPV+ arm and the CRT contacts the participant to move her appointment forward.

HPV+ arm
Participants infected by HPV visit the clinic every 2 months. They cannot switch arm and will remain in the HPV+ arm until clearance or the end of the study. During the visits, the same samples as in the inclusion visit (V_1) are collected by the physician/midwife except for the cervical smear, which is put in PBS instead of PreservCyt.

The nurse then draws 5 mL of blood for HPV antibody titration. If this is the first HPV+ visit following an HPV- visit, the nurse also draws 10 mL of blood for

immunophenotyping. Finally, if a test for additional STIs is needed, the nurse draws 5 mL of blood and collects vaginal self-samples for STI detection.

Importantly, if the participant has been infected by a HR-HPV for more than 12 months and a cytological analysis has not been performed within the last 12 months, the cervical smear is put in PreservCyt fixation medium, instead of PBS, for cytological analysis (cervical lesion screening).

Finally, the participant fills in questionnaires #2 and #3, receives self-samples for home collection and an appointment is set for the next visit in 8 weeks.

Endpoints

The primary endpoint for the study is the inclusion and follow-up of HPV-infected women in order to describe the kinetics of HPV virus load, and the associated immune response.

Secondary endpoints are the characterisation of the interactions between the course of the infection (eg, duration), the HPV type(s), the abundance and taxonomic diversity of bacteria, fungi and viruses in the vaginal microbiota, human genetics (SNPs) and basal immunological status.

Technical procedures

DNA extraction

DNA extraction from cervical smears will be performed using Nuclisens easyMAG from Biomerieux or an equivalent protocol. For the microbiota analyses, special kits involving physical (via beads) and/or enzymatic breaking of the cellular barrier will be favoured following standard protocols to study the vaginal microbiome,[37] for example, the MagAttract PowerMicrobiome DNA/RNA kit from Qiagen.

HPV detection, typing and quantification

The participants' infection status (HPV+ or HPV-) will be assessed using a DNA enzyme immunoassay (DEIA) detection test based on a PCR of the short SPF_{10} amplicon[38], which detects all *Alphapapillomaviruses* with great sensitivity.

If the DEIA test is positive, HPVs will be typed using the LiPA25 kit which is based on the same SPF_{10}-PCR, and has a lower detection threshold compared with other hybridisation-based typing methods.[39]

The reason for basing the detection on the DEIA rather than the LiPA25 is that some *Alphapapillomavirus* may be detected by DEIA but not genotyped by LiPA and also that the DEIA is more sensitive than the LiPA. If the DEIA is positive and the LiPA25 is negative, typing will be performed by sequencing the product of a PGMY09/11 PCR[40] which targets another region of the HPV genome than the SPF10 PCR.

The quantification of HPV DNA genome copy number in the samples will be performed using the qPCR protocol set up by Micalessi *et al*.[41]

Cytokine titration

Cytokines can be used as markers of immune activation or immunosuppression and can also inform us on which components of the immune system are involved. Cervical sponges are centrifuged after the addition of 175 μL of PBS. Cervical secretions are analysed for a set of 5 to 6 cytokines levels using the Meso Scale Discovery Multiplex ELISA platform which has a low detection threshold and a slowly saturating dose–response curve. Based on earlier results,[42 43] we will first investigate a large panel of 20 cyctokines to choose the most relevant ones for a longitudinal follow-up.

Flow cytometry

Analysing immune cells via flow cytometry is extremely challenging on cells as fragile as the ones from cervical smears. However, several studies suggest that this is feasible.[42–44] Here, we follow the protocol described in Ref [45].

Staining is performed using a Duraclone custom mix targeting CD45, CD3, CD4, CD8, CD16, CD56, CD69, CD161 and TCR$\gamma\delta$. The last marker, Live&Dead tests for cellular viability. Samples are acquired using a Navios flow cytometer (Beckman Coulter, three-laser configuration).

Sequencing

Sequencing will be performed for microbiota profiling. It involves PCR amplification of the V3-V4 region of 16S RNA for bacteria[46] and ITS1 for fungi.[47] We anticipate that the bacteria should belong to the operational taxonomic units (OTU) described in the five community state types found in vaginal communities.[48 49] The virome will also be explored using shotgun sequencing and rolling circle PCR amplification.[50] Human genetics are explored using chip sequencing for SNPs.

Statistical analyses

Times series analyses

The core results of the study will come from the longitudinal follow-up of infected women which will generate time series (ie, a set of values collected from the same individual over time) (figure 2). There will be several time series per individual (virus load, number of immune cells, cytokine and antibody levels). These time series will be used to fit mathematical viral kinetics models that describe the interaction between viruses, host target cells (here, in the case of HPV, keratinocytes) and the immune response. These models are commonly developed for viral infections,[14 51–53] including those caused by HPVs.[54] We anticipate our follow-up to yield adequate data for such a fit based on the estimated duration of HPV infections (9–18 months[6 8–10]). Furthermore, the weekly self-samples allow us to increase the resolution if necessary.

We will use non-linear mixed effect models[55] to jointly analyse time series from all participants. More precisely, we will rely on R packages such as nlme[56] or lme4.[56] Note that, in addition to estimating model parameters (eg, life-expectancy of infected cells or virion production rate
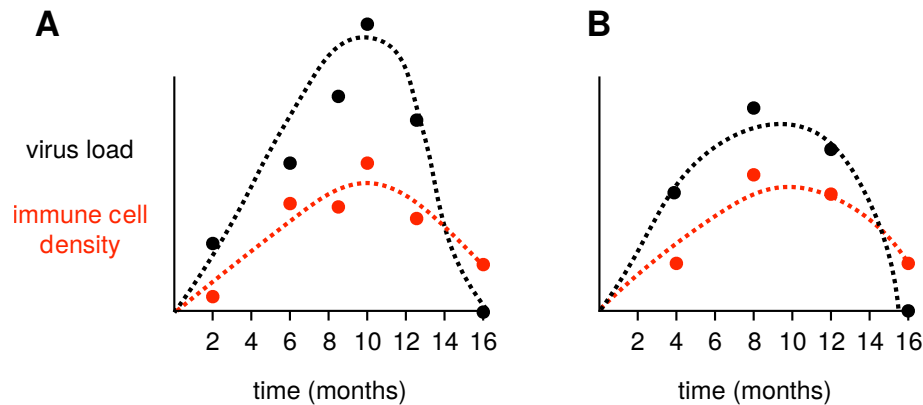
Figure 2 Fitting viral kinetics models to within-host times series. Dashed lines indicate a model fitted using virus load (in black) or immune cells (in red) time series. In panel A, the follow-up is bi-monthly with two missing visits and several delayed visits, whereas in panel B the follow-up is every 4 months without any missing or delayed visits. In spite of missing data this, the situation shown in panel A is clearly the best for inferring parameter values and for fitting the underlying dynamics.

of infected cells), this approach also allows us to compare biological models using statistical tools based on model likelihood such as Akaike Information Criterion. For an example of such analysis in the case of HIV, see Ref.[52]

Microbiota dynamics

The composition of the vaginal microbiota has already been described and shown to exhibit considerably less diversity than the gut microbiota.[48] The dynamics of this microbiota has also been studied and shown to closely follow menstrual cycles.[49]

Using the time series of OTU abundances (measured via 16S RNA sequencing and qPCR) we will infer interaction parameters by assuming an underlying Lotka-Volterra competition model.[57] This work will include time series analysis techniques (eg, auto-correlation or local similarity analysis) and statistical inference methods in order to infer community structure and interactions from the next-generation sequencing (NGS) datasets.[58] Finally, statistical methods from ecology will also be used to study community diversity (eg, diversity indices) and community assembly, such as cluster and ordination analyses.[59]

Genome wide association studies

We will use human SNPs inferred by chip sequencing to look for genetic determinants of key traits (e.g., microbiota composition or HPV infection duration). This is classically done by performing a Genome Wide Association Study (GWAS), which is a complex regression method designed for situations where there are many explanatory variables (here millions of SNPs) for a single trait of interest. GWAS will be performed using classical methods.[60] Earlier GWAS studies have been applied to HPV infections for instance to test for determinants to the ability to seroconvert following infection[61] and cervical cancer (see[62] for a review). Here, our expected sample (n=300 women) is limited but SNPs with large effects have been detected by studies with comparable sizes.[63]

Additional analyses

For all collected variables, descriptive statistics will be calculated according to the level of measurement. For metric variables these statistics can be mean and SD as well as quantiles and more robust statistics.[64] In case of categorical variables group proportions and contingency tables are prepared.

Univariate inferential statistics follow a descriptive analysis. Generally, parametric testing procedures are preferred to non-parametric tests, as the former have higher power. That is why, for metric variables, we will first check whether the data can be assumed to be normally distributed. For normally distributed variables, analysis of variance statistics are done to detect differences between groups. In case of significance, post-hoc analysis (Tukey test) are planned to reveal pairwise differences. If the data are not normally distributed or ordinally scaled, non-parametric analyses will be used. These contain the Kruskal-Wallis test and the Wilcoxon test as a post-hoc test with an appropriate correction of the significance level. Since the cell counts are expected to be small, Fisher's exact test will be performed for contingency tables instead of the asymptotic χ^2 test for categorical variables.

Sample size calculation

The study will enrol a total of n=300 women, with n=150 in a longitudinal study and n=150 in a cross-sectional study. The goal of the longitudinal study is to follow 75 women longitudinally, preferentially before they are infected (see above). For the following calculations, we assumed a high percentage of lost during follow-up (30%).

With 150 enrolments and considering that the prevalence of HPV infection in young women is ≈60% (based on our preliminary data) and 30% of lost to follow-up, we expect to detect (and successfully follow) 63 infections at inclusion (95% CI: 51 to 75, assuming a binomial distribution to calculate the 95% CI).

Among women who are uninfected at the first visit and considering the yearly incidence being close to 30%,[65] we

expect 12 (95% CI 6 to 20) to be infected during the first year of follow-up (still assuming 30% of lost to follow-up).

In the end, with 150 enrolments and assuming a high percentage of lost to follow-up (30%), we expect to successfully follow 75 (95% CI 56 to 95) women infected at different stages of HPV infection: beginning, during and end.

This will be made possible by the probability of transmission of HPV, which is estimated to be ≈90% without condom use and still high with condom use (≈40%).[35]

Finally, regarding potential interference with the HPV vaccines, we do not anticipate any significant problem for two reasons. First, as mentioned above, the vaccine coverage is low in France.[33] Second, and more importantly, the vaccines only target few HPV types, thus leaving open the possibility of infection by dozens of types. Furthermore, studying the kinetics of a non-vaccine HPV type in a vaccinated woman will be extremely informative (eg, to detect any potential cross-reactivity).[66]

To run cross-sectional analyses (especially on the microbiota and human genetics), we will enrol n=150 additional women who will only perform the inclusion and the results visits. This sample size was chosen to reach that of earlier GWAS studies.[62 63]

Trial governance
Sponsor
This study is sponsored by the CHU of Montpellier. The CHU is involved in the implementation of the trial, legal/ethical submissions (see Ethics approval statement) and implementing the clinical database online Case Report Form (eCRF) which is hosted by Ennov-Clinical (ClinSight). The CHU is not involved in the analysis or interpretation of the data. The CHU of Montpellier performs regular quality control assessments. A clinical research assistant will visit the CeGIDD every 4 months to ensure that implementation is in accordance with the protocol. The CHU has taken out insurance from the Société hospitalière d'assurances mutuelles, 18, rue Edouard Rochet-6 9372 Lyon cedex 08 (contract number 138983) through the full research period, covering its own civil liability and that of any agent (clinical or research staff), in accordance with article L.1121–10 of the French Public Health Code.

Scientific committee
The scientific committee comprises the study investigators, clinicians, scientific experts and representatives of the sponsor. The committee meets yearly and is responsible for following research progress, monitoring compliance with good clinical practices and patient safety. It can also decide relevant modification of the protocol. Requests from third parties to access data collected during the study will be evaluated by the committee.

Monitoring
Monitoring is performed during the whole study at CeGIDD according to the sponsor specific SOP. Routine monitoring visits are made by the monitors designated by the sponsor to check compliance with the protocol, the completeness, accuracy and consistency of the data and adherence to GCP. The principal investigator ensures that eCRFs are completed in a timely manner and must allow periodical access to eCRFs, patient records, drug logs and all other study-related documents and materials. The frequency of monitoring visits is determined by factors such as study design and the site enrolment requirements but visits will normally occur at least once every 4 months.

Ethics and dissemination
The protocol has been modified since its initial version and the latest modification was accepted by the Comité de Protection des Personnes on 12 December 2018.

All participants in the study will sign an informed consent form prior to participation.

The results will be published on preprint servers (eg, BioRxiv), peer-reviewed journals, postprint servers (e.g., *Hyper Articles en Ligne*, HAL) and disseminated through conferences.

DISCUSSION
Expected results
Acute infections by HPVs are important to study because vaccination is most effective when performed before the first infection. However, we currently know very little about the early stages of HPV infections. This clinical study will give us an unprecedented level of detail into the natural history of HPV infections in young women. Variations in virus load over time have been studied but in the context of cervical cancer in older women.[67] In addition, we will also describe the nature and the dynamics of the immune response (local immune cells and cytokines, circulating anti-HPV antibodies) and of the vaginal microbiota. Beyond these kinetics, we will also have access to data such as infection clearance or not in 24 months, presence of more than one HPV type or coinfection by other STIs.

These data will be analysed in the light of numerous cofactors. One of the most important will be human genetics, with the sequencing of millions of SNPs. Others will be related to the sexual behaviour (number of partners, contraception methods, sexual practices) and general lifestyle. We, therefore, expect broader insights regarding sexual health in young women.

Practical and operational issues
One of the main practical challenges resides in the analysis of cervical smears by flow cytometry. Indeed, the tissues are known to be fragile, adhesive and auto-fluorescent. Even though standard protocols now exist,[45] they require processing fresh samples in less than 2 hours.

Another potential issue has to do with contamination by HPV DNA between samples, which are frequent in the HPV field due to the robustness of the virions and the sensitivity of the tests. To certify our ability to control for these, we have entered the 2017 GLOBAL HPV DNA Proficiency Panel from the WHO HPV LabNet.[68]

Regarding the enrolment of the participants, we do not expect issues with enrolling 150 women in 28 months for the longitudinal study and 150 for the cross-sectional study. This is due to the number of visitors of the centre who fit the inclusion criteria (more than 3000 per year) and because of earlier high participation rates in the same population ([34] enrolled 1381 participants in 5 months for their study).

Concerning the follow-up, the high incidence rate of HPV can also lead to transient carriage, that is, women who are positive for a type only at a single visit. This has been observed for instance in longitudinal studies with a tight follow-up interval.[22] To control for this, we will run the HPV detection test on the cells from the cervical smear after washing with RPMI.

Trial status

The study began on 1 October 2016 and the first inclusion was on 3 November 2016. On 23 June 2018, 89 participants have been included in the longitudinal study. Inclusions in the longitudinal study will continue until March 2019 and the study is expected to last until August 2021.

Author affiliations
[1]MIVEGEC (UMR 5290 CNRS, IRD, UM), CNRS, Montpellier, France
[2]Center for Free Information, Screening and Diagnosis (CGIDD), Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[3]Department of Obstetrics and Gynaecology, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[4]Department of Research and Innovation (DRI), Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[5]CIBER en Epidemiología y Salud Pública (CIBEResp), Madrid, Spain
[6]Sequencing and Bioinformatics Service, Fundaciónpara el Fomento de la Investigación Sanitaria y Biomédica de laComunidad Valenciana (FISABIO-Salud Pública), Valencia, Spain
[7]Pathogenesis and Control of Chronic Infections, INSERM, CHU, University of Montpellier, Montpellier, France
[8]Department of Virology, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[9]LBPC/PPC- IRMB, CHU de Montpellier and Université de Montpellier, Montpellier, France
[10]Department of Medical Information (DIM), Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[11]Department of Medical microbiology, University of Manitoba, Winnipeg, Manitoba, Canada
[12]Department of pathology and oncobiology, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[13]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA
[14]Department of Infectious and Tropical Diseases, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[15]German Cancer Research Center (DKFZ), Infections and Cancer Epidemiology, Heidelberg, Germany
[16]GIGA-Research, Cellular and molecular immunology, University of Liège, Liège, Belgium

REFERENCES

1. Tota JE, Chevarie-Davis M, Richardson LA, *et al*. Epidemiology and burden of HPV infection and related diseases: implications for prevention strategies. *Prev Med* 2011;53(Suppl 1):S12–21.
2. Chesson HW, Dunne EF, Hariri S, *et al*. The estimated lifetime probability of acquiring human papillomavirus in the United States. *Sex Transm Dis* 2014;41:660–4.
3. Monsonego J, Zerat L, Syrjänen K, *et al*. [Prevalence of genotype-specific HPV infection among women in France: implications for screening and vaccination]. *Gynecol Obstet Fertil* 2013;41:305–13.
4. Brun-Micaleff E, Coffy A, Rey V, *et al*. Cervical cancer screening by cytology and human papillomavirus testing during pregnancy in French women with poor adhesion to regular cervical screening. *J Med Virol* 2014;86:536–45.
5. Bruni L, Diaz M, Castellsagué X, *et al*. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. *J Infect Dis* 2010;202:1789–99.
6. Insinga RP, Dasbach EJ, Elbasha EH, *et al*. Incidence and duration of cervical human papillomavirus 6, 11, 16, and 18 infections in young women: an evaluation from multiple analytic perspectives. *Cancer Epidemiol Biomarkers Prev* 2007;16–709–15.
7. Woodman CB, Collins SI, Young LS. The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer* 2007;7:11–22.
8. Rodríguez AC, Schiffman M, Herrero R, *et al*. Proyecto Epidemiológico Guanacaste Group: Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst* 2008;100:513–7.
9. Trottier H, Mahmud S, Prado JC, *et al*. Type-specific duration of human papillomavirus infection: implications for human papillomavirus screening and vaccination. *J Infect Dis* 2008;197:1436–47.
10. Ramanakumar AV, Naud P, Roteli-Martins CM, *et al*. Incidence and duration of type-specific human papillomavirus infection in high-risk HPV-naïve women: results from the control arm of a phase II HPV-16/18 vaccine trial. *BMJ Open* 2016;6:011371.
11. Houlihan CF, Baisley K, Bravo IG, *et al*. Rapid acquisition of HPV around the time of sexual debut in adolescent girls in Tanzania. *Int J Epidemiol* 2016;45:762–73.
12. Alizon S, Murall C, Bravo I. Why Human Papillomavirus Acute Infections Matter. *Viruses* 2017;9:293.
13. Herrero R, Wacholder S, Rodríguez AC, *et al*. Prevention of persistent human papillomavirus infection by an HPV16/18 vaccine: a community-based randomized clinical trial in Guanacaste, Costa Rica. *Cancer Discov* 2011;1:408–19.
14. Canini L. Perelson, A.S.: Viral kinetic modeling: state of the art. *J Pharmacokinet Pharmacodyn* 2014;41:431–43.
15. Stanley M. Immune responses to human papillomavirus. *Vaccine* 2006;24(S1):S16–S22.

16. Ferenczy A, Franco E. Persistent human papillomavirus infection and cervical neoplasia. *Lancet Oncol* 2002;3:11–16.
17. Hausen zur. H.: Review: Papillomaviruses — to Vaccination and Beyond. *Biochemistry* 2008;73:498–503.
18. Einstein MH, Schiller JT, Viscidi RP, et al. Clinician's guide to human papillomavirus immunology: knowns and unknowns. *Lancet Infect Dis* 2009;9:347–56.
19. Van Hede D, Langers I, Delvenne P, et al. Origin and immunoescape of uterine cervical cancer. *Presse Med* 2014:43:413–21.
20. Stanley M. Immunology of HPV Infection. *Curr Obstet Gynecol Rep* 2015;4:195–200.
21. Gao W, Weng J, Gao Y, et al. Comparison of the vaginal microbiota diversity of women with and without human papillomavirus infection: a cross-sectional study. *BMC Infect Dis* 2013;13:271.
22. Brotman RM, Shardell MD, Gajer P, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis* 2014;210:1723–33.
23. Koutsky LA, Ault KA, Wheeler CM, et al. Proof of Principle Study Investigators: A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med* 2002;347:1645–51.
24. Riethmuller D, Jacquard AC, Lacau St Guily J, et al. Potential impact of a nonavalent HPV vaccine on the occurrence of HPV-related diseases in France. *BMC Public Health* 2015;15:453.
25. Joura EA, Giuliano AR, Iversen O-E, et al. Broad Spectrum HPV Vaccine Study: A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med* 2015;372:711–23.
26. Murall CL, Bauch CT, Day T. Could the human papillomavirus vaccines drive virulence evolution? *Proc Biol Sci* 2015;282:20141069.
27. Alizon S, Méthot PO. Reconciling Pasteur and Darwin to control infectious diseases. *PLoS Biol* 2018;16:e2003815.
28. Moscicki A-B, Ma Y, Wibbelsman C, et al. Shiboski, S.: Rate of and Risks for Regression of CIN-2 in adolescents and young women. *Obstet Gynecol* 2010;116:1373–80.
29. Buck Jr. H.W. Warts (genital). *BMJ Clin Evid* 2015:1602.
30. Herrero R, González P, Markowitz LE. Present status of human papillomavirus vaccine development and implementation. *Lancet Oncol* 2015;16:e206–e216.
31. Maver PJ, Poljak M. Progress in prophylactic human papillomavirus (HPV) vaccination in 2016: A literature review. *Vaccine* 2018;36:5416–23.
32. Fagot JP, Boutrelle A, Ricordeau P, et al. HPV vaccination in France: uptake, costs and issues for the National Health Insurance. *Vaccine* 2011;29:3610–6.
33. Ben Hadj Yahia M-B, Dervaux B, Duport N, et al. *Vaccination contre les infections à papilloamvirus*. Paris, France: Technical report, Haut Conseil de la Santé Publique, 2014.
34. Clarivet B, Picot E, Marchandin H, et al. Prevalence of Chlamydia trachomatis, Neisseria gonorrhoeae and Mycoplasma genitalium in asymptomatic patients under 30 years of age screened in a French sexually transmitted infections clinic. *Eur J Dermatol* 2014;24:611–6.
35. Winer RL, Hughes JP, Feng Q, et al. Condom use and the risk of genital human papillomavirus infection in young women. *N Engl J Med* 2006;354:2645–54.
36. Winer RL, Hughes JP, Feng Q, et al. Incident Detection of High-Risk Human Papillomavirus Infections in a Cohort of High-Risk Women Aged 25-65 Years. *J Infect Dis* 2016;214:665–75.
37. Ravel J, Brotman RM, Gajer P, et al. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome* 2013;1:29.
38. Kleter B, van Doorn LJ, ter Schegget J, et al. Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. *Am J Pathol* 1998;153:1731–9.
39. Geraets DT, Struijk L, Kleter B, et alColau, B.: The original SPF10 LiPA25 algorithm is more sensitive and suitable for epidemiologic HPV research than the SPF10 INNO-LiPA Extra. *J Virol Meth* 2015:216:22–9.
40. Gravitt PE, Peyton CL, Alessi TQ, et al. Improved amplification of genital human papillomaviruses. *J Clin Microbiol* 2000;38:357–61.
41. Micalessi IM, Boulet GAV, Bogers JJ, et al. High-throughput detection, genotyping and quantification of the human papillomavirus using real-time PCR. *Clin Chem Lab Med* 2012;50:655–61.
42. Hunter PJ, Sheikh S, David AL, et al. Cervical leukocytes and spontaneous preterm birth. *J Reprod Immunol* 2016;113:42–9.
43. Shannon B, Yi TJ, Perusini S, et al. Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. *Mucosal Immunol* 2017;10:1310–9.
44. Lajoie J, Juno J, Burgener A, et al. A distinct cytokine and chemokine profile at the genital mucosa is associated with HIV-1 protection among HIV-exposed seronegative commercial sex workers. *Mucosal Immunol* 2012;5:277–87.
45. Juno JA, Boily-Larouche G, Lajoie J, et al. Collection, isolation, and flow cytometric analysis of human endocervical samples. *J Vis Exp* 2014;89:51906.
46. Frank JA, Reich CI, Sharma S, et al. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 2008;74:2461–70.
47. Findley K, Oh J, Yang J, et al. NIH Intramural Sequencing Center Comparative Sequencing Program, Kong, H.H., Segre, J.A.: Topographic diversity of fungal and bacterial communities in human skin. *Nature* 2013;498:367–70.
48. Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1:4680–7.
49. Gajer P, Brotman RM, Bai G, et al. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 2012;4:132ra52–52.
50. Johne R, Müller H, Rector A, et al. Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends Microbiol* 2009;17:205–11.
51. Nowak MA, May R. *M.: Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford, USA: Oxford University Press, 2000.
52. Stafford MA, Corey L, Cao Y, et al. Modeling plasma virus concentration during primary HIV infection. *J Theor Biol* 2000;203:285–301.
53. Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol* 2002;2:28–36.
54. Murall CL, Jackson R, Zehbe I, et al. Epithelial stratification shapes infection dynamics. *PLoS Comput Biol* 2019;15:e1006646.
55. Steimer JL, Vozeh S, Racine Poon A, et al. *Handbook of Experimental Pharmacology*. Balant P, O'Neil, R.: The population approach: rationale, methods and applications in clinical pharmacology and drug development. . Berlin : Springer, 1994:110. 405–51.
56. Bates D, Mächler M, Bolker B, et al. Walker, S.: Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015;67.
57. Bucci V, Tzen B, Li N, et al. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. *Genome Biol* 2016;17:121.
58. Faust K, Lahti L, Gonze D, et al. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr Opin Microbiol* 2015;25:56–66.
59. Fox GA, Negrete-Yankelevich S, Sosa V. *J.: Ecological Statistics: Contemporary Theory and Application. Oxford University Press*. Oxford, USA: Oxford University Press, 2015.
60. Shi Y, Li L, Hu Z, et al. A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nat Genet* 2013;45:918–22.
61. Chen D, Gaborieau V, Zhao Y, et al. A systematic investigation of the contribution of genetic variation within the MHC region to HPV seropositivity. *Hum Mol Genet* 2015;24:2681–8.
62. Chen D, Gyllensten U. Lessons and implications from association studies and post-GWAS analyses of cervical cancer. *Trends Genet* 2015;31:41–54.
63. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007;317:944–7.
64. Huber PJ. The 1972 Wald Lecture Robust Statistics: A Review. *The Annals of Mathematical Statistics* 1972;43:1041–67.
65. Winer RL, Feng Q, Hughes JP, et al. Risk of female human papillomavirus acquisition associated with first male sex partner. *J Infect Dis* 2008;197:279–82.
66. Herrero R. Human papillomavirus (HPV) vaccines: limited cross-protection against additional HPV types. *J Infect Dis* 2009;199:919–22.
67. Depuydt CE, Verstraete L, Berth M, et al. Human Papillomavirus Positivity in Women Undergoing Intrauterine Insemination Has a Negative Effect on Pregnancy Rates. *Gynecol Obstet Invest* 2016;81:41–6.
68. HPV LabNet WHO. World Health Organization. 2008 http://www.who.int/biologicals/areas/human_papillomavirus/WHO_HPV_LabNet/en/.