Y Chromosome and HIV DNA Detection in Vaginal Swabs as Biomarkers of Semen and HIV Exposure in Women

Kerri J. Penrose, MS,* Barbara A. Richardson, PhD,† Guillaume Besson, PhD,* Charlene S. Dezzutti, PhD,*‡ Betsy C. Herold, MD,§ Salim S. Abdool Karim, PhD,¶ John W. Mellors, MD,* and Urvi M. Parikh, PhD,* and on behalf of the MTN Biomedical Sciences Working Group and the HPTN 035 Protocol Team

Background: The inability to quantify sexual exposure to HIV limits the power of HIV prevention trials of vaccines, microbicides, and preexposure prophylaxis in women. We investigated the detection of HIV-1 and Y chromosomal (Yc) DNA in vaginal swabs from 83 participants in the HPTN 035 microbicide trial as biomarkers of HIV exposure and unprotected sexual activity.

Methods: One hundred forty-three vaginal swabs from 85 women were evaluated for the presence of Yc DNA (Quantifier Duo DNA quantification kit; Applied Biosystems) and total HIV-1 DNA (single-copy in-house quantitative polymerase chain reaction assay). Y DNA was paired with self-reported behavioral data with regard to recent coitus (≤ 1 week before collection) and condom usage (100% vs. <100% compliance).

Results: Yc DNA was detected in 62 (43%) of 143 swabs. For the 126 visits at which both behavioral data and swabs were collected, Yc DNA was significantly more frequent in women reporting less than 100% condom usage (odds ratio, 10.69; 95% confidence interval, 2.27–50.32; \( P = 0.003 \)). Notably, 27 (33%) of 83 swabs from women reporting 100% condom usage were positive for Yc DNA. HIV DNA was only detected in swabs collected postseroconversion.

Conclusions: The use of Yc DNA in HIV prevention trials could reliably identify subgroups of women who have unprotected sexual activity and could provide valuable exposure-based estimates of efficacy.

A major limitation of current HIV prevention trial design is the lack of an accurate and sensitive means to measure exposure to HIV.1,2 The evaluation of risk behavior using self-reported condom use and frequency of coitus is unreliable because of miscommunication between interviewer and interviewee, lack of understanding of the questions asked, reporting behavior according to perceived expectations, inability to recall experience, and blatant misreporting.3,4 In addition, self-report of condom use frequency does not capture risk associated with improper condom use or accidental condom breakage. Despite greater than 80% reported condom use in VOICE, FEM-PrEP, and the TDF2 study, pregnancy and HIV incidence rates remained higher than expected.5,7 A cross-sectional study of 910 women in Zimbabwe found that only 52% of participants who tested positive for prostate-specific antigen (PSA) in vaginal swabs reported unprotected sex during the previous 2 days. Audio computer-assisted self-interview technology did not generate significantly different responses about unprotected intercourse compared with face-to-face interview.8

The detection of semen could provide an unbiased measure of unprotected sex in participants of HIV prevention trials. However, the limited sensitivity of the most commonly used biomarker, PSA, makes it impractical for use in clinical studies where swab collection could occur hours to weeks after intercourse. Levels of PSA decline 10-fold by 3 hours postexposure and are undetectable by 24 to 48 hours postexposure.9–11 Rapid stain identification of human semen that detects the presence of semenogelin has also been used in trials to indicate that a woman has been exposed to ejaculate in the previous 48 hours, but it is 10-fold less sensitive than quantitative polymerase chain reaction (PCR) methods to detect Y chromosomal (Yc) DNA.12,13 Despite heterogeneity in both initial deposit of Yc DNA and rate of decline of Yc DNA signal, Yc has the advantage of detectability up to 15 postcoital days from self-collected vaginal swabs without impact from menses on the rate of decay.14,15 Y chromosome DNA is not detected in women using condoms correctly, as demonstrated by a study that showed that only 5 of 56 women had positive Yc DNA results after condom use after a 14-day abstinence period, and the 5 detections were associated with receptive oral sex and digital penetration.16,17

The detection of HIV in genital samples from HIV-negative women could more directly assess HIV infection risk. In a study of cervical dysplasia in US women, HIV-1 env and gag glycoproteins were identified in cervicovaginal lavage samples from women who were confirmed to be HIV negative by serology.18
However, partner HIV status was not known, and the linkage of \textit{env} and \textit{gag} detection with future seroconversion was not verified. HIV-1 viral RNA and proviral HIV-1 DNA sequences can be detected in seminal plasma and nonspermatozoal mononuclear cells in HIV-infected men throughout successful long-term highly active antiretroviral therapy, whereas plasma HIV-1 viral RNA levels remain undetectable.\textsuperscript{19-22} The detection of HIV DNA in genital samples from HIV-negative women has not been studied as a marker of HIV infection risk.

We therefore examined the frequency of Yc DNA and HIV DNA detection in vaginal swabs collected in HIV seroconverters (both preseroconversion and postseroconversion) and nonseroconverters from the HPTN 035 study, using highly sensitive quantitative PCR assays with detection limits of a single copy.

\section*{MATERIALS AND METHODS}

\subsection*{Study Population}

HPTN 035 was a phase II/IIb safety and effectiveness study of the vaginal microbicides BufferGel and 0.5\% PRO2000/5 gel for the prevention of HIV-1 infection in women, conducted from February 2005 through September 2008 (NCT00074425). All participants provided informed consent for swab collection and future testing. The population demographic characteristics, protocol, and trial results are described elsewhere.\textsuperscript{23} Starting in 2008, vaginal swab specimens were collected from participants during each quarterly pelvic examination by applying a Dacron swab to the posterior fornix of the vagina until the tip was saturated with fluid, then placing the swab in a cryovial containing 400 \mu L of phosphate-buffered saline. The cryovials were stored at \textdegree 80°C at the sites and shipped to the MTN Network Laboratory after the primary study results were available. The current study evaluated a case-control subset of swabs collected from women at sites in Zimbabwe (Harare and Chitungwiza), South Africa (Hlabisa and Durban), and Malawi (Blantyre and Lilongwe) at a 1:3 ratio (seroconverters/nonseroconverters). Swabs from seroconverters and nonseroconverters were collected and stored in 89 of 89 wells with a predicted copy number of 2 or higher. Swabs from participants postseroconversion were collected a median of 21 days after detection of seroconversion (range, 5–124 days). Seroconverters were not taking antiretroviral therapy at the time of swab collection. The operator performing the assays was blinded to the subgroup to which the participant belonged.

\subsection*{Nucleic Acid Extraction}

Swabs were processed to isolate the cell pellet as described previously.\textsuperscript{24} Total nucleic acid was extracted by incubating the vaginal swab cell pellet in 2 mg/mL Proteinase K solution (Applied Biosystems) for 30 minutes at 55°C. Guanidinium isothiocyanate (Sigma) and glycogen (Roche) were added to final concentrations of 4.58 M and 0.47 mg/mL, respectively, and incubated at room temperature for 30 minutes. Nucleic acids were precipitated by centrifugation at 15,000 \times g in the presence of a nearly equal volume of isopropanol. Nucleic acid pellets were washed repetitively with 70\% ethanol and air dried before suspending in 5 mM Tris, pH 8. One third of the sample was used as template in the Quantifiler Duo assay, one third was used for testing HIV-1 DNA, and one third was stored.

\subsection*{Detection of HIV-1 DNA}

HIV-1 DNA was detected using a modified version of the single copy assay with primers targeted to a conserved region in the integrase gene (iSCA).\textsuperscript{25} Briefly, 10 \mu L of extracted DNA was diluted with 20 \mu L of 5 mM Tris pH 8 and run in triplicate in a reaction containing 1 \times Roche LightCycler 480 probes master mix, 400 nM of primers iSCA-F (5'-TTG GGA AAG GAC CAG CCA A-3') and iSCA-r (5'-CCT GCC ATC TGT TTT CCA-3') and 200 nM Taqman probe (5'-6FAM AAA GGT GAA GGG GCA GTA GTA ATA C BHQ_1-3'). DNA was amplified at 95°C for 5 minutes followed by 45 cycles of 95°C for 15 seconds and 60°C for 1 minute on a LightCycler 480 (Roche). This assay can detect HIV-1 DNA at a single copy per well, as verified by Poisson distribution statistics in dilution experiments of purified HIV-1 DNA target in mock swab cell pellets prepared using A431 cells spiked with a known quantity of purified HIV-1 DNA and human semen. False-positive results were observed in 0 of 111 reactions using Tris buffer as a no template control. Each sample was run in triplicate and considered “undetectable” if all 3 reactions had no amplification; “detected, quantifiable” if at least 2 of 3 reactions had 1 or more copies detected per reaction; and “detected, not quantifiable” if only 1 of 3 reactions had 1 or more copies detected.

\subsection*{Detection of Yc DNA}

Y chromosome and total human DNA was detected using the commercially available Quantifiler Duo kit (Applied Biosystems). Reactions were prepared according to the manufacturer’s guidelines with the modification of increasing the total number of PCR cycles from 40 to 55 to allow late amplification (\geq 35 cycles) to reach the plateau phase of PCR, and to observe if any cases of nonspecific amplification occurred after 40 cycles. Despite these adjustments, all positive wells had cycle threshold values of 40 cycles or below, consistent with the kit manufacturer’s guidelines. Wells with cycle threshold values of 40 to 41 that had exponential amplification were included as positive only if the calculated copy number was 1 copy or above. No cases of false amplification between 41 and 55 cycles occurred. The assay can detect Yc DNA at a single copy per well, as verified by Poisson distribution statistics in dilution experiments of semen in mock swab cell pellets. False-positive Yc DNA was observed in 0 of 135 reactions using Tris buffer as a no template control and 0 of 65 reactions using A431 female epithelial cell pellets as a negative control. Positive detection using dilutions of semen was obtained in 89 of 89 wells with a predicted copy number of 2 or higher. Additional sensitivity and specificity information can be found in the Quantifiler Duo user’s manual. Each test sample was run in replicates of 5 and considered “undetectable” if all 5 reactions had no amplification; “detected, quantifiable” if at least 3 of 5 reactions had 1 or more copies detected per reaction; and “detected, not quantifiable” if only 1 or 2 reactions had at least 1 copy detected per reaction. Total Human DNA, measured through the RPPH1 target as part of the Quantifiler Duo kit, was used to calculate approximate total cell numbers in each swab cell pellet and quantified according to kit manufacturer’s instructions.

\subsection*{Behavioral Data Collection and Analysis}

Participants were queried on gel and condom use during the last coital act and during all coital acts in the last 7 days at each quarterly visit, and data were collected by participant self-report.\textsuperscript{23} Detected, but not quantifiable Yc DNA samples were excluded from the analysis of Yc DNA detection with self-reported condom use because of the inferior quality of these samples.

\subsection*{Statistical Analysis}

SPSS Version 20 (IBM Corp.) was used for all analyses. Conditional logistic regression was used to compare the odds of being a seroconverter among those with detectable Yc DNA and those with no detectable Yc DNA. The association between Yc DNA detection (vs. no detection) and reported sexual activity was assessed using generalized estimating equation models with a binary link, robust errors, and independent correlation structure.
RESULTS

Swab Collection and Cell Recovery

In the HPTN 035 study, 3524 vaginal swab specimens were collected from 2031 women from the African sites. A randomly selected subset of 110 swabs from 65 women who remained HIV negative throughout the study (controls) and 33 swabs from 18 women who seroconverted postenrollment (cases) were evaluated for the presence of HIV and Yc DNA (Fig. 1). Of 33 swabs from seroconverters, 10 were collected, whereas the women were still HIV negative (before seroconversion) and 23 were collected after the participant was confirmed as having seroconverted. Cell recovery from swabs as assessed through total human DNA detection was a median of 3.0 × 10^5 cells/swab (interquartile range, 1.4 × 10^5 to 6.1 × 10^6 cells/swab). Receiver operating characteristic curve analysis demonstrated no minimal cell pellet size requirement for the detection of Yc DNA. The minimum cell number needed for HIV DNA recovery could not be determined because of small sample size (data not shown).

HIV-1 DNA Detection

HIV-1 DNA was only detected in 10 of 23 swabs collected postseroconversion. Five had quantifiable HIV-1 DNA with a median of 141 copies/sample and a range of 55 to 1593 copies/sample from swabs that were collected postseroconversion, and 5 had HIV-1 DNA that was detected but not quantifiable (Table 1). HIV-1 DNA could not be detected in the 10 samples collected before seroconversion or in the 110 samples from women who remained HIV-1 negative throughout the study.

Yc DNA Detection

All 143 swabs were evaluated for the presence of Yc DNA as a measure of unprotected sexual activity. Y chromosome DNA was detected in 62 of 143 (43%) swabs. Forty-four of 62 samples had quantifiable Yc DNA, with a median of 425 copies/sample and a range of 20 to 11926 copies/sample. Y chromosome DNA was detected but not quantifiable in 18 of 62 (29%) samples (Table 2). Using conditional logistic regression with serostatus as the outcome, no significant differences in Yc DNA detection were found between the HIV seroconverter cases (5/33 [15%]) and HIV nonseroconverter controls (39/110 [35%]) in samples that had quantifiable Yc DNA (P = 0.1). This difference remained non-significant when samples collected before seroconversion were included as cases (odds ratio [OR], 0.48; P = 0.3). There was no correlation between Yc and HIV-1 DNA detection in the sample set.

Association of Yc DNA Detection With Self-Reported Condom Use

Participant responses to frequency of coitus and condom use were also collected at 126 of the 143 visits at which a swab sample was collected. There were 27 (21%) reports of no vaginal sex in the week before sample collection, 83 (66%) reports of vaginal sex in the past week with 100% condom use, and 16 (13%)
report of seminal viral DNA (from an infected male partner) in a vaginal swab sample collected from participants reporting no coitus had Yc DNA (median, 98 copies), compared with 27 (33%) of 83 samples with Yc DNA (median, 518 copies) from participants reporting vaginal sex in the past week with 100% condom use and 9 (56%) of 16 samples from participants with Yc DNA (median, 568 copies) reporting vaginal sex in the past week with less than 100% condom use. The proportion of samples with Yc DNA was significantly higher (OR, 10.69; \( P = 0.003 \)) among those participants with inconsistent condom use compared with those reporting no coitus in the past week (Table 3). Reporting of feminine hygiene practices did not correlate with Yc DNA detection (data not shown).

**DISCUSSION**

The advent of new, highly sensitive quantitative PCR technologies for nucleic acid detection and quantification in genital specimens enabled us to evaluate the feasibility of using HIV and Yc DNA detection as a biomarker for risk behavior in HIV prevention clinical trials. We quantified HIV and Yc DNA from the same vaginal swab and stratified the results against behavioral data on condom use from women both preseroconversion and from women who remained seronegative at study termination.

Despite having an assay with sensitivity down to a single copy of HIV-1 DNA per PCR reaction, we did not find evidence of HIV-1 in specimens from seronegative women, including from those women who eventually seroconverted. The inability to detect seminal viral DNA (from an infected male partner) in a vaginal swab sample could have been influenced by sample quality, timing of swab sample collection from last coital act, and lack of frequent exposure to HIV. Furthermore, CD4⁺ lymphocytes only comprise 2% of the total cell number in an average semen sample and only 0.1% of CD4⁺ T cells carry provirus in an HIV-infected individual. The number of absolute CD4⁺ lymphocytes in semen can also depend on the health of the individual and decrease dramatically in viremic patients. Of note, the maximum number of Yc DNA copies from any swab in our study was 1000 copies/well, with the majority (134/143 [94%]) having Yc copies of 100 or less, meaning that only a small fraction of semen was being detected. Because so little of Yc DNA is being recovered from an ejaculate, it is very unlikely that rare HIV-infected cells would have been detected.

We also did not observe a difference in frequency of Yc DNA detection in seroconverters compared with nonseroconverters (15% vs. 35%), and the number of copies of Yc detected did not predict risk of seroconversion. Walsh et al. showed that there was a significant difference in levels of PSA and sperm counts with different types of risk exposures such as condom breakage, but a similar analysis using Yc DNA has not been done. There are several explanations for Yc DNA not predicting seroconversion including variation in donor sperm count, variable time between last coitus and sample collection, noncoital exposure (e.g. digital), and limited sampling of seroconverters. Larger observations over longer periods may reveal an association of Yc DNA exposure and risk of HIV or other sexually transmitted infections.

Our data do provide further evidence that Yc chromosomal DNA detection serves as a reliable biomarker to monitor sexual activity. A significantly higher proportion of women (\( P = 0.003 \)) reporting unprotected sex in the past week had detectable Yc DNA compared with women reporting no vaginal sex in past week. Interestingly, 33% of women reporting 100% condom use had detectable Yc DNA, with copy numbers at similar levels to those reporting unprotected sex with less than 100% condom use. The proportion of participants with Yc DNA (median, 568 copies) reporting vaginal sex in the past week compared with those reporting no coitus had Yc DNA (median, 518 copies), with a significant difference in levels of PSA and sperm counts with different types of risk exposures such as condom breakage, but a similar analysis using Yc DNA has not been done. There are several explanations for Yc DNA not predicting seroconversion including variation in donor sperm count, variable time between last coitus and sample collection, noncoital exposure (e.g. digital), and limited sampling of seroconverters. Larger observations over longer periods may reveal an association of Yc DNA exposure and risk of HIV or other sexually transmitted infections.

Our data do provide further evidence that Yc chromosomal DNA detection serves as a reliable biomarker to monitor sexual activity. A significantly higher proportion of women (\( P = 0.003 \)) reporting unprotected sex in the past week had detectable Yc DNA compared with women reporting no vaginal sex in past week. Interestingly, 33% of women reporting 100% condom use had detectable Yc DNA, with copy numbers at similar levels to those women reporting less than 100% condom use (518 vs. 568 copies), suggesting that condom usage was overreported in this study or that exposure occurred before the use of a condom. The behavioral questionnaire was administered through face-to-face interviews, which may contribute to inconsistent reporting particularly for

**TABLE 1. Detection of HIV-1 DNA in a Subset of Vaginal Swabs From HPTN 035**

<table>
<thead>
<tr>
<th>HIV-1 DNA*</th>
<th>Seronegative† (n = 110, n (%))</th>
<th>Preseroconversion‡ (n = 10, n (%))</th>
<th>Postseroconversion (n = 23)</th>
<th>n (%)</th>
<th>Median (range) copies/sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undetectable</td>
<td>110 (100)</td>
<td>10 (100)</td>
<td>13 (50)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Detected, quantifiable</td>
<td>0</td>
<td>0</td>
<td>5 (25)</td>
<td>141 (55–1593)</td>
<td></td>
</tr>
<tr>
<td>Detected, not quantifiable</td>
<td>0</td>
<td>0</td>
<td>5 (25)</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Each sample was run in triplicate and considered “undetectable” if all 3 reactions had no amplification; “detected, quantifiable” if at least 2/3 reactions had 1 or more copies detected per reaction; and “detected, not quantifiable” if only 1 of 3 reactions had 1 or more copies detected per reaction.

†Swab sample collected from participant who remained seronegative at study termination.

‡Swab sample collected from participant who was seronegative at time of sample collection, but seropositive at study termination.

**TABLE 2. Detection of Yc DNA in a subset of vaginal swabs from HPTN 035**

<table>
<thead>
<tr>
<th>Yc DNA*</th>
<th>Seronegative† (n = 110)</th>
<th>Seropositive‡ (n = 33)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n (%)</td>
<td>Median (range) copies/sample</td>
<td>n (%)</td>
</tr>
<tr>
<td>Undetectable</td>
<td>62 (56)</td>
<td>—</td>
</tr>
<tr>
<td>Detected, quantifiable</td>
<td>39 (35)</td>
<td>413 (16–7003)</td>
</tr>
<tr>
<td>Detected, not quantifiable</td>
<td>9 (8)</td>
<td>—</td>
</tr>
</tbody>
</table>

*Each sample was run in replicates of 5 and considered “undetectable” if all 5 reactions had no amplification; “detected, quantifiable” if at least 3 of 5 reactions had 1 or more copies detected per reaction; and “detected, not quantifiable” if only 1 or 2 reactions had at least 1 copy detected per reaction.

†Swab sample collected from participant who remained seronegative at study termination.

‡Swab sample collected from participant who was seronegative at time of sample collection, but seropositive at study termination.
sensitive topics. Our data show promise for the use of Yc DNA as an objective measure for condom use. Further study is needed to determine if level of risk or type of risk behavior can be correlated with number of Yc copies detected.

One limitation of this study is that testing for HIV and Yc DNA was done retrospectively on stored swabs from quarterly sampling, where timing of swab collection after coitus and the HIV infection status of the male partner was not known. Testing a larger number of prerescreening swabs or swabs from serodiscordant couples could provide further insight into the feasibility of using HIV or Yc DNA as a biomarker. Self-collected samples could provide the best timing for detecting HIV exposure, but would rely on the ability and willingness of participants to collect high-quality samples. More frequent sample collection could provide risk information to statisticians for refined secondary analysis of clinical trial data in populations of highest risk. Modifying the assay for HIV detection to include HIV-1 RNA or total nucleic acid detection may improve sensitivity.

In summary, we demonstrated that by using highly sensitive quantitative PCR assays, Yc DNA and total HIV-1 DNA can be detected down to a single copy in vaginal swab samples. Y chromosome DNA detection is more frequent among women reporting less than 100% condom use with coitus but can also be detected in a third of women reporting 100% condom usage. These results suggest that Yc DNA detection in vaginal fluids could refine assessments of HIV risk and efficacy of preventive strategies.

REFERENCES

TABLE 3. Correlation of Yc DNA with Self-Reported Condom Usage

<table>
<thead>
<tr>
<th></th>
<th>Yc DNA</th>
<th>Univariate Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Detected,</td>
<td>Detected,</td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>Quantifiable,</td>
</tr>
<tr>
<td>No vaginal sex in past week (n = 27)</td>
<td>19 (70)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Vaginal sex in past week with 100% condom use (n = 83)</td>
<td>44 (53)</td>
<td>27 (33)</td>
</tr>
<tr>
<td>Vaginal sex in past week with &lt;100% condom use (n = 16)</td>
<td>4 (25)</td>
<td>9 (56)</td>
</tr>
</tbody>
</table>

CI indicates confidence interval

Samples that were detected but not quantifiable were excluded from the analysis.


