

Signature sequence validation of human papillomavirus type 16 (HPV-16) in clinical specimens

Sin Hang Lee, Veronica S Vigliotti, Suri Pappu

Milford Hospital, Milford, Connecticut, USA

Correspondence to

Dr Sin Hang Lee, Department of Pathology, Milford Hospital, 300 Seaside Avenue, Milford, CT 06460, USA; sinhang.lee@milfordhospital.org

Accepted 2 October 2009
Published Online First
26 October 2009

ABSTRACT

Aims Persistent infection indicated by detection of human papillomavirus 16 (HPV-16) on repeat testing over a period of time poses the greatest cervical cancer risk. However, variants of HPV-16, HPV-31 and HPV-33 may share several short sequence homologies in the hypervariable L1 gene commonly targeted for HPV genotyping. The purpose of this study was to introduce a robust laboratory procedure to validate HPV-16 detected in clinical specimens, using the GenBank sequence database as the standard reference for genotyping.

Methods A nested PCR with two pairs of consensus primers was used to amplify the HPV DNA released in crude proteinase K digest of the cervicovaginal cells in liquid-based Papanicolaou cytology specimens. The positive nested PCR products were used for direct automated DNA sequencing.

Results A 48-base sequence downstream of the GP5+ priming site, or a 34-base sequence upstream thereof, was needed for unequivocal validation of an HPV-16 isolate. Selection of a 45-base, or shorter, sequence immediately downstream of the GP5+ site for Basic Local Alignment Search Tool sequence analysis invariably led to ambiguous genotyping results.

Conclusions DNA sequence analysis may be used for differential genotyping of HPV-16, HPV-31 and HPV-33 in clinical specimens. However, selection of the signature sequence for Basic Local Alignment Search Tool algorithms is crucial to distinguish certain HPV-16 variants from other closely related HPV genotypes.

INTRODUCTION

Almost all cervical cancers or precancerous lesions harbour a human papillomavirus (HPV) long before a cytohistological diagnosis is established. The risk of developing precancer or cancer is greatest in women positive for the same 'high-risk' genotype of HPV on repeat testing over a period of time as an indication of persistent of HPV infection.^{1–4} Therefore, accurate HPV genotyping may play an important role in clinical management. The College of American Pathologists has urged rigorous validation for all HPV assays performed in clinical laboratories.^{5,6}

Since HPV-16 is consistently detected in 50% of cervical cancers and cervical intraepithelial neoplasia 3 (CIN3) lesions,^{7,8} a reliable method to detect and validate HPV-16 in Pap cytology specimens would be a valuable tool to follow the patients with persistent HPV-16 infection before a precancer Pap cytology becomes obvious.

Accurate genotyping of HPV, especially of HPV-16, is challenging in clinical pathology. Two commercial HPV genotyping kits: the SPF10-line

probe assay (SPF10-LiPA; DDL Diagnostic Laboratory, Voorburg, The Netherlands) and the line blot assay (Roche Molecular Systems, Pleasanton, California, USA), have been evaluated against each other with only poor to intermediate agreement in genotyping results.⁹ When the same specimens were tested in parallel for comparison, the SPF10-LiPA detected more HPV-31 than HPV-16, while the line blot assay detected more HPV-16 than HPV-31.^{9,10} These results indicate that some HPV-16 and HPV-31 isolates from clinical samples might have been classified as one or the other interchangeably. The direct automated Sanger sequencing method^{11–21} may validate type-specific signature DNA sequences for HPV-16 differential genotyping.

MATERIALS AND METHODS

A total of 2740 alcohol-preserved Cytoc or Surepath liquid-based specimens submitted by the gynaecologists practicing in southern Connecticut for HPV PCR testing were included in this analysis. These HPV tests were primarily ordered by the physicians affiliated with Milford Hospital for their patients 30 years and older (up to 65 years) as adjunctive screening to routine Pap cytology and for patients below the age of 30 years who had a cytology result of atypical squamous cells of undetermined significance or more severe changes. In this rural and suburban population, the cervical cancer prevalence is less than 6.8/100 000 women.²² The HPV positive prevalence rate for the patients below 30 years old was found to be 36.1%, and that for those 30 years and older was 7.3%.¹⁸ Publication of laboratory data with concealed patient identities was approved by the Milford Hospital Institutional Review Board.

For HPV detection and genotyping, the pellet derived from 5% of the liquid-based Pap cytology sample was digested in proteinase K, and 1 µl of the digestate was used for nested PCR amplification, first with a pair of MY09/MY11 degenerate outer primers and then with a nested PCR primer pair (ie, a GP5+/GP6+ pair or a GP6/MY11 pair (or its equivalent)).^{17–19,21}

The positive nested PCR products were subjected to direct automated DNA sequencing without further purification.^{17–21} An exclusive unique '100% identities' match between the query and subject sequences, returned by the Basic Local Alignment Search Tool (BLAST) online algorithm, was required for genotyping except for variants not yet recorded in the GenBank, as reported previously.^{17–19,21}

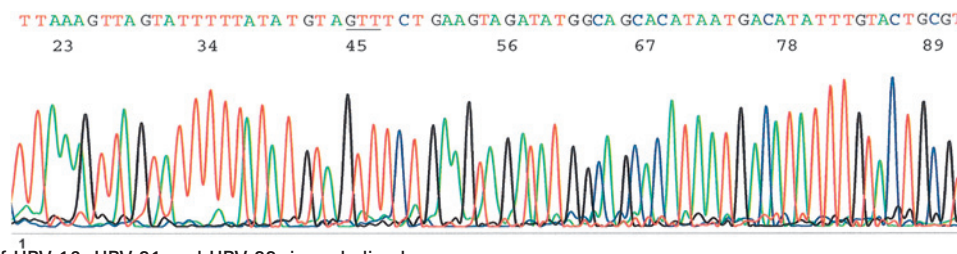
RESULTS

Analysis of a DNA sequence of 34 bases immediately downstream of the GP5+ priming site was



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jcp.bmj.com/site/about/unlocked.xhtml>.

Figure 1 Computer-generated electropherogram showing a 72-base sequence of the L1 gene of a human papillomavirus 16, Locus FJ006723 or its variant. A GP5+/GP6+ nested PCR amplicon was the template. The sequence of the GP5+ priming site has been deleted from the upstream 5' end on the right. The 3-base crucial sequence GTT, for differential genotyping of HPV-16, HPV-31 and HPV-33, is underlined.



adequate for accurate genotyping of all HPV variants encountered, except for HPV-16, HPV-31 and HPV-33 (figure 1). The online BLAST sequence algorithm reports always noted that some variants of HPV-16, HPV-31 and HPV-33 in the GenBank database share a short sequence homology for up to 45 bases in this region (figure 2), indicative of an ambiguous typing result, when a sequence of the HPV-16 L1 gene in this region of less than 46 bases was submitted for BLAST algorithm. Genotyping distinction between these three HPV types depended on identifying the sequence of the adjacent three 'crucial' bases further downstream. A crucial three-base sequence 'GTT' identified one of the HPV-16 variants represented by HPV-16 GenBank Locus FJ006723; a sequence 'GGT' identified one of another group of HPV-16 variants represented by HPV-16 GenBank Locus AF134178; a sequence 'GGC' identified an HPV-31 variant (GenBank Locus EF140820); and a sequence 'CGT' identified an HPV-33 variant (GenBank Locus DQ448214).

Using the crude proteinase K digest of clinical materials of a complex nature for PCR and direct automated DNA sequencing might generate electropherograms of varying qualities. When the quality of an electropherogram was high, recognition of the 3-base 'crucial' sequence (figure 1) for differential genotyping of HPV-16 was straightforward by selecting at least 48 bases downstream of the GP5+ priming site for a BLAST analysis. However, when the electropherogram was less than perfect (figure 3), an inexperienced sequence analyst might have selected the first 34 bases for BLAST sequence alignment algorithms, leading to genotyping ambiguities or errors. For the latter reason, a GP6/MY11 heminested or an equivalent HiFi nested PCR primer amplicon was used to generate an extended electropherogram with an additional base-calling stretch upstream of the GP5+ priming site. In this upstream stretch, the GP5+ priming site became part of the sequence useful for base calling.

Any 34-base sequence in the extended upstream segment, including the entire 23 bases at the GP5+ binding site or a fraction of it, was found to be sufficient for distinguishing any HPV-16 variants from other closely related HPV types without a reasonable doubt. As illustrated in the two sample electropherograms selected for demonstration (figure 4 and figure 5), all

HPV-16 isolates detected in this population had a '100% identity' match in sequence with a variant of one of the two HPV-16 prototypes: Locus FJ006723 and Locus AF134178. These latter two representative prototypes of HPV-16 differ from each other by only 1 base at nucleotide 47 downstream of the GP5+ priming site. But each of them in turn depends on the 3-base crucial sequence at this site for distinction from certain variants of HPV-31 or HPV-33 (figure 2).

Of the 2740 specimens analysed, 202 (7.4%) were found to be positive for one of the 13 'high-risk' HPV genotypes targeted by the FDA-approved Digene HC2 assay. Specimens with mixed HPV infections were excluded from this study. In this local patient population, mixed HPV infections were found in 4.7%–8.5% of the HPV-positive specimens.^{18 19}

Each nested PCR amplicon was confirmed by matching its sequence with a unique genotype-specific DNA sequence through online BLAST algorithms, using the HPV DNA sequences deposited in the GenBank database as the reference. There were 35 single HPV genotypes identified, including the high-risk and low-risk types. The general distribution pattern of the individual HPV genotypes in this rural and suburban population has been published elsewhere.¹⁸

HPV-16 was found to be the most prevalent among the 'high-risk' HPV infections (*n*68; 33.6%). The HPV-16 DNA in 28 of 68 positive specimens (41.2%) relied on a nested PCR for detection. The prevalence of the individual 'high-risk' HPV genotypes is summarised in table 1.

There were 10 high-grade squamous intraepithelial lesion (HSIL) results in the 2740 companion Pap cytology reports. Correlation of the HPV genotyping data with the Pap cytology results showed that six cases of HSIL were associated with a single HPV-16, two cases of HSIL with a HPV-31, one HSIL with a HPV-52, and one HSIL with a HPV-69 infection. No HSIL cytology was associated with a mixed HPV infection.

Colposcopic biopsies were performed on 18 of the 68 patients positive for HPV-16, including the six HPV-16 positive patients whose Pap cytology results were HSIL. Of these six HPV-16 positive HSIL cases, two (2/6) were finally confirmed by histology to harbour a CIN3/2 lesion, and the other four (4/6)

Figure 2 Alignment of four 72-base sequences representing two common variants of human papillomavirus 16 (HPV-16), one variant of HPV-31 and one variant of HPV-33, immediately downstream of their GP5+ priming site (not shown). The three crucial bases for differential genotyping are bold.

DNA sequence homologies retrieved from the National Center for Biotechnology Information database.

```

HPV-16 (1)      ttaaagtagtagtatttttatatgtagtttctcgaagtagatatggcagcacataatgacataattgtactgcgt 5'
HPV-16 (2)      -----ggt-----gatatggcagcacataatgac----- 5'
HPV-31 (3)      -----ggc-----gatatggcagcacataatgac----- 5'
HPV-33 (4)      -----cgt-----gatatggcagcacataatgac----- 5'

```

GenBank Locus: (1) FJ006723, (2) AF134178, (3) EF140820, and (4) DQ448214.

A 45-base DNA sequence homology immediately downstream of GP5+ priming site (5') is shown on the right for certain variants of HPV-16, -31, and -33. A 24-base sequence homology is shown on the left. A 3-base crucial sequence (boldfaced) for differential genotyping of certain variants of HPV-16, HPV-31 and HPV-33 is flanked by these two sequence homologies. There is only 1-base difference between sequences (2) and (3), and between sequences (2) and (4) in this region.

gatatggcagcacataatgac Sequence of a probe used for HPV-16 genotyping in the Roche LBA [27]
 gtatgttctcgaagtagatatgg Sequence of HPV-16 capture probe by Motorola Life Sciences [28]

Figure 3 This is a less than perfect computer-generated electropherogram similar to that shown in figure 1. An imperfect base-calling sequence tracing may prevent selection of a long enough sequence inclusive of the three crucial bases GTT (underlined) for Basic Local Alignment Search Tool (BLAST) algorithm, causing ambiguity in differential genotyping of human papillomavirus 16.

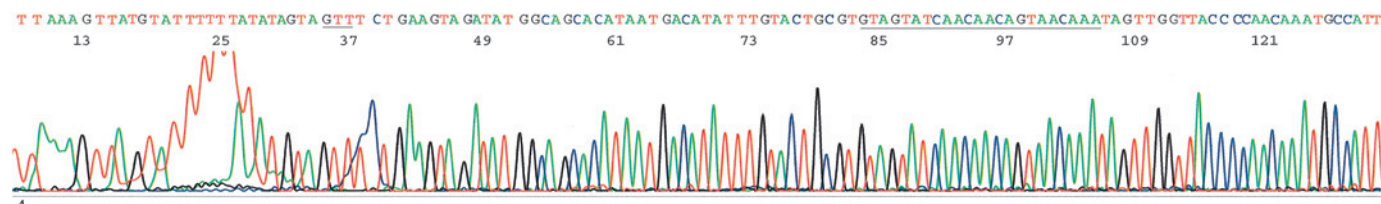
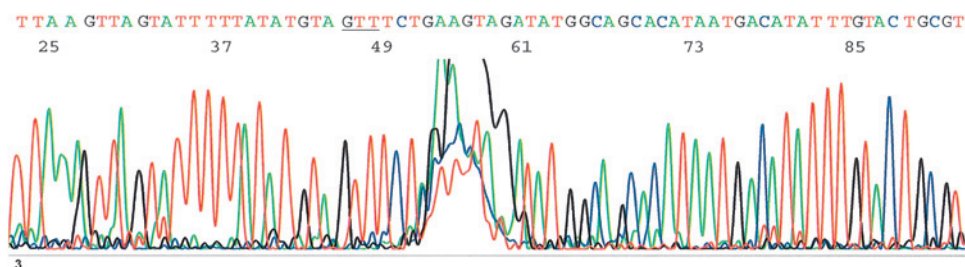


Figure 4 This is the extension of a sequence electropherogram to include the 23-base GP5+ priming site (underlined) and 27 bases further upstream. All variants of human papillomavirus 16 (HPV-16) contain a 34-base type-specific '100% identity' matched sequence selected randomly from the 50 bases on the right. Note the crucial three-base sequence GTT (underlined) in the region downstream of the GP5+ priming site for HPV-16 Locus FJ006723 and variants.

were found to be histologically negative on colposcopic biopsy, indicating that reversible HSIL cytology was associated with an HPV-16 infection in these four patients. Therefore, the positive predictive value (ppv) of HSIL cytology for CIN3/2 was 33.3%.

Among the other 12 HPV-16 positive patients with a Pap cytology lower than HSIL, but selected by the gynaecologists for cervical biopsies, four (4/12=33.3%) were found to harbour a CIN3/2 histology.

There was no invasive cervical cancer recorded in the 68 patients who were found to be positive for HPV-16 DNA in their Pap cytology specimens. Since only 12 of the 62 patients whose cytology was positive for HPV-16 and whose Pap cytology was lower than HSIL were subjected to colposcopic biopsies, the gynaecologists in this community apparently did not rely on a one-occasion HPV-16 genotyping test result as the determining triage tool for referral to colposcopic biopsy. The availability of a routine HPV genotyping for follow-up of persistent HPV infections and a well-known low cervical cancer prevalence rate in this local patient population with an above-average level of healthcare education might have influenced the practice of the local gynaecologists.

DISCUSSION

The role of persistent infection by HPV as a tumour promoter in cervical cancer induction is well recognised.^{1-4 23} The risk of developing precancerous lesions or cancer is greatest in women positive for the same genotype of HPV on repeat testing over a period of time.¹⁻⁴ The medical community is now urged to

emphasise the persistence of a cervical HPV infection, not the single-time detection of HPV, in management strategies and health messages.²⁴

According to the definition adopted in virology and used by the GenBank, a genotype of HPV differs in its L1 gene DNA sequence by at least 10% from every other known HPV type.²⁵ However, sequence dissimilarities are unevenly distributed with scattered short sequence homologies between variants of different papillomavirus genotypes in various segments of the L1 gene.²⁶ As shown in the GenBank database, there are a few short sequence homologies, including one of 45 bases, between the HPV-16, HPV-31 and HPV-33 in their L1 genes downstream of the GP5+ priming site (Figure 2).

Some of the hybridisation probes used for HPV-16 genotyping, including the HPV-16A probe (GATATGGCAGCACATAATGAC) published by Roche Molecular Systems, Alameda, California, USA,²⁷ are within this 45-base sequence homology. Such a probe will hybridise with any variants of HPV-16, HPV-31 or HPV-33 with a DNA strand complementary to its sequence (figure 2). Another HPV-16 capture probe (GTAGTTTCTGAAGTAGATATGG) used for chip development²⁸ contains a strand of seven nucleotides of the aforementioned Roche HPV-16A probe with an extension downstream to include the 3-base crucial sequence GTT. This latter HPV-16 capture probe would theoretically be able to discriminate against the HPV-31 and HPV-33 variants, which do not have a 3-base sequence complementary to GTT in this location (figure 2). By the same token, however, this probe will also fail to capture those HPV-16

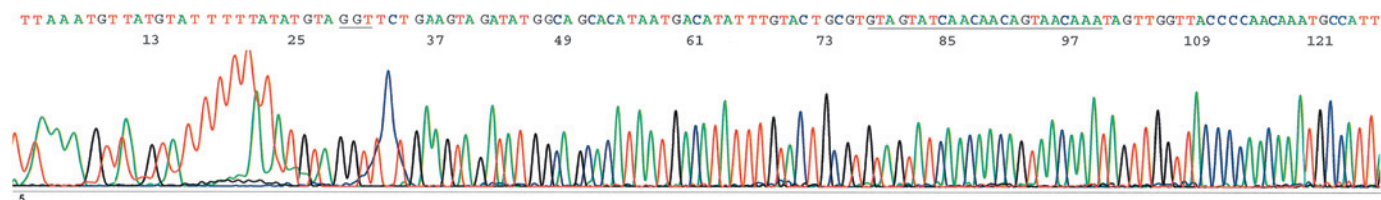


Figure 5 Sequence electropherogram almost identical to that in figure 4, characteristic of another common human papillomavirus 16 (HPV-16) variant with the 23-base GP5+ priming site underlined. Note the three-base crucial sequence GGT (underlined) in the region downstream of the GP5+ priming site for genotyping of HPV-16 Locus AF134178 and variants.

Table 1 Genotype prevalence of 202 'high-risk' HPV isolates from 2740 specimens in southern Connecticut

HPV type	Cases (n)	Isolates (%)
16	68	33.6
18	22	10.9
31	15	7.4
33	5	2.5
35	7	3.5
39	12	5.9
45	9	4.5
51	2	1.0
52	25	12.4
56	15	7.4
58	9	4.5
59	12	5.9
68	1	0.5
Total	202	100

The 13 human papillomavirus (HPV) genotypes listed are targeted by the Digene high-risk HPV HC2 assay. The positive rate is 202/2740=7.4%.

variants with a GGT complementary in this location, for example, HPV-16 GenBank Locus AF134178 and its variants (figure 2 and figure 5). According to the GenBank database, the unique signature sequence of HPV-16 in this region for genotyping differentiation from HPV-31 and HPV-33 is located in the 50-base segment further upstream, including the 23 bases of the GP5+ binding site (underlined in figure 4). This information appears to have been overlooked by some manufacturers of commercial test kits using GP5+ and GP6+ PCR primers for HPV-16 DNA amplification although a 22-base hypervariable strand of DNA upstream of the GP5+ binding site has been used for probe design by the SPF10-LiPA.

When relying on 'high-risk' HPV DNA testing in combination with cytological screening for triage, more than 95% of referrals to colposcopic biopsy for detection of cancer and precancer have been found to be excessive or unnecessary.²⁹ It has been suggested that increase in test specificity may lead to large changes in referral rates.²⁹ Our experience shows that using the most sensitive and specific nested PCR/DNA sequencing method for HPV DNA testing and a HSIL cytology result as the endpoint for evaluation, a one-occasion HPV test result may have a 100% clinical sensitivity, a 94% clinical specificity, and a 100% negative predictive value, but its ppv is only 3.4%–6% in a rural and suburban US population, with a cervical cancer prevalence rate of <6.8/100 000 women.^{18 19} According to the data presented in this series, if all 68 women found to be positive for HPV-16 were subjected to colposcopic biopsy work up, the one-occasion HPV-16 test would have a ppv of 8.9% (6/68) provided CIN3/2 is used as the endpoint for evaluation. The appropriate use of HPV-16 genotyping is to identify potentially carcinogenic persistent infections,³⁰ not as a referral tool to colposcopic biopsy.

Take-home message

- Accurate human papillomavirus 16 (HPV-16) genotyping plays an important role in cervical cancer risk management. However, errors in HPV-16 genotyping may occur because certain variants of HPV-16, HPV-31 and HPV-33 share short sequence homologies in the L1 gene commonly targeted for probe hybridisation. This study shows that the signature DNA sequence for differential genotyping is located in a 50-base segment outside of this common target.

Competing interests Dr Sin Hang Lee declares that he is a pathologist at Milford Hospital, Milford, Connecticut, USA, and the director of Milford Medical Laboratory. Dr Lee receives a fixed salary from the hospital which charges fees for cancer biopsies and HPV testing. Dr Lee is also the president and a shareholder of HiFi DNA Tech, LLC (www.hifidna.com), a company specialised in transferring the Sanger DNA sequencing technology to community hospital laboratories for DNA testing. Veronica S Vigliotti: None to declare. Dr Suri Pappu: None to declare.

Ethics approval Ethics approval was obtained from the Milford Hospital Institutional Review Board.

Patient consent Not obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Wallin KL, Wiklund F, Angström T, *et al*. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. *N Engl J Med* 1999;**341**:1633–8.
2. Kjaer SK, van den Brule AJ, Paull G, *et al*. Type specific persistence of high risk human papillomavirus (HPV) as indicator of high grade cervical squamous intraepithelial lesions in young women: population based prospective follow up study. *BMJ* 2002;**325**:572–6.
3. Cuschieri KS, Cubie HA, Whitley MW, *et al*. Persistent high risk HPV infection associated with development of cervical neoplasia in a prospective population study. *J Clin Pathol* 2005;**58**:946–50.
4. Brummer O, Hollwitz B, Bohmer G, *et al*. Human papillomavirus-type persistence patterns predict the clinical outcome of cervical intraepithelial neoplasia. *Gynecol Oncol* 2006;**102**:517–22.
5. Seabrook JM, Hubbard RA. Achieving quality reproducible results and maintaining compliance in molecular diagnostic testing of human papillomavirus. *Arch Pathol Lab Med* 2003;**127**:978–83.
6. Fischer AH. College of American Pathologists—Practice patterns in HPV testing. *CAP Today* 2008.
7. Bosch FX, Manos MM, Muñoz N, *et al*. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J Natl Cancer Inst* 1995;**87**:796–802.
8. Kulasingam SL, Hughes JP, Kiviat NB, *et al*. Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. *JAMA* 2002;**288**:1749–57.
9. Klug SJ, Molijn A, Schopp B, *et al*. Comparison of the performance of different HPV genotyping methods for detecting genital HPV types. *J Med Virol* 2008;**80**:1264–74.
10. van Doorn LJ, Quint W, Kleter B, *et al*. Genotyping of human papillomavirus in liquid cytology cervical specimens by the PGMV line blot assay and the SPF(10) line probe assay. *J Clin Microbiol* 2002;**40**:979–83.
11. Vernon SD, Unger ER, Williams D. Comparison of human papillomavirus detection and typing by cycle sequencing, line blotting, and hybrid capture. *J Clin Microbiol* 2000;**38**:651–5.
12. Feoli-Fonseca JC, Oligny LL, Brochu P, *et al*. Human papillomavirus (HPV) study of 691 pathological specimens from Quebec by PCR-direct sequencing approach. *J Med Virol* 2001;**63**:284–92.
13. Kosel S, Burggraf S, Mommsen J, *et al*. Type-specific detection of human papillomaviruses in a routine laboratory setting—improved sensitivity and specificity of PCR and sequence analysis compared to direct hybridisation. *Clin Chem Lab Med* 2003;**41**:787–91.
14. Johnson T, Bryder K, Corbet S, *et al*. Routine genotyping of human papillomavirus samples in Denmark. *APMIS* 2003;**111**:398–404.
15. Speich N, Schmitt C, Bollmann, *et al*. Human papillomavirus (HPV) study of 2916 cytological samples by PCR and DNA sequencing: genotype spectrum of patients from the west German area. *J Med Microbiol* 2004;**53**:125–8.
16. Asato T, Maehama T, Nagai Y, *et al*. A large case-control study of cervical cancer risk associated with human papillomavirus infection in Japan, by nucleotide sequencing-based genotyping. *J Infect Dis* 2004;**189**:1829–32.
17. Lee SH, Vigliotti VS, Vigliotti JS, *et al*. Routine human papillomavirus genotyping by DNA sequencing in community hospital laboratories. *Infect Agent Cancer* 2007;**2**:11.
18. Lee SH, Vigliotti VS, Pappu S. Human papillomavirus (HPV) infection among women in a representative rural and suburban population of the United States. *Int J Gynaecol Obstet* 2009;**105**:210–14.
19. Lee SH, Vigliotti VS, Pappu S. Molecular tests for human papillomavirus (HPV), *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in liquid-based cytology specimen. *BMC Womens Health* 2009;**9**:8.
20. Lee SH, Vigliotti VS, Pappu S. DNA sequencing validation of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* nucleic acid tests. *Am J Clin Pathol* 2008;**129**:852–9.
21. Lee SH, Vigliotti VS, Vigliotti JS, *et al*. Validation of human papillomavirus genotyping by signature DNA sequence analysis. *BMC Clin Pathol* 2009;**9**:3.
22. Hovey D. State must do more to prevent cervical cancer. http://www.housegop.ct.gov/%5Cpressrel/%5CHoveyD112%5C2007%5C20070116_HoveyD112_01.pdf (accessed 8 December 2009).
23. Lu H, Ouyang W, Huang C. Inflammation, a key event in cancer development. *Mol Cancer Res* 2006;**4**:221–33.
24. Rodríguez AC, Schiffman M, Herrero R, *et al*. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst* 2008;**100**:513–17.

25. **de Villiers EM**, Fauquet C, Broker TR, *et al*. Classification of papillomaviruses. *Virology* 2004;**324**:17–27.
26. **Chan SY**, Delius H, Halpern AL, *et al*. Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. *J Virol* 1995;**69**:3074–83.
27. **Gravitt PE**, Peyton CL, Apple RJ, *et al*. Genotyping of 27 human papillomavirus types by using L1 consensus PCR products by a single-hybridization, reverse line blot detection method. *J Clin Microbiol* 1998;**36**:3020–7.
28. **Vernon SD**, Farkas DH, Unger ER, *et al*. Bioelectronic DNA detection of human papillomaviruses using eSensor: a model system for detection of multiple pathogens. *BMC Infect Dis* 2003;**3**:12.
29. **Stout NK**, Goldhaber-Fiebert JD, Ortendahl JD, *et al*. Trade-offs in cervical cancer prevention: balancing benefits and risks. *Arch Intern Med* 2008;**168**:1881–9.
30. **Naucner P**, Ryd W, Törnberg S, *et al*. Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening. *J Natl Cancer Inst* 2009;**101**:88–99.