Published in final edited form as:

Gynecol Oncol. 2013 March ; 128(3): 433–441. doi:10.1016/j.ygyno.2012.11.028.

Genotyping human papillomaviruses: Development and evaluation of a comprehensive DNA microarray^{☆, ☆☆}

Jane Shen-Gunther^{a,*} and Jennifer Rebeles^{b,1}

Jane Shen-Gunther: jane.shengunther@us.army.mil; Jennifer Rebeles: rebeles@live.unc.edu

^aGynecologic Oncology & Clinical Investigation, Dept. of Clinical Investigation, Brooke Army Medical Center, Fort Sam Houston, TX, USA

^bGreehey Children's Cancer Research Institute, 8403 Floyd Curl Drive, San Antonio, TX 78228, USA

Abstract

Goals—To define the analytical and clinical performance of a human papillomavirus (HPV) custom-designed microarray targeting the HPV L1 gene for viral genotyping.

Methods—Microarray probes were designed by cataloging the genome sequence of all 120 known HPV types to generate tiling probes using eArray[®] software against the unique L1 capsid gene segments targeted by MY09/11 and FAP59/64 primers. The microarray (1 slide×8 arrays×60 K features) synthesized in situ by inkjet printing was tested using synthetic type-specific HPV DNA and existing HPV DNA from cervical cytology. The synthetic HPV L1 segments (genotypes 6, 11, 16, 18, 31, 33, 35, 45, 53, 58, 66, 73, 83) were manufactured from sequences stored in the NCBI taxonomy database. Using the hybridization patterns of the synthetic HPV DNA as the Support Vector Machine classifier, HPV DNA from patient samples were genotyped and compared to antecedent DNA sequencing/BLAST[®] results for concordance.

Results—16 cytology-derived HPV DNA samples and 13 synthetic type-specific HPV DNA samples were tested singly, in duplicate, or in combination on 40 arrays. The synthetic HPV DNA hybridization patterns were found to be uniquely distinctive to serve well as a classifier of unknown HPV-containing specimens. For the 16 HPV DNA+ samples classified, 15 were concordant with DNA sequencing results. In 6/16 (38%) samples, the microarray hybridization pattern revealed 2 concurrent HPV infections.

*Corresponding author at: Gynecologic Oncology & Clinical Investigation, Department of Clinical Investigation, Brooke Army Medical Center, ATTN: MCHE-CI. Dept. of Clinical Investigation, 3400 Rawley E. Chambers Ave., Ste A. Fort Sam Houston, TX 78234-6315, USA

¹Present address: Dept. of Nutrition, University of North Carolina at Chapel Hill, Gillings School of Global Public Health, Chapel Hill, NC 27599.

Conflict of interest statement

The author has no conflicts of interest to declare.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ygyno.2012.11.028>.

The view(s) expressed herein are those of the author(s) and do not reflect the official policy or position of Brooke Army Medical Center, the U.S. Army Medical Department, the U.S. Army Office of the Surgeon General, the Department of the Army, Department of Defense or the U.S. Government.

Conclusion—The novel “HPV Array” was sensitive and specific for detecting single and multiple infections. This proof-of-principle project demonstrated the accuracy and advantages of microarray technology for HPV genotyping.

Keywords

HPV; HPV genotyping; HPV microarray; HPV molecular detection; DNA microarray

Introduction

Since the initial isolation of papillomavirus (PV) from warts of wild cottontail rabbits in 1932, over 189 species-specific PV have been characterized to date [1,2]. The most recent addition to the PV phylogenetic tree is the *Camelus dromedaries* isolated from warts on the lips and muzzles of young Sudanese camels [3]. Similarly, the list of human papillomavirus (HPV) genotypes has grown to ~120 subsequent to the isolation and cloning of HPV-16 by zur Hausen and colleagues in 1983 [4]. As scientists continue to investigate and sequence HPV from less well-studied, extragenital sites using modern molecular techniques, novel genotypes will undoubtedly be discovered. Recently, Botalico et al. detected a remarkable spectrum of HPV from healthy subjects totaling 59 unique types from the oral cavity and 53 from cervical cytology [5].

Currently, PCR plus amplicon sequencing is the most widely used method for genotyping HPV. This method is also the preferred reference standard for validation of new in vitro diagnostic devices (IVDs) by the U.S. Food and Drug Administration (FDA) [6]. However, PCR/Sequencing is labor intensive and impractical for high throughput screening required in clinical laboratories. Instead, HPV IVDs based on hybridization technology that offers HPV qualitative detection has dominated the market. These IVDs primarily detect the 12 high-risk (HR) “carcinogenic” HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59) and “probably” or “possibly carcinogenic” types (66, 68, 73) as categorized by the International Agency for Research on Cancer (IARC) [7]. The 1st FDA-approved IVD is Hybrid Capture® 2 (Qiagen, CA) which utilizes labeled RNA probes to detect 13 HR HPV DNA [8]. Cervista™ (Hologic, MA) and Cobas® (Roche Molecular Systems, CA) were licensed in 2009 and 2011, respectively, for detection of 14 HR types and identification of HPV-16/18 [9,10]. APTIMA® (Gen-Probe, CA) detects viral E6/E7 mRNA from 14 HR types [11]. These 4 FDA-approved tests are effective for qualitative detection, but notable for two major limitations 1) constrained spectrum of HPV genotypes detected, and 2) non-specific dichotomous result except for HPV-16/18. These limitations, therefore, urge the development of a comprehensive, sensitive, specific, automated test that has broad clinical applications from epidemiological surveillance to vaccine development.

In this study, we aimed to develop a comprehensive HPV genotyping assay based on DNA microarray technology. DNA probes were designed to detect all 120 known HPV genotypes. Analytical performance of the microarray was evaluated by testing with synthetic type-specific HPV L1 DNA sequences. Clinical performance was determined by direct comparison of genotyping results between the microarray and PCR/Sequencing method using HPV-containing cervical cytology.

Materials and methods

Microarray design, manufacture, and hybridization

Microarray oligonucleotide probes were designed to target the HPV L1 capsid gene which defines the HPV genotype. All 120 HPV genotype sequences were acquired from GenBank® and aligned by Jalview2 (www.jalview.org) [12]. The HPV L1 amplicon sequences flanked by 2 consensus primer pairs, MY09/11 and FAP59/64, were located using Jalview2, extracted and submitted to eArray® software (<https://earray.chem.agilent.com/earray>) [13] for probe design (Fig. 1). The specific priming locations of MY09/11 and FAP59/64 on the HPV-16 and -8 genomes respectively, are located between nucleotide positions 6584/7035 and 5981/6436 [14,15]. Tiling probes, 60-mer in length and 15 nucleotides apart were created and grouped according to each HPV genotype and sub-grouped by the L1 target sequence (named “MY or FAP probes”). Due to noncontiguous segments in the HPV genome flanked by the FAP primers, the two longest contiguous segments were selected for probe design and named “FAP L1-S1 and L1-S2” for segment 1 or 2. The final CGH microarray (1 slide×8 arrays×60 K features) contained 2689 MY- and 1627 FAP-custom probes in 6 replicates distributed in a random layout. An additional 10,148 Agilent normalization, control and replicate probes totaled 36,044 occupied features on the array. The microarray was manufactured by Agilent Technologies (Santa Clara, CA) using inkjet printing technology for in situ synthesis of DNA probes on a slide. Hybridization of MY and FAP amplicons (clinically or synthetically derived) to the array was conducted according to Agilent’s enzymatic labeling protocol v. 6.2.1. In brief, MY and FAP amplicon DNA (200 ng each) were fluorescent-labeled with cyanine 3-dUTP (Cy3) or cyanine 5-dUTP (Cy5). Due to insufficient amplicon DNA for samples #40 and 55, only 100 ng (1/2 the recommended quantity) of the MY and FAP amplicons were labeled. The labeled DNA combined with hybridization master mix was deposited onto the array slide, sandwiched between the array and gasket slides via a clamp assembly, and incubated in a hybridization oven at 65C for ~36 h. This was followed by the manufacturer’s standard washing procedures. All array slides were covered with ozone-barrier to prevent ozone-induced dye degradation and placed in the slide holder prior to scanning.

Microarray scanning, feature extraction and data analysis

Arrays were tested and scanned using the Agilent G2565CA dual laser scanner at Greehey Children’s Cancer Research Institute (GCCRI) in San Antonio, Texas. Image data were extracted by Feature Extraction software (v10.7) (Agilent Technologies) and saved for analysis by other analytical software listed in the statistical section.

Synthetic HPV L1 sequence design and production

Thirteen HPV genomes (types 6, 11, 16, 18, 31, 33, 35, 45, 53, 58, 66, 73, 83) were acquired from GenBank® and aligned by Jalview2 [12]. The amplicon sequences flanked by the MY and FAP primer pairs on the L1 gene, were located using Jalview2, extracted and submitted to Eurofins MWG Operon (Huntsville, AL) for gene synthesis. The amplicon sequences of MY (450-bp) and FAP (480-bp) for each HPV type were inserted into the vector backbone (pCR2.1) for cloning and production of purified PCR products. The type-specific MY- or

FAP-amplicons were used to test 1) analytical performance of the array and 2) clinical performance by serving as positive controls on the array.

Samples with HPV DNA

Cervicovaginal cytology specimens (N=16) containing an assortment of HPV genotypes identified during the first phase of the research project were used to test the clinical performance of the microarray. The clinical and cytological characteristics of the study population (N=90), as well as, HPV DNA sequencing results have been published [16]. To recap our methods, genomic DNA was extracted from cervicovaginal cells and tested for HPV DNA by PCR using 3 primer sets, i.e. GP 5+/6+, MY09/11 and FAP59/64. Samples deemed HPV positive by gel electrophoresis were subjected to DNA sequencing and identification with the Basic Local Alignment Search Tool (BLAST®) by querying against HPV sequences in GenBank®. Of these HPV containing specimens, 16 samples were selected for array testing based on 1) maximal diversity of HPV genotypes, and 2) chromatogram patterns indicative of pure or mixed infections. Specifically, maximal phylogenetic diversity meant testing and studying the hybridization patterns of 11 distinct HPV genotypes (types 6, 16, 18, 31, 33, 53, 58, 59, 66, 73, 81) from 7 disparate species (15 total) belonging to the Alpha-papillomavirus genus. These samples along with the library of 13 synthetic HPV DNA were tested on 40 arrays.

Support Vector Machines for HPV classification

Support Vector Machines (SVM) is a form of supervised learning models that analyze data for pattern recognition and classification. The SVM mathematical and conceptual framework is complex and detailed in mathematical and computing realms [17,18]. In simple terms, SVM algorithms of classification are derived from the dataset of known samples (“training set”) belonging to one of two categories (X and not-X) (Fig. 3B). The trained algorithm then classifies unknown samples into one category or the other by a linear separating hyperplane. The SVM method determines the location of the hyperplane which maximizes the distance between the 2 closest data points on either side of the plane. For this study, the microarray hybridization pattern of each type-specific synthetic HPV DNA was characterized and then used as a “genotyping” classifier for the clinical specimens.

Statistical analysis

Data were summarized using means (95% CI), medians (IQR), and proportions. Pairwise correlation was used to compare replicate inter-array mean signals of each feature. Agreement between 2 dichotomous categorical variables was calculated using % and kappa. Gapped alignment significance values (E-value) were generated by WU-BLAST. A *p*-value <0.05 was considered statistically significant. Statistical analyses were performed using Excel (Microsoft, Redmond, WA), STATA/IC 11.0 (StataCorp, College Station, TX), GeneSpring GX (Agilent Technologies, Santa Clara, CA), and WU-BLAST (European Bioinformatics Institute, Cambridgeshire, UK) [19].

Results

Synthetic HPV L1 DNA hybridization patterns

The fluorescence intensity of each feature (spot) on a hybridized array was quantified by mean “signal intensity” units ranging from 0 to 65,527 in green or red channel and visualized collectively in histogram format (Fig. 2A). The characteristic positively skewed distribution was summarized by the median (IQR) and presented as a boxplot (Fig. 2B). Of note, for each feature, the mean green or red signals (abbreviated as gMeanSig or rMeanSig) were calculated by taking the average intensity of inlier pixels (~42 pixels/spot). The median (IQR) signal intensity of the manufacturer’s negative control probes (n=252/array) were 35 (32–37) and 49 (47–52) for green and red fluorescence, respectively.

For each synthetic HPV DNA tested, the corresponding type-specific probes were identified, grouped, and counted to characterize the hybridization pattern of the “probe group.” Table 1 presents the hybridization results of 13 individual and 3 paired HPV sequences. The HPV type-specific tiling probes designed to target the MY amplicons revealed a highly significant rate of matching with 24/24 probes per group fluorescing maximal green signals (~65,527) (Table 1, column 3, 4). Similarly, for probes targeting the FAP amplicons, the HPV specific probe groups fluoresced maximal red signals (~65,527) (Table 1, column 8, 9) except for 4 probe groups. For HPV 33, 35, and 45, four contiguous probes targeting the FAP S2 amplicon failed to hybridize as demonstrated by low median signals ~50 (range, 45–57). For HPV 58, eleven probes targeting the FAP S1 amplicon failed to hybridize with a median signal of 52 (range, 42–463). Upon review of our records, the synthetic HPV plasmid (types -33, -35, -45, -58) for the FAP sequences and its PCR products were generated on the same date by the manufacturer. Hence, the most plausible explanation for failure of the FAP target-probe hybridization was HPV DNA sequence error either in the plasmid or PCR products.

The results clearly demonstrate a significant difference in signal intensities between HPV target-specific probes (HPV-X) and non-target specific probes (~HPV-X) after hybridization with respective median signals of ~65,527 versus ~30 (green) or ~60 (red) (Table 1). The features with the highest intensity (65,527) in the array, usually indicating a perfect “probe-target” match, were filtered for detailed analysis. A representative plot of probes with the brightest signals in an array (green and red) is presented graphically (Fig. 3A). In the diagram, the HPV-16 hybridization profile revealed maximum target-specific probe signals for MY-Cy3 and FAP-Cy5 labeled amplicons indicative of correct hybridization. This analysis also revealed cross-hybridization by non-target HPV probe groups, but the number of probes per group was usually <10. Based on the phylogenetic tree of HPV, the varying degrees of target-probe cross hybridization reflect sequence similarity or difference among genotypes. This was expected and demonstrated by our results. For the 3 paired HPV sequences (types 6 & 11; 16 & 18; 53 & 66), the correct and cross hybridization patterns were nearly the same as the individual HPV sequences tested.

The tiling probe method used for the design of the microarray utilizes the “sequencing by hybridization” principle. Hybridized probes when “strung together” decipher the nucleotide identities and length (bp) of an unknown sequence. Such a sequence may then be aligned

with known sequences for identification. Table 1 lists the sequence similarity search results by alignment of each HPV-specific MY or FAP amplicon sequence targeted by our designer probes. MY (~410 bp) and FAP (~300 bp) sequence alignments were performed singly or in series (~710 bp) as a gapped alignment in WU-BLAST [19]. The length-dependent specificity of alignments increased exponentially by including the FAP with the MY segment. The statistical significance of alignments reported as E-value (similar to p-value) ranged from 5.2E-109 to 7.6E-150 for MY+FAP sequences (Table 1).

Cytology-derived HPV L1 DNA hybridization patterns

HPV DNA extracted and amplified from the cytology of 16 different subjects was tested on 24 arrays. The samples were tested singly (n=8) or in duplicate (n=8), after labeling with Cy3 or Cy5, and without/with positive controls (n=4) to explore all possible permutations of experimental conditions. Eight samples (ID# 21, 23, 43, 49, 54, 55, 56, 71) were tested in duplicate (inter-array technical replicates) initially to evaluate the repeatability of inter-array experiments (Table 2). Samples #23 and 54 (2/8 samples) revealed poor hybridization in 1 of the replicates with non-specific binding of Cy3 or Cy5 labeled MY or FAP amplicons showing low feature signals. For 6 samples (ID #21, 43, 49, 55, 56, 71), the MY-Cy3 and FAP-Cy5 labeled amplicons bound well to the target specific probes with signal intensities of ~65,527; whereas, the non-target probes exhibited low signals consistent with non-binding probes or negative controls (median ~30 or ~60 for respective green or red signals). A two-way scatterplot demonstrating the high correlation between replicate arrays (ID# 43, 21 containing HPV-53 and -83) for both green and red signals is presented in Fig. 4.

Five HPV-16 containing samples (ID# 11, 33, 40, 50, 55) identified previously by direct DNA sequencing were hybridized with Cy3-labeled MY-amplicons alone and/or with Cy5-FAP amplicons (if available). All target-specific probes exhibited high signal intensities ~of 65,527. In 4/5 samples (ID# 11, 33, 40, 55), high signals were detected in other HPV-specific probe group(s) indicative of a co-infection (Table 2, column 2). Sample #37 containing HPV-18 was the 5th sample to contain a co-infection with HPV-53.

Four samples (ID# 37, 50, 52, 57) were hybridized concurrently with a synthetic HPV-35 or -45 MY-amplicon labeled with Cy5. The particular HPV-type chosen was based on phylogenetic distance between the sample and control in an attempt to avoid cross hybridization. For each array, both the sample and control fluoresced maximally ~65,527 without evidence of interference or cross-hybridization.

Comparison of microarray and PCR/sequencing for HPV genotyping

A total of sixteen HPV-containing samples were tested by the microarray. The HPV genotype identified by PCR/Sequencing was designated as the primary infection since direct sequencing is considered as the reference standard. Additional HPV types identified by hybridization were designated as concurrent infection(s). Only 1 out of 16 samples was discordant with prior HPV typing results by direct sequencing. Sample #21 was sequenced after PCR DNA amplification using MY primers. BLAST® alignment identified the amplicon sequence as HPV-81. The maximum identity of the sequence alignment (ratio of query to subject sequence length), % homology, and E-value was 268/302 (88%), 4.0E-97.

In contrast, the microarray results indicated the presence of HPV-83 by detecting both MY and FAP amplicon sequences. The array result was more specific as demonstrated by a gapped alignment E-value of $3.7E-151$ (Fig. S1). The microarray also detected a second ($n=4$) and multiple HPV infections ($n=2$) in 6 samples. In 4 HPV-16 containing samples, HPV-53, -62, and -86 were detected as secondary infections and in sample #55, four additional HPV types-52, -53, -54, and -73 were detected in replicate testing (Table 2). Samples #37 and 28 also had mixed infections detected by the microarray. Overall, microarray “sequencing” was highly concordant with the primary HPV genotype identified by direct sequencing with an agreement of 93.75% ($kappa=0.9324$). However, the arrays detected mixed infections in 6/16 samples (38%) and separate amplicons (MY and FAP) that increased “sequencing” specificity to bestow a diagnostic advantage.

Discussion

This study revealed the performance characteristics of our custom-designed microarray. The analytical performance using 13 synthetic HPV L1 DNA sequences exhibited maximal signal intensities at 65,527 for the anticipated probe/target matches suggesting “perfect” matches between DNA–DNA strands. Additionally, all probes within an HPV type-specific probe group bound equally well with signals at or near 65,527. Conversely, the analytical specificity was demonstrated by low signals (30–60) expressed by non-target HPV probes. Various degrees of cross-hybridization occurred as anticipated due to sequence similarity among target DNA. However, by using the SVM concept for analysis of microarray hybridization patterns, a hyperplane was drawn between the probes with the highest intensity (signal) and those with lower signals (noise) to filter out the noise. The hybridization profile of the selected probes was then generated for each synthetic HPV for use as a genotyping reference for unknown HPV type(s) in clinical specimens. Furthermore, “sequencing by hybridization” which is analogous to gapped alignment of two target sequences, i.e. FAP and MY amplicons, was shown to dramatically increase the specificity and statistical significance of genotype identification.

For clinical performance and validation of the microarray, the sensitivity between DNA sequencing and microarray were found to be similar in that both detected the presence of HPV DNA within the input DNA (~200 ng). Of interest, two samples (ID #40 and 55) were tested with only 100 ng of labeled MY and FAP amplicons due to low quantities of template DNA. The target-specific probe signals remained at or near the maximum level (~65,527) for both MY and FAP amplicons (Table 2). The finding suggests lower amounts of input DNA than recommended by the manufacturer (200–500 ng) may be used without signal degradation. However, formal 10-fold dilution studies of input DNA quantity are warranted to identify the lowest limit of detection to establish analytical sensitivity. Technical replicates of 8 arrays demonstrated excellent correlation between signal intensities for 6 of the arrays suggesting replication is unnecessary. For the 2 replicates that failed to hybridize well, microarray processing may have been the contributing factor. Synthetic HPV DNA served well as positive controls in 4 samples by binding to its specific probes. Hence, this type of +control with the built-in negative control may serve as quality checks for each array. In comparison to direct DNA sequencing, the microarray offered a diagnostic advantage in

that it detected co-infections and distinct amplicons (MY and FAP) that increased HPV genotyping specificity.

The strength of this study lies in the design of the microarray. Up to now, no molecular diagnostic assay has been developed to detect all known HPV genotypes. With plenty of room on the microarray for additional features, all 189 species-specific PV in the animal kingdom may be added without difficulty. The incorporation of FAP probes was critical for detecting HPV in MY-negative samples. Furthermore, FAP probes will be useful for testing oral and dermatological samples often containing *Beta* or *Gamma* HPV [5]. One limitation of the microarray design is lack of target-specific probes for E6/E7 genes. Detection of L1, E6, and E7 genes will widen the breadth of detection for low- and high-grade dysplasia and cancer. Additionally, the E6/E7 DNA sequences provide genetic information regarding oncogenic potential. In high-grade lesions and cancer, HPV DNA is integrated into the cellular genome with loss of most viral genes except for E6 and E7 [20,21]. Hence, assays that only detect HPV L1 may intrinsically lower the clinical sensitivity for severe disease due to systematic “non-detection” bias. To design a truly comprehensive molecular diagnostic tool, viral genes present in various stages of disease must be recognized and detected.

Since this proof-of-principle study tested the microarray as a “prototype,” the limitations of this study stem from the number of arrays tested. This may be explained by the expense of the reagents, microarrays, and processing costs of the specimens. In comparison to other studies using oligonucleotide microarrays for virus genotyping from measles to potato viruses, our results affirm the utility of this technology for potential clinical use [22,23]. In fact, commercially available kits are already available for identification of respiratory viruses using different microarray platforms [24].

In conclusion, we evaluated the analytical and clinical performance characteristics of a custom-designed DNA microarray. Synthetic HPV L1 DNA served well for testing the sensitivity and specificity of the array and for establishing the hybridization profile of HPV-specific probe groups. The novel array in comparison to direct DNA sequencing offers the greatest specificity and breadth of detection for known single and mixed infections. To extend and translate this work to into the clinical realm, probes targeting the HPV E6 and E7 genes will be incorporated in our next prototype.

Acknowledgments

This work was supported in part by a CTSA grant (UL1TR000149) from the University of Texas Health Sciences Center at San Antonio, San Antonio, Texas. Agilent Technologies provided 3 microarrays via a Material Transfer Agreement. GeneSpring GX software support was provided by the Computational System Biology Core at University of Texas at San Antonio, funded by the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health.

References

1. Shope RE, Hurst EW. Infectious papillomatosis of rabbits: with a note on the histopathology. *J Exp Med.* 1933; 58:607–24. [PubMed: 19870219]

2. Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*. 2010 May 25; 401(1):70–9. [PubMed: 20206957]
3. Ure AE, Elfadl AK, Khalafalla AI, Gameel AAR, Dillner J, Forslund O. Characterization of the complete genomes of *Camelus dromedarius* papillomavirus types 1 and 2. *J Gen Virol*. 2011; 92:1769–77. [PubMed: 21471319]
4. zur Hausen H. Papillomaviruses in the causation of human cancers — a brief historical account. *Virology*. 2009; 384:260–5. [PubMed: 19135222]
5. Botalico D, Chen Z, Dunne A, Ostolozza J, McKinney S, Sun C, et al. The oral cavity contains abundant known and novel human papillomaviruses from the Betapapillomavirus and Gammapapillomavirus genera. *J Infect Dis*. 2011; 204:787–92. [PubMed: 21844305]
6. FDA Guidance for Industry (US). Establishing the performance characteristics of in vitro diagnostic devices for the detection or detection and differentiation of human papillomaviruses [internet]. Silver Spring (MD): US Food and Drug Administration; 2011. [[cited 2012 Oct 1]. Available from: <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm181509.htm>].
7. Schiffman M, Clifford G, Buonaguro FM. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infect Agent Cancer*. 2009; 4:8. [PubMed: 19486508]
8. QIAGEN Sample Assay Technologies (US). HC2 high-risk HPV DNA test: Digene Catalog Number 5101-1296 [internet]. Valencia (CA): QIAGEN; 2007. [[cited 2012 Oct 1]. Available from: <http://www.qiagen.com/products/digenehpvtesthc2.aspx>].
9. Hologic (US). Cervista HPV HR and 16/18 package inserts [internet]. Marlborough (MA): 2010. [[cited 2012 Oct 1]. Available from: http://www.cervistahpv.com/us_caribbean.html].
10. FDA Drugs@FDA (US). Cobas HPV Test package insert [internet]. Silver Spring, (MD): US Food and Drug Administration; 2011. [[cited 2012 Oct 1]. Available from: http://www.accessdata.fda.gov/cdrh_docs/pdf10/P100020c.pdf].
11. Hologic Gen-Probe (US). APTIMA HPV Assay package insert [internet]. San Diego (CA): Hologic GEN-PROBE; 2011. [[cited 2012 Oct 1]. Available from: <http://www.gen-probe.com/pdfs/pi/503745-EN-RevB.pdf>].
12. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2: a multiple sequence alignment and analysis workbench. *Bioinformatics*. 2009; 25:1189–91. [PubMed: 19151095]
13. Agilent Technologies (US). eArray product page [internet]. Santa Clara, (CA): [cited 2012 Oct 1]. Available from: <https://earray.chem.agilent.com/earray>.
14. Resnick RM, Cornelissen MT, Wright DK, Eichinger GH, Fox HS, ter Schegget J, et al. Detection and typing of human papillomavirus in archival cervical cancer specimens by DNA amplification with consensus primers. *J Natl Cancer Inst*. 1990; 82(18):1477–84. [PubMed: 2167986]
15. Forslund O, Antonsson A, Nordin P, Stenquist B, Hansson BG. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J Gen Virol*. 1999; 80:2437–43. [PubMed: 10501499]
16. Shen-Gunther J, Yu X. HPV molecular assays: defining analytical and clinical performance characteristics for cervical cytology specimens. *Gynecol Oncol*. 2011; 123:263–71. [PubMed: 21839499]
17. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci*. 2001; 98:15149–54. [PubMed: 11742071]
18. Zvelebil, M.; Baum, J. Cluster methods and statistics Understanding bioinformatics. New York: Garland Science; 2007. p. 662-3.
19. European Bioinformatics Institute (EBI). Sequence Similarity Searching WU-BLAST program. <http://www.ebi.ac.uk/Tools/sss/wublast/nucleotide.html>.
20. Howley, PM.; Lowy, DR. Papillomaviruses. In: Knipe, DM.; Howley, PM., editors. *Fields virology*. 5th. Philadelphia: Lippincott Williams & Wilkins; 2007. p. 2323-30.

21. Matlashewski, G.; Banks, L. Papillomaviruses. In: Acheson, NH., editor. Fundamentals of molecular virology. 2nd. Hoboken: John Wiley & Sons; 2011. p. 263-71.
22. Neverov AA, Riddell MA, Moss WJ, Volokhov DV, Rota PA, Lowe LE, et al. Genotyping measles virus in clinical specimens on the basis of oligonucleotide microarray hybridization patterns. J Clin Microbiol. 2006; 44:3752–9. [PubMed: 17021105]
23. Sip M, Bystricka D, Kmoch S, Noskova L. Detection of viral infections by an oligo-nucleotide microarray. J Virol Methods. 2010; 165:64–70. [PubMed: 20100517]
24. Miller MB, Tang YW. Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev. 2009; 22:611–33. [PubMed: 19822891]

HIGHLIGHTS

- ▶ A HPV genotyping microarray designed to detect all 120 known genotypes was developed.
- ▶ Analytical and clinical performance of the microarray was evaluated using synthetic and cytological type-specific HPV L1 DNA sequences.
- ▶ The novel array in comparison to DNA sequencing offers greater specificity and breadth of detection for single and mixed infections.

Fig. 1.

Conceptual illustration of the location of the L1 gene segment, MY09/11 and FAP59/64 priming sites, and tiling probes in relation to the HPV genome (8 kbp). Tiling probes 60-mer in length were created to target the amplicon sequences of MY (450 bp) and FAP (480 bp).

Fig. 2.

Distribution of fluorescence intensities on a hybridized array. Synthetic HPV-16 L1 MY- and FAP-amplicons labeled with Cy3 (green) and Cy5 (red), respectively, were used as the target. (A). Histogram of Log₁₀ intensity (x-axis) versus number of features (y-axis) segregated by fluorescence color (left panel: green) (right panel: red). The distribution of intensity signals were characteristically positively skewed for all hybridized arrays. Probes that fluoresced at or near the maximum intensity (65,527 or log₁₀ value ~4.82) are highlighted by the dashed outline. (B). Boxplot of Log₁₀ intensity (x-axis) stratified by 4 probe groups: 1) HPV-16 L1 (green signal) 2) HPV-16 L1_S1 & S2 (red signal) 3) Not-HPV-16 L1 (green signal) 4) Not-HPV-16 L1_S1 & S2 (red signal). The target-specific probes fluoresced at or near the maximum intensity (65,527 or log₁₀ value ~4.82) as highlighted by the dashed outline. The median intensities of the non-target probes for MY- and FAP-amplicons labeled respectively as Non-HPV-16 L1 and Non-HPV-16 L1_S1S2 were significantly lower at 28 (log₁₀ value ~1.45) and 59 (log₁₀ value ~1.77).

Fig. 3.

(A) Representative plot of HPV-specific probes with maximum fluorescence post array hybridization. Synthetic HPV-16 L1 MY-amplicons (Cy3 labeled) and FAP-amplicons (Cy5 labeled) were used as the target. The HPV-16 hybridization profile revealed maximum probe-target signals (65,527) for almost all designer probes, i.e. HPV-16 L1 MY- and FAP-probes (labeled as HPV_16_L1 and HPV_16_L1_S1S2 on the x-axis, respectively). Specifically, 24/24 (MY) and 12/14 (FAP) probes fluoresced maximally (□). Cross-hybridization with non-target probes for FAP-amplicons was characterized by maximal fluorescence, but sparseness of probes in each HPV defining group (□). This phenomenon is explained by the sequence homology within the L1 capsid gene among HPV genotypes. (B) Conceptual Venn diagram illustrating the Support Vector Machine (SVM). A hyperplane or line separates and classifies unknown samples into one of two categories (X or not-X). The SVM method determines the location of the hyperplane which maximizes the distance between the 2 closest data points on either side of the plane. The closest data points to the hyperplane are called support vectors. For this study, the signal intensity values of matching HPV probe:target's were clustered at or near the maximum of 65,527 versus <65,527 for non-target probes. Hence, the hyperplane derived from HPV-X (●) and not-HPV-X (○) probe signals was 65,527 intensity units.

Fig. 4.

Two-way scatterplot of inter-array fluorescence intensities on concurrently hybridized arrays. (A) Patient sample (ID#43) containing HPV-53 showed high correlation ($r=0.99$) between array signals in $\log_2$ scale (default) (left panel: green signal) (right panel: red signal). (B) Patient sample (ID# 21) containing HPV-83 showed similar high correlation ($r=0.99$) between arrays for both green and red signals. Image was prepared with GeneSpring GX.

Table 1

Microarray hybridization profile: synthetic HPV L1 DNA.

DNA		HPV-X ^a			~HPV-X ^b			HPV-X ^a			~HPV-X ^b			WU-BLAST ^c probability			
HPV		MY	Feature	Probe Grp	MY	Probe Grp	FAP	Feature	Probe Grp	FAP	Probe Grp	MY	FAP	MY+ FAP			
Type	(#)	probe (N)	gMeanSig ranged ^d	Med Sig (IQR) ^d	probe (N)	Med Sig (IQR)	probe (N)	rMeanSig ranged ^d	Med Sig (IQR) ^d	probe (N)	Med Sig (IQR)	E-val	E-val	E-val	E-val		
6		24	65,527	65,527	2665	29 (27–33)	14	65,527	65,527	1613	62 (57–86)	1.7E-86	4.2E-62	9.4E-148			
11		24	64,427–65,527	65,527	2665	29 (26–33)	14	63,292–65,527	65,527	1613	61 (56–81)	1.7E-86	6.7E-42	1.1E-130			
16		24	65,527	65,527	2665	28 (25–57)	14	65,317–65,527	65,527	1613	59 (53–125)	9.0E-85	1.5E-64	1.3E-131			
18		24	65,172–65,527	65,527	2665	29 (26–31)	14	64,224–65,527	65,527	1613	61 (54–131)	4.9E-84	1.1E-65	7.6E-150			
31		24	64,789–65,527	65,527	2665	32 (29–41)	15	64,765–65,527	65,527	1612	65 (56–152)	1.2E-86	1.0E-65	1.4E-130			
33		24	34,176–65,527	65,527	2665	29 (26–35)	11 ^e	64,575–65,527	65,527	1612	62 (53–129)	2.1E-85	1.3E-40	4.3E-129			
35		24	65,527	65,527	2665	29 (26–37)	11 ^e	65,188–65,527	65,527	1612	57 (50–111)	2.3E-86	1.5E-37	2.7E-128			
45		24	65,527	65,527	2665	36 (33–54)	11 ^e	65,527	65,527	1612	55 (50–95)	9.0E-87	2.3E-42	3.2E-132			
53		24	65,527	65,527	2665	30 (27–37)	14	63,971–65,527	65,527	1613	67 (55–179)	2.1E-85	4.2E-38	1.3E-126			
58		24	65,146–65,527	65,527	2665	30 (28–37)	4 ^f	64,790–65,527	65,172(64,810–65,527)	1612	51 (42–58)	1.6E-86	3.4E-18	5.2E-109			
66		24	64,964–65,527	65,527	2665	30 (27–36)	14	64,852–65,527	65,527	1613	63 (54–159)	5.8E-86	7.3E-60	9.6E-149			
73		24	64,826–65,527	65,527	2665	28 (26–32)	14	64,363–65,527	65,527	1613	58 (53–86)	1.8E-87	9.9E-42	2.7E-132			
83		24	65,249–65,527	65,527	2665	28 (26–33)	15	65,138–65,527	65,527	1612	63 (57–90)	4.4E-86	1.2E-61	3.7E-151			
6 & 11 ^g		24 ⁱ	64,819–65,527	65,527	2641	30 (27–39)	14 ⁱ	64,903–65,527	65,527	1599	69 (59–155)	1.7E-86	4.2E-62	9.4E-148			
16 & 18 ^g		24 ⁱ	64,321–65,527	65,527	2641	27 (24–42)	14 ^h	65,177–65,527	65,527			1.7E-86	6.7E-42	1.1E-130			
		24 ⁱ	65,065–65,527	65,527	2641		14 ⁱ	64,141–65,527	65,527	1599	70 (57–498)	9.0E-85	1.5E-64	1.3E-131			
		24 ^h	65,527	65,527			14 ^h	65,143–65,527	65,527			4.9E-84	1.1E-65	7.6E-150			

DNA	HPV-X ^a			~HPV-X ^b			HPV-X ^a			~HPV-X ^b			WU-BLAST ^c probability			
	MY	probe (N)	Feature	Med Sig (IQR) ^d	Probe Grp	MY	probe (N)	Feature	Med Sig (IQR) ^d	Probe Grp	FAP	Med Sig (IQR)	MY	E-val	FAP	MY+ FAP E-val
53 & 66 ^g	24'	65,527		65,527	29 (26–43)	2641	14'	65,527	65,527	74 (57–359)	1600		2.1E-85	4.2E-38		1.3E-126
	24''	65,527		65,527			13''	65,527	65,527				5.8E-86	7.3E-60		9.6E-149

Note. Synthetic HPV L1 DNA (13 types) were tested singly or in couplets (n=3) on 16 microarrays. HPV type-specific MY- and FAP-amplicons were labeled with Cy3 (green) and Cy5 (red) dye, respectively.

BLAST, Basic Local Alignment Search Tool; E-val, E-value; MY, MY 09/11 primers; FAP, FAP59/64 primers; gMeanSig, green mean signal; HPV, Human Papillomavirus; HPV-X, HPV type-X; ~HPV-X, not HPV type-X; IQR, interquartile range; Med Sig, median signal; Probe Grp, probe group; rMeanSig, red mean signal; WU-BLAST, Washington University BLAST.

^aHPV-X: HPV type-specific probes on the microarray targeting the same HPV type-specific MY- or FAP-amplicons listed under column heading “HPV TYPE.”

^b~HPV-X: HPV probes on the microarray not targeting the HPV type-specific MY- or FAP-amplicons listed under column heading “HPV TYPE.”

^cWU-BLAST: a sequence similarity search tool that permits sequences to have gaps (gapped alignments) and reported as E-value (statistical significance).

^dRange or IQR: For range and IQR values that are identical to the mean or median, only the central tendency is listed.

^eFour contiguous FAP probes failed to bind to the HPV type-specific amplicon as demonstrated by low median signal intensities ~50 (range, 45–57). These probes were excluded from the analysis.

^fEleven contiguous FAP probes failed to bind to the HPV type-specific amplicon as demonstrated by low median signal intensities ~52 (range, 42–463). These probes were excluded from the analysis.

^gTwo HPV type-specific amplicons were tested simultaneously on one microarray. The probes of the 1st and 2nd HPV types are designated by (') and (''), respectively.

Table 2

Microarray hybridization profile: Sample HPV L1 DNA.

Sample Info		HPV-X ^a				-HPV-X ^b				HPV-X ^a				-HPV-X ^b			
Sample ID ^c (#)	HPV Type ^d (#)	DNA label (Cy) ^e	MY probe (N)	Feature gMeanSig range ^f	Probe Grp Med Sig (IQR) ^f	MY probe (N)	Feature Med Sig (IQR)	DNA label (Cy) ^e	FAP probe (N)	Feature rMeanSig range ^f	Probe Grp Med Sig (IQR) ^f	FAP probe (N)	Med Sig (IQR)				
23	6	MY-Cy3	24	65,527	65,527	2665	41 (36–77)	FAP-Cy5	14	65,527	65,527	1613	125 (64–930)				
23 rep	6	MY-Cy3	24	20–228	29 (26–33)	2665	30 (26–36)	FAP-Cy5	14	44–4443	53 (46–417)	1613	50 (46–60)				
11	16	MY-Cy3	24	65,527	65,527	2641	32 (28–53)	–	–	–	–	–	–				
[53]		MY-Cy3	24	65,527	65,527	–	–	–	–	–	–	–	–				
33	16	MY-Cy3	24	63,769–65,527	65,527	2641	32 (28–53)	–	–	–	–	–	–				
[62]		MY-Cy3	24	34,515–65,527	65,507 (65,040–65,527)	–	–	–	–	–	–	–	–				
40	16	MY-Cy3	24	65,172–65,527	65,527	2665	34 (30–64)	FAP-Cy3	14	16,256–65,527	60,434 (47,337–65,527)	1613	45 (32–383)				
[86]		MY-Cy3	24	121–65,527	65,527	–	–	–	–	–	–	–	–				
55	16	MY-Cy3	24	65,527	65,527	2665	42 (31–94)	FAP-Cy5	14	28,912–65,527	65,527 (57,964–65,527)	1613	134 (59–2287)				
[52]		MY-Cy3	24	65,527	65,527	–	–	FAP-CY5	15	65,137–65,527	65,527	–	–				
[53]		MY-Cy3	24	65,290–65,527	65,527	–	–	FAP-CY5	14	65,527	65,527	–	–				
[54]		MY-Cy3	24	53,300–65,527	65,527 (64,732–65,527)	–	–	FAP-CY5	14	64,952–65,527	65,527	–	–				
[73]		MY-Cy3	24	64,109–65,527	65,527	–	–	FAP-CY5	15	65,151–65,527	65,527	–	–				
55 rep	16	MY-Cy3	24	22,127–65,527	65,527	2665	42 (31–94)	FAP-Cy3	14	16,579–65,527	61,580 (35,762–65,482)	1613	62 (34–1518)				
[52]		MY-Cy3	24	64,176–65,527	65,527	–	–	FAP-Cy3	15	65,527	65,527	–	–				
[53]		MY-Cy3	24	65,107–65,527	65,527	–	–	FAP-Cy3	14	47,062–65,527	65,525 (64,577–65,527)	–	–				
[54]		MY-Cy3	24	30,921–65,527	65,527	–	–	FAP-Cy3	14	64,458–65,527	65,527	–	–				
[73]		MY-Cy3	24	65,219–65,527	65,527	–	–	FAP-Cy3	15	65,527	65,527	–	–				
50	16	MY-Cy3	24	65,527	65,527	2665	31 (27–73)	–	–	–	–	–	–				

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Sample Info		HPV: X ^a			~HPV: X ^b			HPV: X ^a			~HPV: X ^b		
Sample	HPV	DNA	MY	Feature	Probe Grp	MY	Feature	DNA	FAP	Feature	Probe Grp	FAP	Probe Grp
ID ^c (#)	Type ^d (#)	label (Cy) ^e	probe (N)	gMeanSig range ^f	Med Sig (IQR) ^f	probe (N)	Med Sig (IQR)	label (Cy) ^e	probe (N)	rMeanSig range ^f	Med Sig (IQR) ^f	probe (N)	Med Sig (IQR)
	SYN 45	MY-Cy5	24	65,527	65,527	2665	53 (49–70)	–	–	–	–	–	–
37	18	MY-Cy3	24	65,527	65,527	2665	36 (29–103)	–	–	–	–	–	–
	[53]	MY-Cy3	24	65,527	65,527	–	–	–	–	–	–	–	–
	SYN 35	MY-Cy5	23	65,527	65,527	–	–	–	–	–	–	–	–
54	31	MY-Cy3	24	65,527	65,527	2665	37 (33–54)	FAP-Cy5	15	65,527	65,527	1612	80 (57–449)
54 rep	31	MY-Cy3	24	21–179	44 (27–37)	2665	30 (25–41)	FAP-Cy5	15	41–114	50 (47–52)	1612	51 (30–64)
56	31	MY-Cy3	24	65,527	65,527	2665	38 (35–48)	FAP-Cy5	14	65,527	65,527	1612	85 (58–454)
56 rep	31	MY-Cy3	24	65,527	65,527	2665	37 (34–53)	FAP-Cy5	14	65,527	65,527	1612	77 (56–376)
52	33	MY-Cy3	24	29,602–65,527	65,527	2665	32 (27–59)	–	–	–	–	–	–
	SYN 45	MY-Cy5	24	64,730–65,527	65,527	2665	51 (47–64)	–	–	–	–	–	–
43	53	MY-Cy3	24	65,527	65,527	2665	35 (31–45)	FAP-Cy5	14	64,100–65,527	65,527	1613	107 (58–995)
43 rep	53	MY-Cy3	24	65,527	65,527	2665	41 (37–53)	FAP-Cy5	14	64,330–65,527	65,527	1613	93 (57–670)
57	58	MY-Cy3	24	65,001–65,527	65,527	2665	32 (28–59)	–	–	–	–	–	–
	SYN 45	MY-Cy5	24	63,598–65,527	65,527	–	–	FAP-Cy5	14	65,527	65,527	–	–
28	59	–	–	–	–	–	–	FAP-Cy5	15	65,527	65,527	1612	84 (57–474)
	[42]	–	–	–	–	–	–	FAP-Cy5	14	65,527	65,527	–	–
	[54]	–	–	–	–	–	–	FAP-Cy5	13	65,527	65,527	–	–
71	66	MY-Cy3	24	65,527	65,527	2665	42 (37–64)	FAP-Cy5	13	64,108–65,527	65,527	1614	94 (57–717)
71 rep	66	MY-Cy3	24	65,527	65,527	2665	42 (37–62)	FAP-Cy5	13	65,188–65,527	65,527	1614	99 (55–734)
49	73	MY-Cy3	24	64,936–65,527	65,527	2665	42 (38–44)	FAP-Cy5	15	63,584–65,527	65,527	1612	59 (51–134)

Sample Info			HPV-X ^a			~HPV-X ^b			~HPV-X ^b		
Sample	HPV		DNA	MY	Feature	Probe Grp	MY	Feature	FAP	Feature	Probe Grp
ID ^c (#)	Type ^d (#)		label (Cy) ^e	probe (N)	gMeanSig range ^f	Med Sig (IQR) ^f	label (Cy) ^e	Med Sig (IQR)	probe (N)	rMeanSig range ^f	Med Sig (IQR) ^f
49 rep	73		MY-Cy3	24	65,527	65,527	FAP-Cy5	34 (41–44)	15	64,759–65,527	65,527
21 ^g	81(83)		MY-Cy3	24	65,527	65,527	FAP-Cy5	34 (30–57)	15	62,452–65,527	65,527
21 rep	81(83)		MY-Cy3	24	65,527	65,527	FAP-Cy5	35 (32–60)	14	65,527	65,527

Notes to Table 2

BLAST, Basic Local Alignment Search Tool; Cy, cyanine dye; Dash (–), not applicable; E-val, E-value; MY, MY09/11 primers; FAP, FAP59/64 primers; gMeanSig, green mean signal; HPV, Human Papillomavirus; HPV-X, HPV type-X; ~HPV-X, not HPV type-X; IQR, interquartile range; Med Sig, median signal; MY- or FAP-Cy3 or Cy5, MY or FAP amplicons labeled with Cy3 or Cy5; Probe Grp, probe group; rep, inter-array technical replicates; rMeanSig, red mean signal; SYN, synthetic HPV L1 DNA.

^aHPV-X: HPV type-specific probes on the microarray targeting the same HPV type-specific MY- or FAP-amplicons listed under column heading “HPV TYPE.”

^b~HPV-X: HPV probes on the microarray not targeting the HPV type-specific MY- or FAP-amplicons listed under column heading “HPV TYPE.”

^cCytologically derived DNA samples that were tested simultaneously on 2 separate arrays (inter-array technical replicates) are designated by the identifier “rep.”

^dConcurrent type-specific HPV infections identified by the microarray, but not by direct sequencing are designated by square brackets []. Synthetic HPV-35 or -45 amplicons were used as positive controls.

^eCyanine 3 (Cy3) or 5 (Cy5) fluorescent dyes were used to label MY- or FAP-amplicons derived from each clinical sample.

^fRange or IQR: For range and IQR values that are identical to the mean or median, only the central tendency is listed. For samples labeled with Cy-3 or Cy-5, respectively, green or red mean signal intensities are reported.

^gSample 21: Direct sequencing identified the MY-amplicon sequence as HPV-81. Microarray () identified the sample as HPV-83 using both MY and FAP amplicon sequences.