Journal of Clinical Virology 58 (2013) 437–442

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Next generation sequencing for human papillomavirus genotyping

L. Sara Arroyo[a], Vitaly Smelov[a,b], Davit Bzhalava[a], Carina Eklund[a], Emilie Hultin[a], Joakim Dillner[a,*]

[a] Department of Laboratory Medicine, Karolinska Institutet and Karolinska University Hospital, 141 86 Stockholm, Sweden
[b] Department of Urology and Andrology, Northwestern State Medical University n.a. I.I. Mechnikov, 191015 St. Petersburg, Russia

ARTICLE INFO

Article history:
Received 29 April 2013
Received in revised form 6 July 2013
Accepted 16 July 2013

Keywords:
Human papillomavirus
GS junior
Next generation sequencing
High throughput sequencing
Luminex

ABSTRACT

Background: Human papillomavirus (HPV) genotyping using next generation sequencing (NGS) could be useful to study the HPV variant-specific epidemiology, including monitoring for possible emergence of new HPV variants after introduction of HPV vaccination programs.
Objectives: We wished to design and validate a method for rapid HPV detection, typing and sequencing in clinical samples.
Study design: Plasmids of 15 different HPV types were mixed and serially diluted in human DNA in concentrations ranging from 1 to 100 copies per sample, amplified using the HPV general PCR primer pair PGMY and sequenced using 454 technology. Sixty cervical samples were tested both with the NGS-based method and with a comparison method based on genotyping using type-specific probes bound to fluorescent beads (Luminex). Thirty-three clinical samples were repeat tested using NGS to evaluate reproducibility.
Results: The NGS-based method correctly identified all 15 mixed HPV types when present in 100 copies/sample and 13/15 types when present in 10 copies/sample. For 36/60 cervical samples genotyping results using NGS and Luminex were identical. For 12/60 samples the NGS method was more sensitive than the Luminex test and most of the remaining discrepancies could be explained by the different type coverage of the assays. Reproducibility testing found complete or partial concordance in 30/33 samples.
Conclusions: NGS provides a sensitive and accurate method for genotyping of HPV. The fact that also the amplimer sequence is obtained could be important for studying the epidemiology of viral variants and monitoring of HPV vaccination.

1. Background

Because infection with oncogenic human papillomavirus (HPV) genotypes is a necessary risk factor for cervical cancer [1], HPV testing could improve the cervical screening programs and HPV vaccination could reduce the cervical cancer risk [2].

Multiple HPV genotyping methods are available [3], most of which are based on PCR amplification of conserved regions in the L1 gene by consensus primers followed by HPV genotyping identification by hybridization of amplimers to type-specific probes [4,5] or by amplimer sequencing [6].

Development of efficient methodology for sequencing of HPV amplimers could be important both for management of infections, for epidemiological studies of viral variants and for surveillance of HPV in monitoring of vaccination program effectiveness.

2. Objectives

This study aimed to detect and type HPV using next generation sequencing (NGS) of PCR amplimers, to investigate sensitivity and reproducibility of the method and to compare it with the HPV genotyping method used by the WHO HPV LabNet global reference laboratory.

3. Study design

3.1. Clinical samples

Sixty-two cervical samples were selected from Swedescreen [7,8], a population-based randomized controlled trial of HPV DNA testing in primary cervical screening. Also, 14 urethral and 11

Abbreviations: CIN, cervical intraepithelial neoplasia; emPCR, emulsion PCR; GS, genome sequencer; HPV, human papillomavirus; MID, multiplex identifier; NGS, next generation sequencing.

* Corresponding author at: Department of Laboratory Medicine, Karolinska Institutet, Karolinka University Hospital, Huddinge F56, 141 86 Stockholm, Sweden. Tel.: +46 76 8871126; fax: +46 40 337312.
E-mail address: joakim.dillner@ki.se (J. Dillner).

1386-6532/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jcv.2013.07.013

prostatic samples from a university outpatient clinic from St. Petersburg, Russia, were used.

3.2. Sample preparation

Cervical samples were collected following endo/ectocervical swabbing with a cytobrush. The brush was swirled in 1 ml sterile 0.9% NaCl afterwards and immediately frozen. After thawing and centrifugation (10 min at 3000 × *g*), the cell pellet was dissolved in 1 ml 10 mM Tris–HCl (pH 7.5) and frozen at −20 °C.

For urethral sampling, a urinary swab was first inserted into the urethra and rotated 180° right- and leftwards. Then, the swab was rinsed in 1000 μl phosphate buffer and stored at −20 °C. Before collecting prostate samples, patients were asked to urinate and a digital rectal examination with massage of the prostate was done. The prostate secretion dropping from the urethra was stored at −20 °C.

Sample handling retaining sterility was ensured for all sample types.

3.3. DNA extraction

Frozen samples were thawed and centrifuged for 10 min at 3000 × *g*. The pellet was dissolved in 1 ml 10 mM Tris–HCl (pH 7.5) and aliquots of 100 μl used for DNA extraction by a freeze-thaw-boiling procedure [9].

Amplification of the β-globin gene with real-time PCR evaluated sample adequacy [10]. β-Globin negative samples were extracted using SDS/Proteinase K [12], whereafter the real-time PCR was repeated.

3.4. Genotyping with Luminex

Five μl of the samples were amplified with MGP primers [11]. HPV detection and genotyping used multiplex bead-based hybridization with Luminex [12]. Biotinylated amplicons were hybridized to type-specific oligonucleotide probes coupled to the beads. After hybridization, streptavidin-F-Phycoerythrin binds to the biotin molecule. A 2-color laser allows recognizing HPV types by detecting the fluorescence of streptavidin-F-Phycoerythrin (identifying that hybridization occurred) and the color of the bead (identifying the HPV type).

Beads with probes for 13 oncogenic (high-risk, HR-HPV) types (16, 18, 31, 33, 35 39, 45, 51, 52, 56, 58, 59, 68a and 68b) and 23 non-oncogenic types (6, 11, 26, 30, 40, 42, 43, 53, 54, 61, 66, 67, 70, 73, 74, 81, 82, 83, 86, 87, 89, 90 and 91) were used. To control for possible contamination and accuracy, 11 negative controls (H_2O) and 8 positive controls (HPV plasmid pools) were run for every PCR and Luminex plate. All samples were tested twice.

3.5. NGS

Sixty cervical samples (45 HPV positive and 15 HPV negative samples) previously genotyped with Luminex, were selected randomly for the HPV type concordance evaluation.

We used the HPV general primer set PGMY for amplification, following the WHO HPV laboratory manual [13], but using non-biotinylated PGMY primers and without using HLA_DQ primers. PGMY primers amplify a 450 bp fragment within the L1 region (6500–7000 bp HPV genome region).

PCR was performed with 5 μl DNA in a 50 μl reaction containing 1.25U Amplitaq Gold, 1× PCR II buffer, 0.2 mM dNTPs, 80 nM each PGMY09, PGMY11 and HMB01 and 3 mM $MgCl_2$. A pre-heat of 95 °C for 9 min was followed by 45 cycles at 95 °C 30 s, 55 °C 1 min 30 s and 72 °C, 2 min, with a final extension at 72 °C for 5 min.

Amplicons were purified with MinElute PCR Purification kit (Qiagen, Germany) following the manufacturer's instructions.

Rapid library preparation followed the Rapid Library Preparation methods manual for 454 GS Junior Titanium (Roche). Because the amplicon length fits to the typical read length of 454 Titanium, no fragmentation of DNA was necessary. 500 ng of purified DNA was eluted in 16 μl of TE Buffer and amplimers were polished and end-repaired to create 3′A ends to allow TA ligation of library adaptors. A total of 240 different oligonucleotide adaptors (120 Adaptor A and 120 Adaptor B) were obtained from Integrated DNA technologies (IDT, Iowa) and 120 multiplex identifiers (MIDs) were prepared according to Technical Bulletin 2010-010 (Roche). Each DNA sample was ligated to a different MID (60 different MIDs were used in this study), according to the manual.

After adaptor ligation, small DNA fragments were removed and samples were purified by the AMPure beads (Agencourt Bioscience Corporation, USA), quantified and diluted to 1×10^7 molecules/μl.

Three pools of 20 DNA libraries (each DNA library corresponding to one sample) were combined with capture beads (an input of 2 molecules of each library DNA per capture bead) and then amplified by emulsion PCR (emPCR), following the emPCR Amplification Manuals for 454 GS Junior Titanium. After bead recovery and bead enrichment, the bead-attached DNAs were denatured and sequencing primers were annealed.

A single clonally amplified DNA bead was deposited per well in the Pico Titer Plate device. Sequencing followed the manual for the GS Junior Titanium.

If NGS detected HPV genotypes that had not been detected by Luminex, the NGS was repeated.

3.6. NGS sensitivity

The same HPV plasmid stocks that are also used in the WHO HPV LabNet global proficiency panel [14,15] (HPV 6,11, 16, 18, 31, 33, 35, 45, 51, 52, 56, 58, 59, 66 and 68) pooled at concentrations of 0.2, 2, 5, 10, 15 and 20 copies/μl diluted in TE (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0) with human placental DNA (10 μg/μl) were used for the sensitivity assessment.

3.7. NGS reproducibility

Thirty-three clinical samples were selected for reproducibility testing. DNA extraction was carried out by freeze/thaw/boiling procedure followed by E.Z.N.A.™ HP Viral DNA/RNA kit (Omega BIO-TEK, Germany), according to the manufacturer's instructions, scaling 10:1 the volume of samples, reagents and buffers. PCR was performed with 5 μl DNA in 25 μl reaction containing 1.5U Amplitaq Gold, 1× PCR buffer, 0.2 mM dNTPs, 80 nM each PGMY09 primer, HB01 and non-biotinylated PGMY11 and 4 mM $MgCl_2$. The thermal program was as described [16].

Library preparations were performed twice in order to assess reproducibility.

3.8. Analysis of NGS data

Bases from the 454 sequencing with a Phred quality score less than 30 were considered as ambiguous and trimmed off. A Phred quality score of 30 is equivalent to a base calling accuracy of 99.9% (1 error in 1000 bases) [17].

After removal of primer sequences, reads with a length of more than 80 bp and less than 20% of ambiguous bases were analyzed for human and bacterial DNA using SSAHA2 software [18]. Reads with at least 95% identity over 75% of their length to human or bacterial DNA were removed. Remaining sequences were assembled to contiguous sequences (contigs) using MIRA [19] software (with the parameters '-job = denovo, genome, accurate, 454 454_SETTINGS