Elucidation of integrated HPV DNA structure in a cervical carcinoma cell line by combined high resolution and long range sequencing analyses

Anne Van Arsdale MD MS[a], Jack Lenz PhD[b], Nicole Patternson[b], Elaine Maggi[b], Mark H. Einstein MD MS[c], Dennis Y. Kuo MD[a], and Cristina Montagna PhD[b]

[a]Department of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center, Bronx, NY
[b]Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY
[c]Rutgers-New Jersey Medical School, Department of Obstetrics, Gynecology & Women's Health, Newark, NJ

LB-334

BACKGROUND

- Integration of human papillomavirus (HPV) DNA into the human genome after viral infections of the female reproductive tract is a near ubiquitous factor in cervical tumorigenesis.[1-3]
- HPV DNA integration generates viral-host DNA junctions at the genomic sites of insertion. These junctions are frequently accompanied by local rearrangements of human DNA ranging from small to large indels, other chromosomal rearrangements, and establishment of episomal viral-host genome fragments.
- Next-generation sequencing (NGS) methods provide powerful tools for understanding of the role of HPV DNA integration in tumorigenesis, most notably identification of human genome oncogene targets. While short read approaches are potent means for junction identification, they can be susceptible to ligation and polymerase-jumping artifacts and are unable to elucidate genomic complexity near the junctions.[4]
- Long read, nanopore sequencing (Oxford Nanopore Technologies, ONT) can generate single molecule template, sequencing reads from tens to hundreds of kilobases and can thereby provide a tool to resolve complex genomic rearrangements accompanying HPV DNA integration.

OBJECTIVES

- To combine HPV16-targeted hybridization capture-Illumina NGS with long read (ONT) sequencing to CaSki. CaSki is a well studied cervical carcinoma cell line containing many HPV16 insertions. Our ultimate goal is to map with single nucleotide resolution the genomic rearrangements associated with HPV16 integration.

METHODS

Targeted hybridization capture/NGS:

•We designed a custom HPV DNA capture assay using the Roche Nimblegen SeqCap EZ system[5] targeting the full length genomes HPV16 and other high risk HPV types.

•CaSki genomic DNA (gDNA) was extracted using the QIAamp DNA mini kit (Qiagen), mechanically fragmented to 900 bp (Covaris), ligated to Illumina adapters, hybridized to our custom probe panel and sequenced by Illumina HiSeq 2500 using the paired end 300 bp mode.

•Read pairs were aligned using BWA MEM[6] to a custom human genome assembly of human GRCh37/hg19 plus HPV16 reference genome. Junction fragments were computationally identified using Delly[7] and SplazerS[8].

Nanopore Sequencing:

•CaSki gDNA was mechanically sheared to 20 kb using the g-TUBE (Covaris) and libraries were prepared using the Oxford Nanopore 1D ligation library prep kit SQK-LSK108.

•Libraries were loaded onto one R9.4 flow cell and sequenced on MinION Mk1b device (ONT) using the standard 48 hour scripts. Post sequencing base calling and fastq file extraction were performed with Albacore v2.0.

•Reads were aligned to the custom reference genome using Ngmlr[9] and structural variants called using Sniffles[9] with parameter adjustment for expected low coverage.

Fluorescent in situ hybridization (FISH):

•Two BAC clones (RP11-348M15 and RP11-625K18) flanking the 5' and 3' ends of one HPV16 integration mapping to chromosome 10 identified by NGS and long range sequencing were fluorescently labeled to visualize both junctions of the integration event. A plasmid containing HPV16 full length sequence was also labeled and co-hybridized to CaSki cells.

•Signals for HPV16 and the 5' and 3' ends of the chromosome 10 junction sites were visually inspected and evaluated in 26 cells.

a)

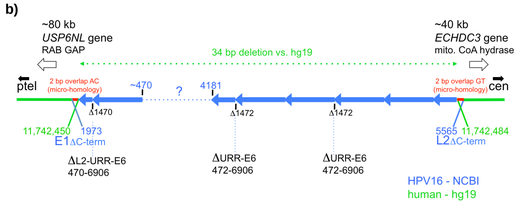
Illumina

Chromosome breakpoint	Chromosome strand	HPV breakpoint	HPV strand	HPV location	Number supporting reads
11,689,563	+	5631	+	L2	225
11,725,022	+	4636	+	L2	184
11,742,232	+	1468	+	E1	682
11,742,489	-	5706	-	L1	1209
11,812,562	+	5820	+	L1	8

Nanopore

11,742,450	+	1973	+	E1	3
11,742,484	+	5565	+	L2	3

b)



c)

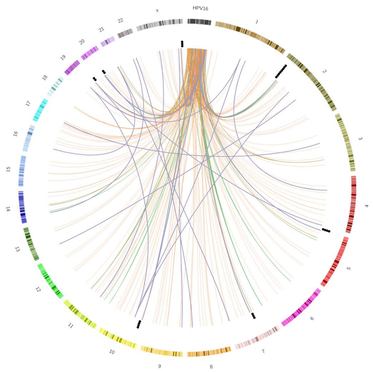
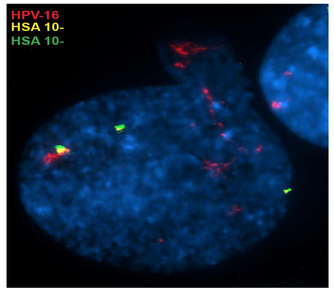


Figure 1. a) Positions of Chromosome 10, HPV16 junctions. **b)** Architecture of chromosome 10 integration based on ONT sequencing results. Human genome positions and distance were from hg19. **c)** Circos plot of inter-chromosomal rearrangements between the host genome and HPV16. Orange = Illumina novel junctions. Green = Illumina prior validated junctions. Purple = Nanopore junctions. Black = number of tandem HPV genomes by Nanopore long read sequences.

a)



b)

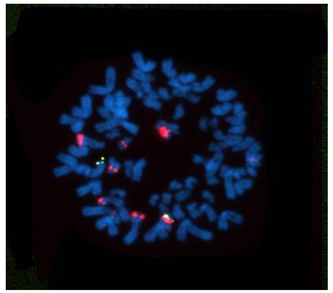


Figure 2. Representative FISH depicting CaSki cells: Red = HPV16, Yellow = RP11-625K18 (5' junction end), Green = RP11-348M1 (3' junction end) **a)** Interphase with trisomy chromosome 10, one copy with HPV16. **b)** Metaphase with chromosome 10 disomy, one copy with HPV16.

Phenotype	1	2	3	4	5	6	7	8
Chr 10 - Yellow	0	1	1	2	2	2	3	5
Chr 10 - Green	0	1	2	2	2	2	3	5
HPV 16 - Red	0	0	1	0	1	2	1	2

Table 1. Signal counts in n=26 CaSki cells

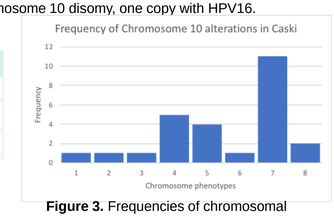


Figure 3. Frequencies of chromosomal phenotypes (n=26 cells)

RESULTS

- Illumina sequencing results: 34 million demultiplexed read pairs were obtained with average HPV16 coverage depth of 2.8 *10^6. All previously known and validated CaSki-human junctions (n=47) were identified; in addition we mapped n=193 novel junctions. Of the novel junctions, 4 out of 6 have been validated by PCR methods. Clustering of junctions at nearby human genome positions was identified for specific chromosomal locations, with n=30 (13.5%) called less than 1 kb apart on the human genomic side.
- Nanopore sequencing results: ~150K reads were obtained with a median length of 17kb [Range: 3-110kb]. Average coverage depth of HPV16 and the human genome was 220X and 0.60X respectively. Integrations were mapped to chromosomes 2,5,7,10,19,20, and X, confirming 4 additional novel junctions called by Illumina sequencing. All of these integration sites contained HPV16 in tandem arrays. The number of HPV tandem repeats ranged from 2 to 8 viral genome copies (Figure 1).
- FISH analysis confirmed the presence of HPV16 integration mapping to chromosome 10 (Figures 1 and 2). Only one cell failed to to show co-localization of both chromosome 10 probes, thus indicative of a low false negative frequency. However, both HPV16 signals at the chromosome 10 locus and chromosome ploidy varied widely among the cells. 8 different chromosome 10 phenotypes were identified (Table 1 and Figure 3), strongly suggestive of ongoing genomic variation within this cultured cell line. The most frequently observed phenotype was 3 copies of chromosome 10 with 1 of those copies harboring the HPV16 insertion.

CONCLUSIONS

- The hybridization capture-NGS approach here described is highly sensitive and specific for detecting and mapping individual HPV-human DNA junctions. Combining this molecular approach with long read, ONT analysis generated a direct examination of the tandem arrays of HPV16 DNA in the CaSki cell line.
- The clusters of HPV-human DNA junctions at nearby positions within the human genome for many HPV insertion sites are suggestive of frequent, ongoing, genomic instability. Sites with fewer supporting reads on short read sequencing may reflect low frequency, subclonal populations rather than false positive results, as demonstrated by technical validation of a fraction of them using PCR or long range sequencing.
- FISH analysis of one integration site mapping to chromosome 10 identified 8 different phenotypes involving this locus, thus providing further evidence of ongoing, genomic instability at HPV16 integration sites.
- We are currently applying our combined molecular approaches to determine if these findings hold true in primary lesions from patients.

REFERENCES

1. IARC monographs on the Evaluation of Carcinogenic Risks to Humans. Vol 90: Human Papillomaviruses. Lyon (France); 2005.
2. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189(1):12-9.
3. Clifford GM, Smith JS, Plummer M, Munoz N, Franceschi S. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. Br J Cancer 2003;88(1):63-73.
4. Dyer N, Young L, Ott S. Artifacts in the data of Hu et al. Nature Genetics, 2016;48(1):2-4.
5. Dello M, Patel K, Maslov A, Marion RW, McDonald TV, Cadoff EM, Golden A, Greally JM, Morrow B, Montagna C. Development of a targeted multi-disorder high-throughput sequencing assay for the effective identification of disease-causing variants. PLoS One 2015. 10(7):e0133742.
6. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60.
7. Rausch T, Zichner T, Schlatti A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333-9.
8. Emde AK, Schulz MH, Weese D, Sun R, Vingron M, Kalscheuer VM, Haas SA, Reinert K. Detecting genomic indel variants with exact breakpoints in single- and paired-end sequencing data using SplazerS. Bioinformatics. 2012;28(5):619-27.
9. Sedlazeck FJ, et al. Accurate detection of complex structural variations using single molecule sequencing. bioRxiv 169557.