Samplix

Detecting HPV18 integration sites in the human genome using Xdrop®

Esben Bjørn Madsen[1], Thomas Kvist[1], Marie Mikkelsen[1], Ida Höijer[2], and Adam Ameur[2]

[1] Samplix Aps, Birkerød, Denmark

[2] Science for Life Laboratory, Uppsala, Uppsala University, Sweden

Introduction

Targeted DNA enrichment is often necessary to sequence viral integration sites within complex DNA regions. However, available enrichment methods often fail with high-complexity, unknown, or rearranged regions.

Our Xdrop technology encapsulates long DNA fragments in double-emulsion droplets for highly targeted enrichment that supports such complex sequencing.

Experimental setup for HPV18 integration enrichment

The HeLa cancer cell line HPV18 integration sites were enriched by designing a single primer pair (amplifying the Detection Sequence) on the HPV18 virus sequence (Figure 1).

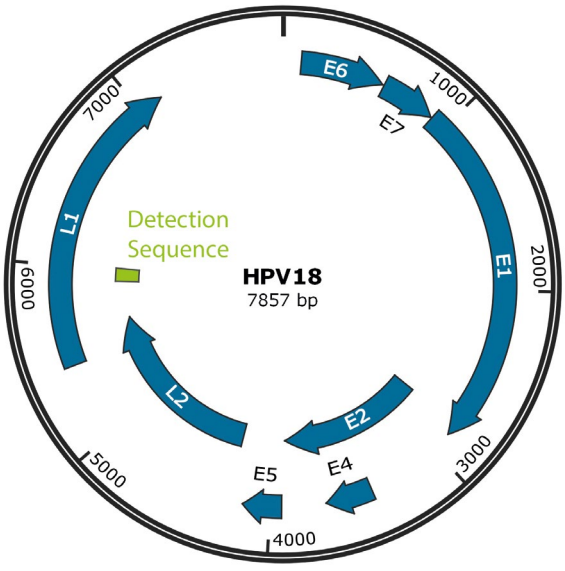

Figure 1. The circular genome of HPV18, 7857 bp. Viral open reading frames are depicted as blue arrows. The location of the Detection Sequence is shown in green.

Library preparation and sequencing

DNA libraries were prepared directly from Xdrop® enriched DNA samples and sequenced on three platforms: 1) PacBio RSII, 2) Oxford Nanopore MinION, and 3) Illumina MiSeq.

The sequencing reads were mapped to both HPV18 and the human genome (GRCh38) and three integration sites were detected on chromosome 8.

Mapping to the HPV18 genome

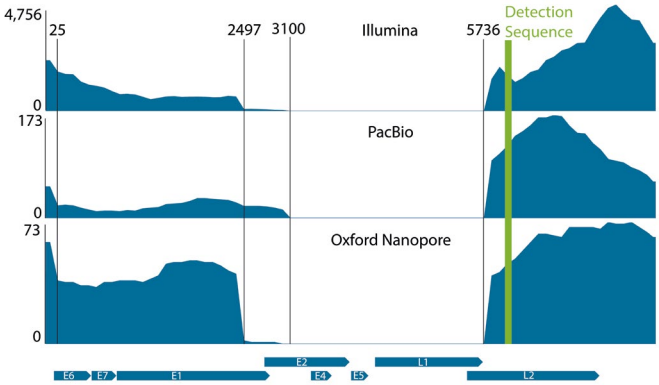

Figure 2. Mapping of sequenced reads by three different platforms to the HPV18 genome. The vertical black lines indicate the four breakpoints and their position in the HPV18 genome.

All three datasets (from PacBio, Oxford Nanopore and Illumina) showed that the central part of the viral genome (E2-L2 genes) was absent in the integrations, as previously reported (Adey et al. 2014) (Figure 2). The sharp drops in the HPV18 mapping coverage graphs above (positions 25, 3100 and 5736) are caused by sequence reads that have only one end mapping to HPV18, while the other end maps to an insertion breakpoint in the human genome. A fourth less frequent breakpoint with a lower coverage drop was identified at position 2497.

Mapping to the human chromosome 8

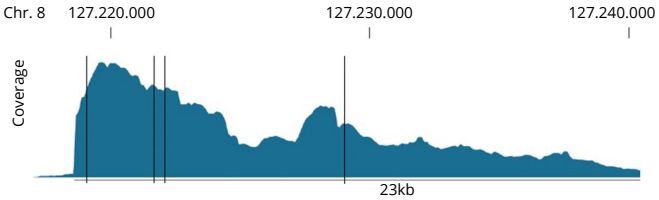

Figure 3. Coverage plot of reads mapped to a large region of the human genome on chromosome 8. Vertical black lines show the HPV18 integrations sites.

More than 20 kb of the viral integration site region in the human genome (HeLa) could be retrieved (Figure 3). Vertical lines pinpoint the integration sites. In total, regions of ~30 kb were characterized, demonstrating the efficiency of Xdrop® in enriching for long DNA fragments. The exact size of the enriched region could be even larger, as it is highly rearranged with multiple sequence repeats.

1000-fold enrichment with Xdrop®

The HeLa sample enrichment for HPV18 has been calculated to be 500-1000x fold as determined by sequencing (Figure 4).

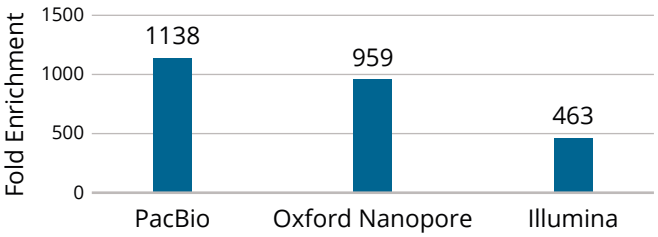

Figure 4. Fold enrichment of the region of interest compared to the input DNA. Measured by three different sequencing platforms.

Resolving HPV18 integration into the human chromosome 8

Detailed analysis of the reads spanning the breakpoints revealed the presence of several HPV18 integration sites in the region on chromosome 8 (Figure 5). The identified integration sites were supported by PacBio, Oxford Nanopore and Illumina datasets.

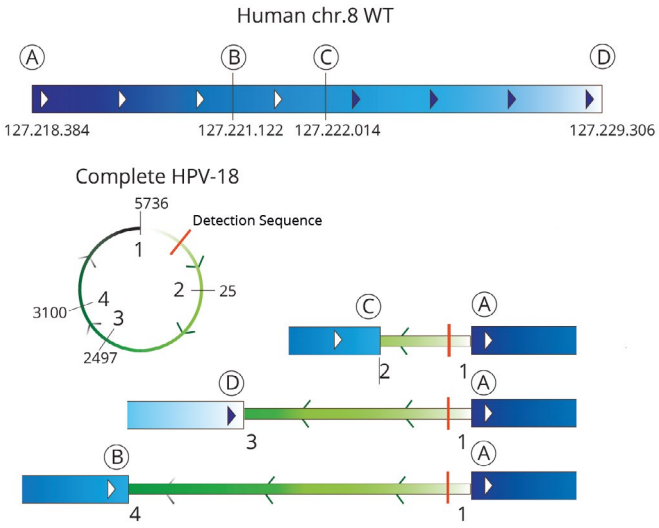

Figure 5. Top: Region of the human chromosome 8 where HPV18 is integrated in HeLa cells. Characters in circles show breakpoints. Middle: The circular HPV18 genome. The red line indicates the Detection Sequence. A single primer pair in the Detection Sequence was used to enrich the HPV18 integration sites and the surrounding HeLa genome. Below: Schematic representation of three inegration sites of HPV18 (green) into the human genome (blue) in HeLa cells.

Conclusion

The Xdrop® technology is a novel targeted DNA enrichment method with the unique feature that native DNA fragments can be selected, enriched and sequenced. This allows enriching for large genomic portions (~100kb), including unknown regions. The long fragments generated with Xdrop® method are not only suitable for long-read sequencing but also can provide valuable context information with short-read sequencing.

Peer-reviewed paper

This data is also published here:

Madsen EB, et al. (2020) Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting. *Hum Mutat*. 2020;41(9):1671-1679. doi:10.1002/humu.24063

Acknowledgements

This application was developed collaboration with Ida Höijer and Adam Ameur from Science for Life Laboratory, Uppsala, Uppsala University, Sweden, and partially funded by Eurostars and EU Horizon 2020 research and innovation program (project E!10942 DNANext).

References and notes

Adey A, Burton JN, Kitzman JO, et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature*. 2013;500(7461):207–211. doi:10.1038/nature12064