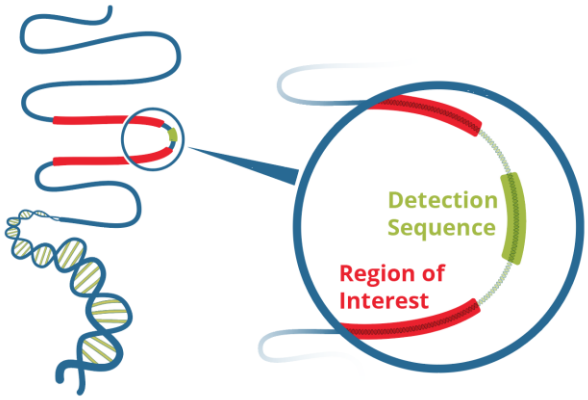
Samplix

Resolving complex structural variation with Xdrop®: Detecting HPV18 integration sites in the human genome

Background

Targeted DNA enrichment is often necessary when it comes to sequencing complex DNA regions or rare genetic elements. However, available enrichment methods often fail to enrich for high-complexity, unknown and rearranged regions.

Xdrop®, a novel DNA enrichment technology based on microfluidics, is able to enrich for long (~100 kb) regions of interest. This technology requires the design of a single primer pair amplifying the Detection Sequence, within the Region of Interest (ROI) or in flanking regions. This amplicon is exclusively used for detection, selection and enrichment of the ROI.



We evaluated the Xdrop® technology on the human cancer cell line HeLa, designing an enrichment experiment targeting integrations of the human papilloma virus 18 (HPV18) in the genome. Xdrop™ allowed to successfully enrich and identify the HVP18 virus integration sites in the human genome.

The Xdrop® Technology

The Xdrop® technology combines high-resolution droplet PCR (dPCR) with droplet sorting and Multiple Displacement Amplification in droplets (dMDA).

Firstly, Xdrop® partitions the DNA into millions of double emulsion droplets. Droplets containing the target DNA molecules are identified by a 120-160 bp droplet PCR specific to the Detection Seuence within or adjacent to the region of interest.

The detection and sorting of droplets are performed using a standard flow cytometer, which allows the PCR positive droplets containing the ROI to be collected. The sorted long DNA fragments are finally amplified in droplets (dMDA) to ensure unbiased DNA amplification.

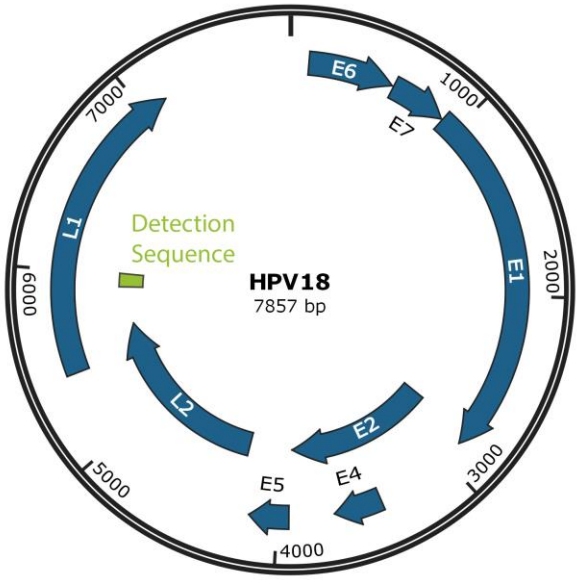
The Xdrop® enrichment and amplification technology are compatible with both long- and short-read library preparation and sequencing.

Main Applications of Xdrop® targeted enrichment

- Structural Variations
- Tandem Repeats
- GC-rich Regions
- Gap-closing
- CRISPR edits verification

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.

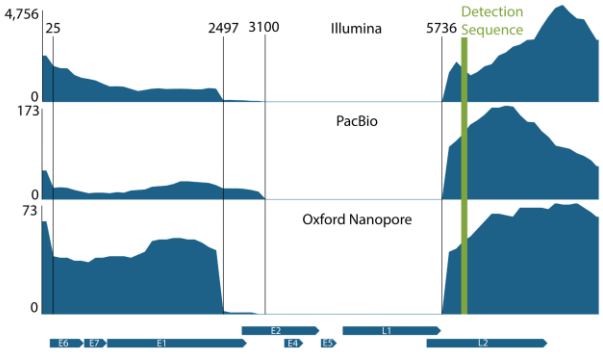


Library Preparation and Sequencing

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.

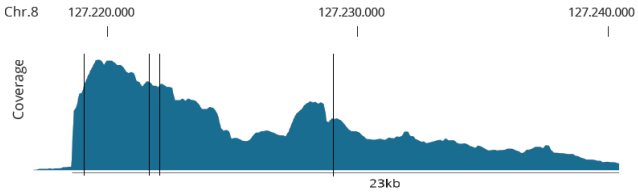
The sequencing reads were mapped to both HPV18 and the human genome (GRCh38) and three integration sites were detected on chromosome 8.

Mapping to the HPV18 Genome



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). 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.

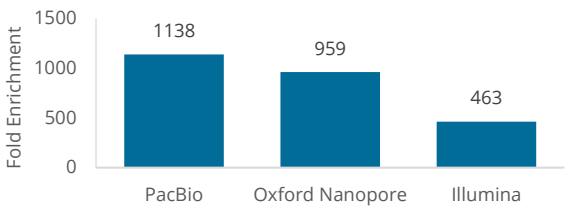
Mapping to the human chromosome 8



More than 20 kb of the viral integration site region in the human genome (HeLa) could be retrieved. 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.

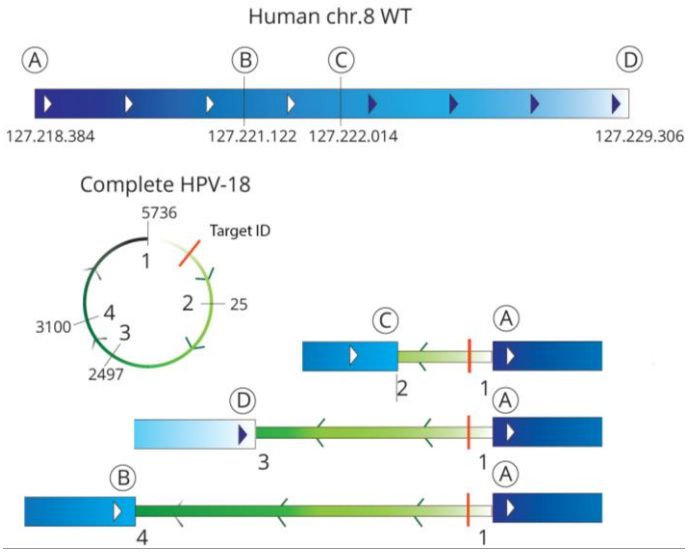
1000-fold enrichment with Xdrop®

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



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. The identified integration sites were supported by PacBio, Oxford Nanopore and Illumina datasets.



Conclusions

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

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