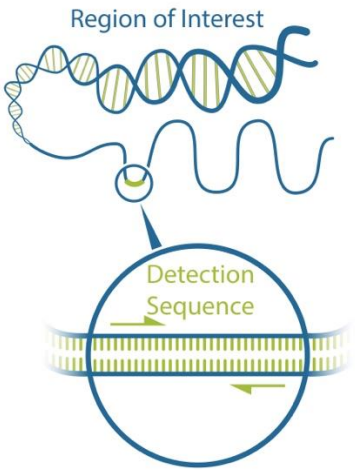
Resolving repetitive and GC-rich regions with Xdrop®: Indirect Sequence Capture on Epstein Barr Virus

Background

PCR and probe-based targeted DNA enrichment approaches require intensive design optimization and the knowledge of the Region of Interest (ROI) in full length. These conventional approaches also fail in characterizing complex genome contexts (e.g. repetitive regions, structural variations, GC-rich regions), as well as unknown and rearranged regions. To circumvent these limitations Samplix has developed a novel microfluidic approach, the Xdrop® enrichment workflow, based on Indirect Sequence Capture.

Xdrop® can enrich for long (~100 kb) target DNA regions requiring the design of a single primer pair over the Detection Sequence, corresponding to a small portion of ROI or in flanking regions. This amplicon is exclusively used for detection, selection and enrichment of the full-length ROI, which is ultimately captured and sequenced.



We evaluated the Xdrop® by designing an enrichment experiment targeting an unknown repetitive and GC-rich region of the Epstein Barr Virus (EBV) from a complex sample including human genomic DNA and EBV genetic material, mimicking an infected individual.

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 targeted dPCR specific to a sequence (Detection Sequence) 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
- Unintended CRISPR edits detection
- Integration of Transgene and Virus

Experimental Setup for EBV enrichment

Epstein Barr Virus (EBV) has been associated with several cancers, such as Hodgkin's lymphoma, Burkitt's lymphoma and gastric cancer. However, the genomic characteristics and diversity of the EBV is relatively poorly understood, also because of the low viral copy number in patient samples.

To address this, the Xdrop® technology was applied to enrich for EBV DNA using a primer-set targeting 89 bp on the EBV genome (Detection Sequence, Positions: 50165- 50251, Accession: NC_007605).

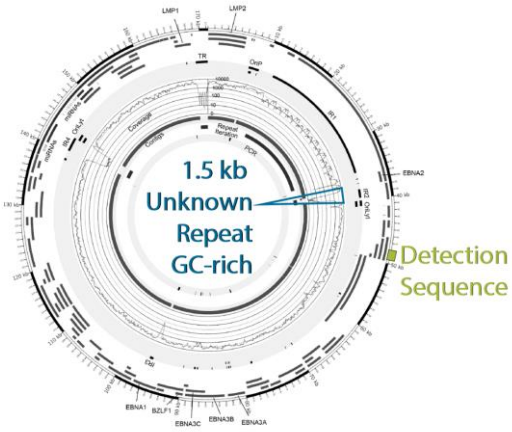
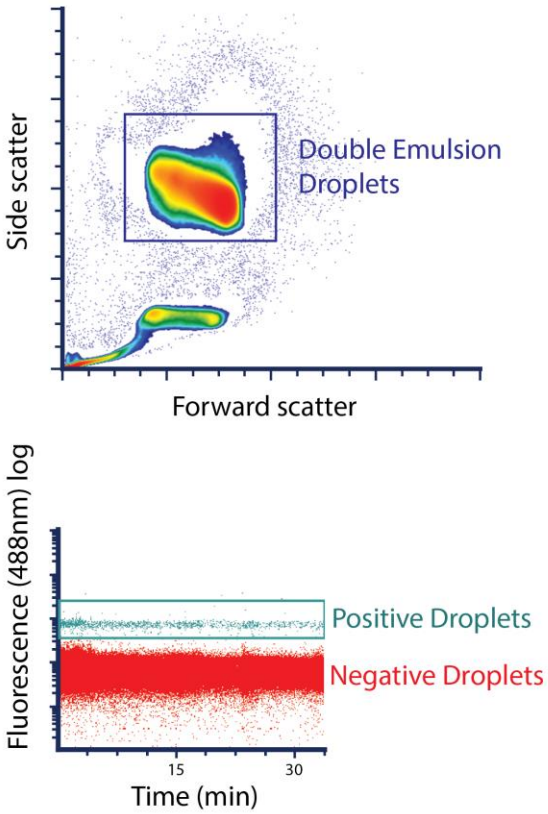


Image modified from Lin Z, et al. 2013.

Selection of positive droplets with flow cytometry

Double emulsion droplets were sorted using a BioRad S3e flow cytometer sorter. The top graph shows Side versus Forward scatter (height), with a clear cloud of double emulsion droplets (in the blue square). The bottom graph shows Fluorescence versus Time for the Double Emulsion Droplets selected. The top green cloud of Positive Droplets can be clearly distinguished from the Negative droplets cloud and sorted out.



Library Preparation and Sequencing

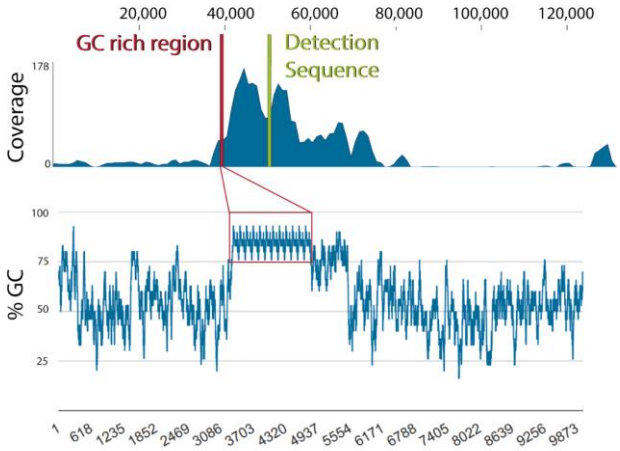
After sorting and dMDA amplification, DNA libraries were prepared and sequenced on PacBio RSII without size selection using the SMRTbell™ Template Prep Kit.

Mapping to EBV genome

The sequencing reads were mapped to EBV reference genome (NC_007605) and the unknown region was reconstructed by long-reads mapping.

Xdrop® enriched for a region of ~40 kb around the Detection Sequence, where the target selection primer pair were designed. The region included a 1,5 kb sequence with high GC-content (76-91% GC) consisting of 12 repeats, each 125 bp in length. The repeats are positioned in the EBV-IR2 region, approximately 10 kb from the enrichment primers. This specific region is known to be highly challenging in sequencing due to the combination of high GC-content and repeat sequences (Treangen et al. 2012).

The Xdrop® approach allowed to successfully characterize the region of interest, including the 1,5 kb GC-rich region using a single primer set amplifying the Detection Sequence for the detection and sorting of long fragments surrounding it (see figure below).



The ability to resolve such challenging regions with the Xdrop® technology opens the door to resolving many other challenging regions, which might represent a big proportion of genomes. Combining Xdrop® enrichment with long and short-read sequencing technologies can provide high resolution needed to resolve complex genomic scenarios with a quick and affordable approach, bypassing whole genome sequencing.

Conclusions

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

Acknowledgements

This application was developed collaboration with Ida Höijer and Adam Ameur from Science for Life Laboratory, Uppsala, Uppsala University, Sweden.

References

Lin Z, Wang X, Strong MJ, et al. Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. J Virol. 2013;87(2):1172–1182. doi:10.1128/JVI.02517-12

Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions [published correction appears in Nat Rev Genet. 2012 Feb;13(2):146]. Nat Rev Genet. 2011;13(1):36–46. Published 2011 Nov 29. doi:10.1038/nrg3117