APPLICATION NOTE

Esben Bjørn Madsen, Thomas Kvist, Marie Mikkelsen
Samplix ApS, Birkerød, Denmark

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

Summary

- Epstein Barr Virus (EBV) has been associated with several cancers as well as a wide range of other diseases. However, the genomic characteristics of the EBV are relatively poorly understood.
- Here, we use Xdrop for enriching a repetitive and GC-rich region of the EBV from a complex sample.
- We resolve the sequence of a previously unknown genomic region demonstrating that Xdrop enables the study of complex genomic regions.

Introduction

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, we have developed Xdrop, a novel microfluidic approach for Indirect Sequence Capture. Xdrop encapsulates the DNA fragments into millions of droplets that can then be sorted and collected. The DNA can be released from the droplets and amplified thereby enriching the ROI.

With Xdrop, we resolved the DNA sequence of 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.

Experimental Setup for EBV enrichment

Epstein-Barr virus (EBV) is associated with a variety of diseases ranging from infectious mononucleosis to various types of cancer. However, the understanding of the genomic features and variety of EBV is relatively inadequate, in part due to the low viral copy number in patient samples.

To address this, Xdrop was applied to enrich for EBV DNA in a sample of human genomic DNA mixed with EBV, using a primer-set targeting 89 bp on the EBV genome (Detection Sequence, Positions: 50165- 50251, Accession: NC_007605) (Figure 1). With Xdrop and a DE20 Cartrdige, DNA and primers were encapsulated together with PCR reagents in millions of picolitre double-emulsion droplets. DE20 droplets generated by Xdrop are uniform in size and compatible with thermal cyclers and flow cytometry.

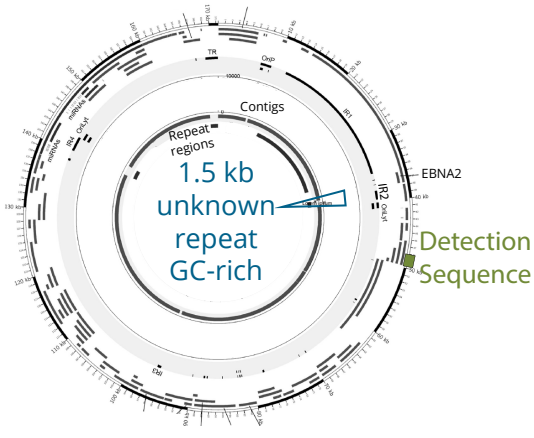

Figure 1. Representation of the circular EBV genome showing (from outside to inner circle) coding exons, regulatory elements, contigs generated by Lin et al., repeat regions. The detection sequence is shown in green. The location of the unknown repeat region is marked in blue. Modified from Lin Z, et al. 2013.

Selection of detection sequence-positive droplets with flow cytometry

After droplet PCR, the droplets were sorted using a Bio-Rad S3e flow cytometry cell sorter (Figure 2). All droplets were stained with a fluorescent dye (Samplix droplet dye). The double-emulsion droplets were gated from the total mix of all produced droplets and accompanying pure oil droplets based on their physical properties (forward and side scatter). The double-emulsion droplets containing the detection sequence were excited by the 488 nm laser and could easily be distinguished in the green channel.

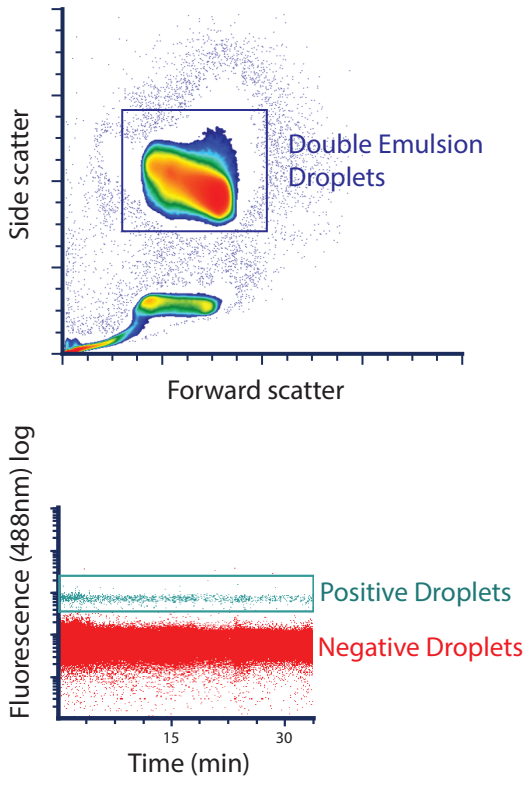

Figure 2. Flow cytometry sorting of double-emulsion droplets. Top plot shows the identification of double-emulsion droplets (blue box). Bottom plot demonstrates the fluorescence of gated double-emulsion droplets excited at 488 nm. A clear population of fluorescent droplets containing DNA from the detection sequence (positive droplets, green) can be discerned from the negative droplets.

Mapping to EBV genome

After sorting and droplet MDA amplification, DNA libraries were prepared and sequenced on PacBio RSII.

The sequencing reads were mapped to the 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 was designed (Figure 3). 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 to sequence due to the combination of high GC-content and repeat sequences.[2]

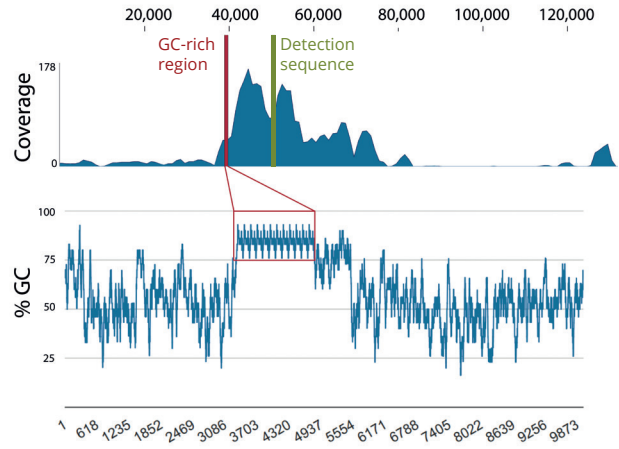

Figure 3. Mapping of obtained reads to the EBV genome. The green line shows the position of the Detection Sequence defined by the enrichment primer pair. The red line indicates the position of the 1.5 kb GC-rich repeat region, which is depicted closer in the bottom graph.

Conclusion

With Xdrop, we successfully characterize a complex region of interest in the EBV genome, including a 1,5 kb GC-rich region using a single primer set for the detection and sorting of long DNA fragments (Figure 3).

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

For more information about Xdrop products and applications, visit samplix.com.

References

1. 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
2. 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