Samplix

Xdrop™ vs. standard hybridization capture

Unparallel flexibility in target selection and more even coverage across repeats and GC-regions

Background

In this Technical Note, we compare Xdrop™ with standard hybridization capture, the most widely used approach for target enrichment to date. Our goal is to describe how Xdrop™ compares to a well-established technology in terms of performance and workflow.

Hybridization capture typically uses tiled biotinylated cRNA or DNA baits in a multiplexed manner to enrich for many short DNA sequences spanning several targets. Xdrop™ is based on **Indirect Sequence Capture, to enrich for long regions (~100 kb) surrounding a Detection Sequence (~150 bp).**

This makes the Xdrop™ very flexible, complementing standard hybridization capture technologies, with a better performance in case of difficult targets, i.e. GC-rich, repeat expansion, homologous genes, etc. Moreover, hybridization capture relies on short-read sequencing, while **Xdrop™ is compatible with both long and short read sequencing.**

Hybridization capture is the right choice to enrich many genes concurrently and in a high number of samples. Xdrop™, in contrast, is an **agile platform that allows you to easily switch from one target to another** and explore each with a handful of samples. This flexibility and throughput make Xdrop™ particularly well-suited for the new generation of personal sequencers, like iSeq and miniSeq.

The Xdrop™ workflow

The Xdrop™ technology combines Indirect Sequence Capture using high-resolution droplet PCR (dPCR) and droplet sorting followed by multiple displacement amplification in droplets (dMDA) to enrich specific long genomic regions in a sample.
If the molecular weight of the input DNA is > 60 kb, one **single Detection Sequence can capture ~100 kb of the surrounding region, with an enrichment of ~ 1000-fold.**

Xdrop™ partitions the DNA into millions of double emulsion droplets. During the dPCR, droplets containing the target DNA molecules accumulate amplicons of a 120-160 bp Detection Sequence within or adjacent to a region of interest, which is indirectly captured.

As a result of the amplified Detection Sequence, positive droplets emit fluorescence signal after staining and can be sorted by standard cell sorter. The sorted long DNA fragments are released, mixed with amplification reagents and then re-encapsulated individually for unbiased droplet multiple displacement amplification (dMDA).

The resulting enriched and amplified long fragments containing the region of interest are compatible with both long- and short-read library preparation and sequencing.

Overall experimental setup

Here we compare short-read sequencing data from three genes, enriched with hybridization capture probe and Xdrop™ platforms: BRCA2, CLU and DMPK. BRCA2 is a large gene with variable GC content and various repeats. CLU displays high GC content, while the DMPK includes Alu repeats and a trinucleotide repeat expansion. The fact that Xdrop™ is not only compatible with short-read, but also with long-read sequencing, represents an advantage for such *loci* associated to difficult mapping.
In the three gene comparison we present here we used NA14170 (BRCA2 6174delT) Cell line DNA for the BRCA2 comparison and HeLa Cell line DNA for the CLU and DMPK comparison. For hybridization capture, we used the SureSelect XT HS system and designed our probes with the SureDesign DNA Standard Design Wizard (https://earray.chem.agilent.com/earray/).

We applied the following probe tilling parameters in the design:

- Tiling density: 2x
- Masking: least stringent
- Boosting: balanced

The Tiling density was 2x, which indicates that designs were aimed at two probes to cover each nucleotide in each location. For the Masking we choose the Least Stringent type, which means only sequences masked by all three databases used to evaluate repeats in the SureDesign software (RepeatMasker, WindowMasker, and the Duke Uniqueness 35 track) would be

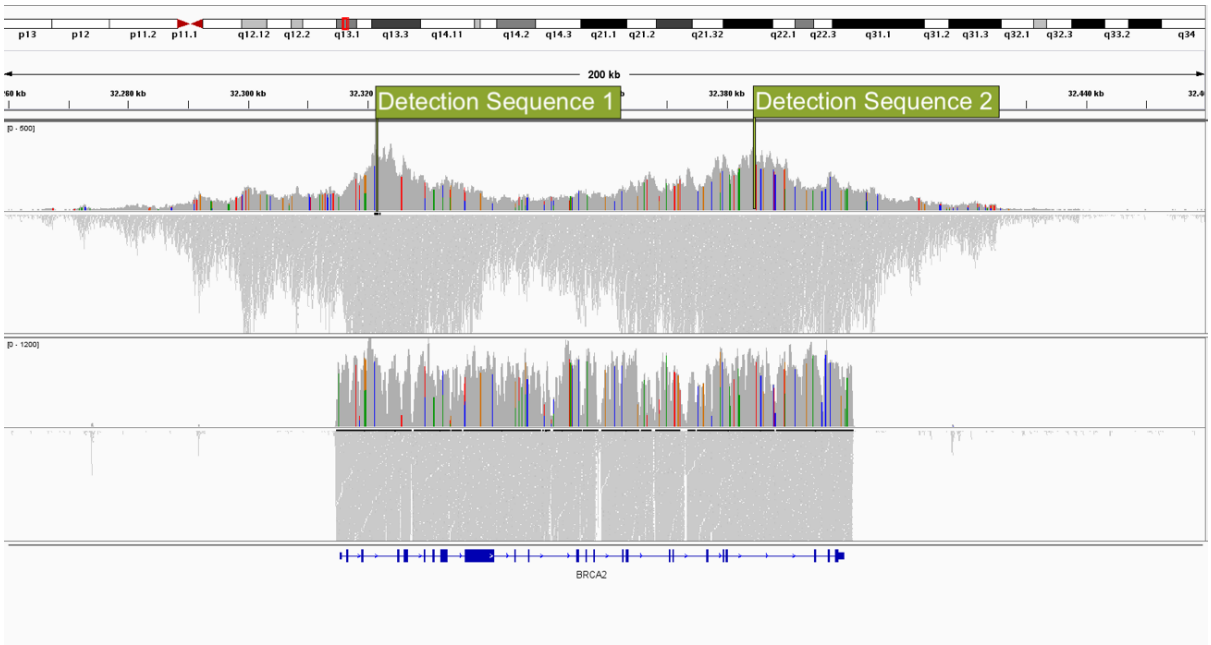
masked, hence forcing the design of probes in the other cases. For Boosting criterium we selected Balanced, which means that SureDesign replicates any GC-rich probes present in the probe group by an Agilent-defined factor, potentially improving the probe group's capture of genomic fragments. The SureSelect target capture and subsequent sequencing was performed by Eurofins Genomics A/S to the best of their standard. We designed primers for both the Detection Sequence and the Validation Sequence for the Xdrop™ workflow using the Samplix primer design tool available at samplix.com.

Note that Xdrop™ workflow including library preparation is PCR-free which increase chances of good coverage in difficult regions. Hybridization capture applies PCR during library preparation.

BRCA

Hereditary breast cancer is responsible for 5-10% of total breast cancer cases and up to 30% of early on-set breast cancer. BRCA2 is a tumor suppressor gene and one of the major susceptibility genes for breast cancer. Though breast and ovarian cancer are most commonly associated with BRCA2, it is known to increase susceptibility to other cancer types as well. BRCA2 is a large gene that covers ~85 kb on chromosome 13q12 and includes 27 exons that code for a protein that binds to DNA and is involved in homologous recombination-based DNA repair. Hundreds of BRCA2 mutations are known in various populations around the world contributing to cancer susceptibility.

For the BRCA2 comparison, we used genomic DNA from the Coriell cell line NA14170 (BRCA2 6174delT). To support a high coverage throughout the entire 85 kb of the BRCA2 gene, two Detection Sequences were designed for the BRCA2 Xdrop™ enrichment in contrast the other smaller genes we analyzed here which only required a Single Detection Sequence. Combined, the two Detection Sequences provided enrichment for a >150 kb stretch of DNA that included all of BRCA2 whereas hybridization capture just provided the ~85 kb "on-gene" coverage. Short-read sequencing presented here has an average depth of 500-600x for hybridization capture (~85 kb) and 100-300 for Xdrop™ (~150 kb) . Noteworthy is the considerable variation in sequencing depth across BRCA2 resulting from the hybridization capture enrichment compared to the very even coverage achieved with Xdrop™ (where less sequencing data was analyzed). The specific mutation BRCA2 6174delT was detected on both platforms.

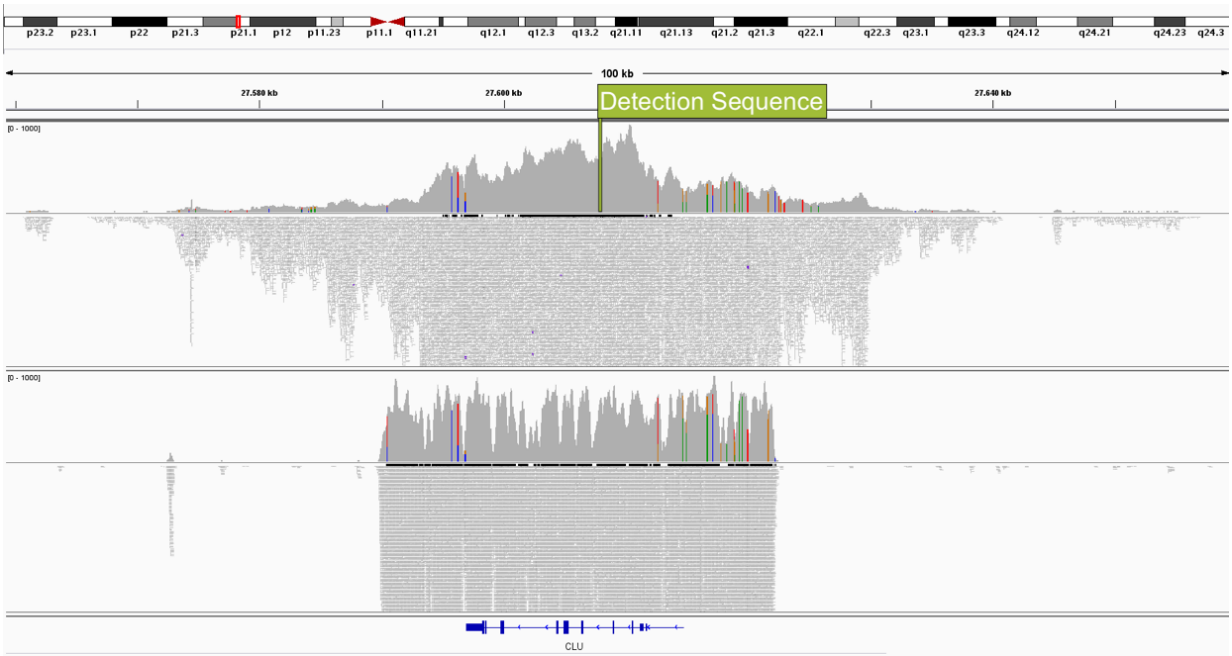


CLU

The CLU or Clusterin gene is ~20 kb in size and located on Chromosome 8p21. The gene is very GC rich and exhibits alternative splicing to generate two major protein isoforms. One localizes to the nucleus (CLU1) to induce apoptosis, while the second is secretory and inhibits apoptosis. Clusterin is implicated in Alzheimer's disease where it modulates pathways such as inflammation and apoptosis and acts as an amyloid-beta (Aβ) chaperone that alters Aβ aggregation and clearance.

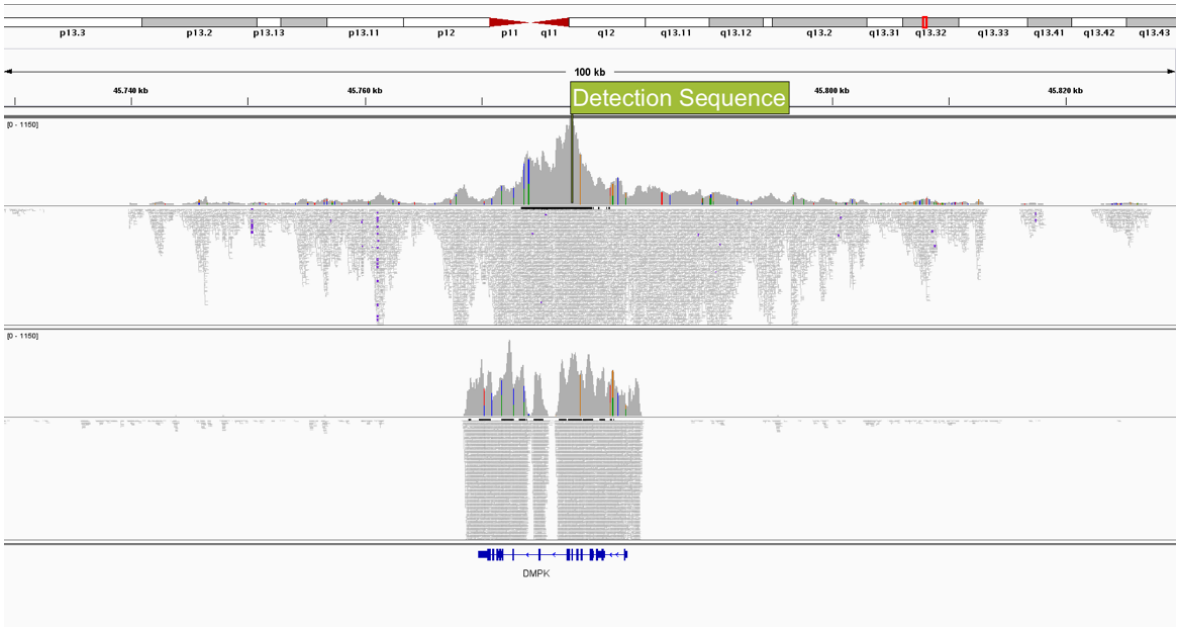
For the CLU comparison, we used genomic DNA from HeLa Cell line (New England Biolabs). With high CG content, CLU

can be problematic during enrichment and sequencing. However, in our comparison, both hybridization capture and Xdrop™ perform very well. Both methods in the figure below display an average 500–800x sequencing depth, though hybridization capture enrichment resulted in higher depth variation across the sequenced 20 kb DNA stretch. Xdrop™ generated very even coverage of CLU and captured sequence information from >80 kb around the target.



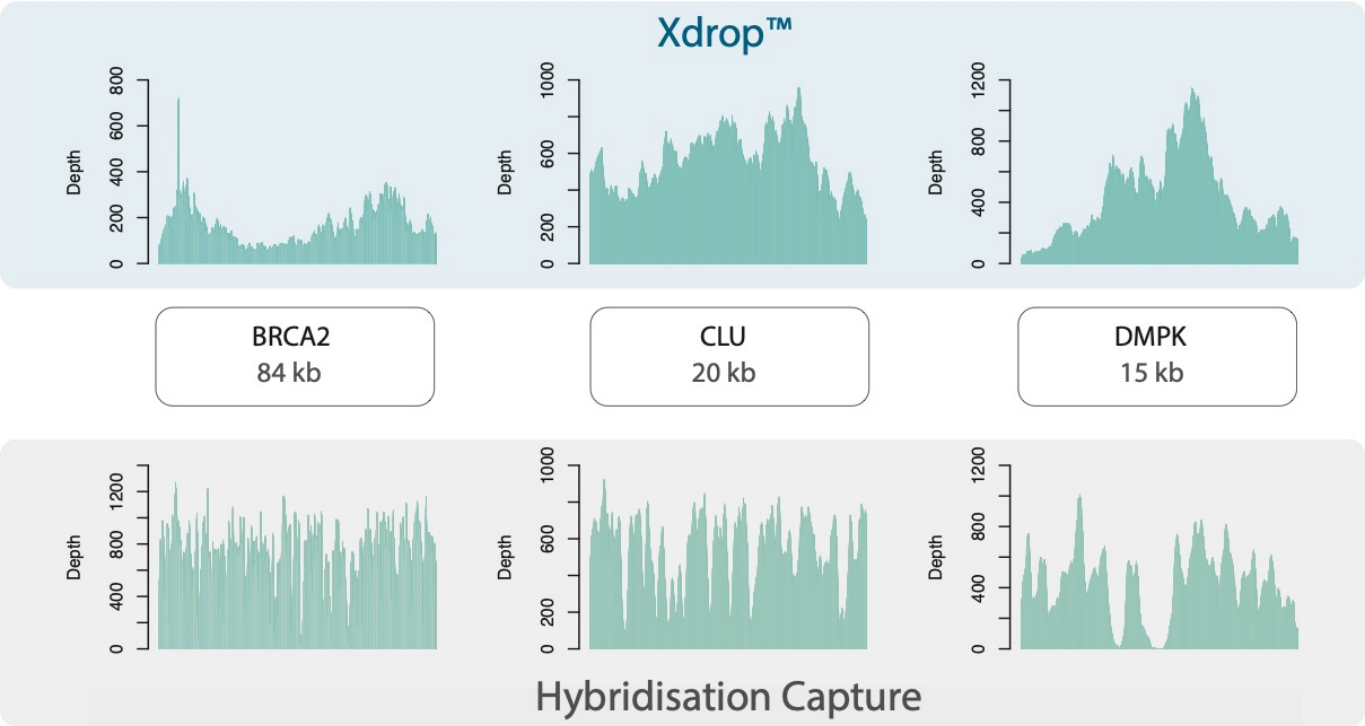
DMPK

The DMPK or myotonic dystrophy protein kinase gene is ~15 kb gene located on chromosome 19q13.32 and contains both Alu repeats and unstable CTG trinucleotide repeat. Located in the 3' untranslated region of DMPK, these trinucleotide repeats cause mytonic dystrophy type I when the normal 5–37 copies expand to 50–5,000 copies. Disease severity is correlated with copy number. Though the specific function of DMPK is unknown, it appears to play a role in muscle, heart, and brain cells. Sequencing depth along the 15 kb DNA stretch enriched with hybridization capture was overall successful but failed to cover central regions of DMPK that contain Alu repeats where the SureSelect design algorithm is unable to design capture probes. The Xdrop™ enrichment provided coverage across a genomic region >50 kb including DMPK. Both platforms display in the figure below a comparable average depth of 650–1000x for the target gene region.



Conclusions

Turnaround time from design to sequencing results could be as short as one week for Xdrop™, while 5-7 weeks are needed only for the design and synthesis of hybridization capture probes. Moreover, Xdrop™ can be exploratory applied to one or few samples with a **small investment** of few hundred dollars, while hybridization capture focuses high samples throughput (with min. 16 samples) and multiplexing, with higher associated cost. Moreover **Xdrop™ can successfully explore the broader genomic context of sequence variants including difficult regions, thanks to its indirect sequence capture technology and the compatibility with both long- and short-read sequencing.**



	Xdrop™	***Hybridization capture***
Sequencing read length	Short- and long-read sequencing	Short-read sequencing
Overage across GC rich regions and repeats	Even	Uneven
Flanking regions included	Yes	No
PCR	PCR-free (no bias)	PCR-based
Oligos required	2	Thousands
Turnaround time for custom assay	1 day	5–7 weeks
Cost per sample (without sequencing)	200 USD/sample	390 USD/sample (min. 16 samples)
Best use case	Flexible system to examine in detail few targets in <100 samples	Multiplexed system to sequence many (pre-designed) targets in many samples
Breadth of coverage	~100 kb per Detection Sequence (only one primer set needed)	Probe-dependent

1. Gnirke, A., et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotech. 27: 18