Incorporating sequence capture into library preparation for MinION™, GridION™ and PromethION™

Hybrid sequence capture allows users to select thousands of loci of interest simultaneously prior to sequencing, making more efficient use of the sequencing run

Contact: **publications@nanoporetech.com** More information at: **www.nanoporetech.com** and **publications.nanoporetech.com**

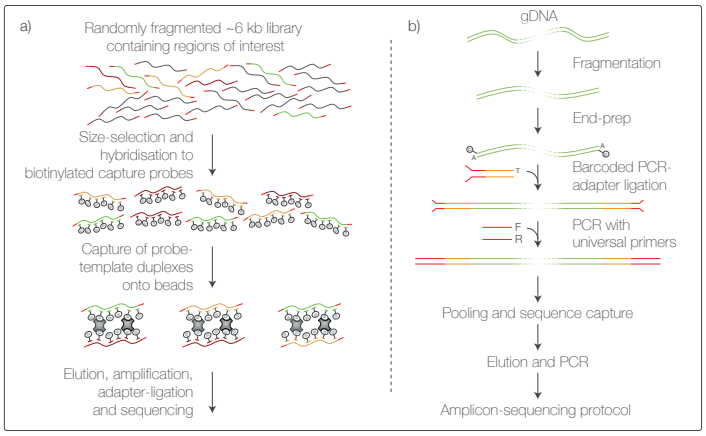

Fig. 1 Long-read sequence capture a) workflow overview b) multiplexed sequence capture

Sequence capture uses complementary probes to enrich for targets of interest

Sequence capture is a technique which allows the enrichment of specific regions of interest from a genome. It is useful when:
i) the user is not interested in analysing the entire genome
ii) the genome is too large for the throughput of the sequencer
iii) the user wishes to save money and time on sequencing and analysis
iv) the regions are longer than can be amplified by PCR, or too many PCRs would be required.

Sequence capture is performed during library preparation by hybridising the library fragments to probes which are specific to the regions of interest (Fig. 1).

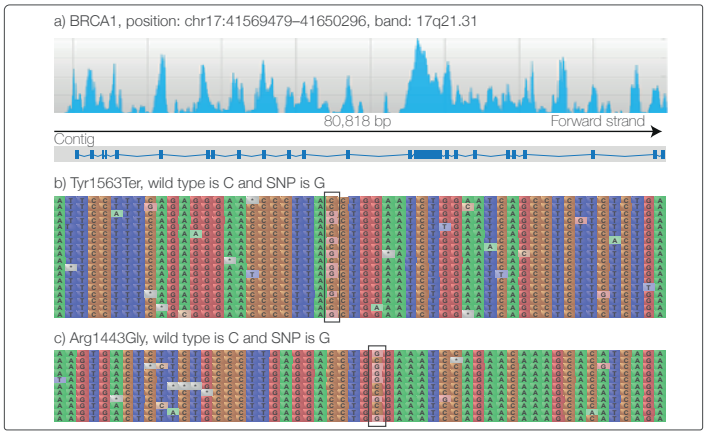

Fig. 3 Plots showing reads enriched for the BRCA1 gene and SNP-calling

Enrichment of Comprehensive Cancer panel genes, allowing detection of SNPs

We evaluated Agilent's ClearSeq Comprehensive Cancer panel using DNA from two BRCA1 SNP-carriers with a positive family history of breast cancer (NA13708 and NA13710, Coriell NIGMS Human Genetic Cell Repository). We performed sequence capture as described, sequenced the library and basecalled reads using the 1D workflow. We aligned all reads to the reference sequence from Agilent, using BWA with standard parameters, and visualised reads with Savant (Fig. 3a). Figs. 3b and 3c show alignments of reads from exon 16 and exon 13, respectively, of the BRCA1 gene. Sequencing errors are distributed randomly, allowing SNPs to be seen clearly in the data.

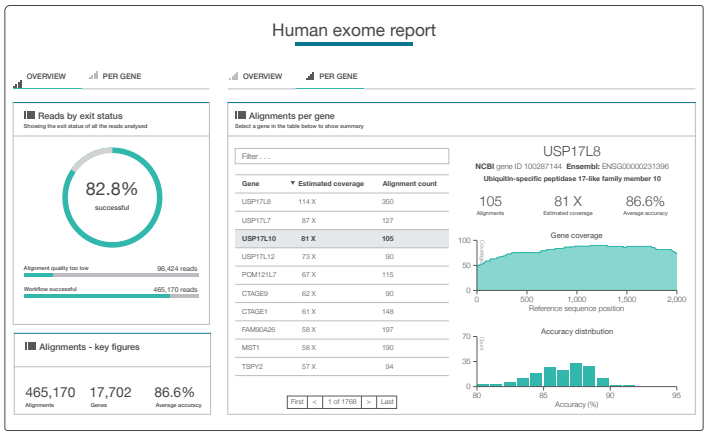

Fig. 2 Analysis report of sequence-capture data for the human exome

Resequencing analysis workflow for sequence-capture experiments

We have released an updated analysis workflow for sequence-capture experiments. To illustrate this, we captured the human exome using Agilent's SureSelect Human All Exon V6 panel, and generated ~ 2.35 Gb of 1D sequence data from a MinION run, representing ~40x average coverage. We analysed the data using the resequencing analysis workflow (Fig. 2). Following basecalling, reads are mapped to the human exome reference sequence. Individual gene information is displayed, including coverage and read-accuracy distribution at that position. Future releases of this application will support the uploading of target regions, highlight known SNPs in the target regions and allow those SNPs to be displayed with a confidence value.

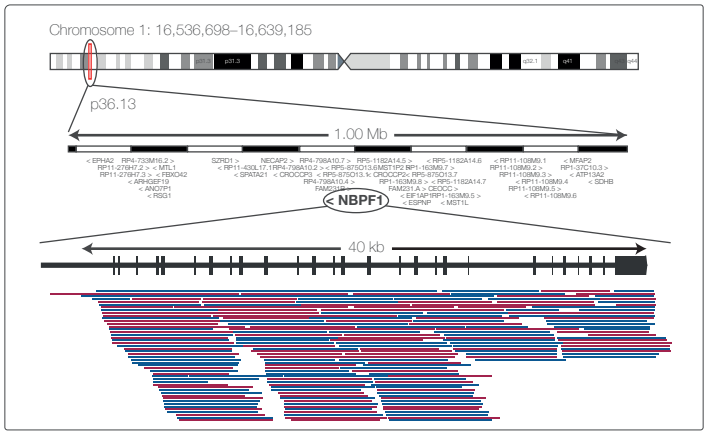

Fig. 4 Long reads mapping to exons 1–27 of NBPF1 after whole exome sequence capture

Long fragment capture protocol generates reads which can span multiple exons

We performed sequence capture on human genomic DNA, using Agilent's SureSelect Human All Exon V6, using a randomly fragmented library which had been size-selected using the Blue Pippin automated gel fractionation system, leaving only fragments above 6 kb. The library was then amplified by PCR prior to performing hybrid capture. The library was sequenced on a single run and the reads were mapped to the human genome (H19). Fig. 4 shows reads mapping to the neuroblastoma BPF1 gene, and it can be seen that multiple exons are spanned by many of the reads, meaning that both exon- and intron-specific variations can be detected and phased haplotypes can be obtained.