QIAseq® xHYB Long Read Hereditary Cancer Panel targeted-capture accurately detects large structural variants through a streamlined workflow and straightforward analysis pipeline

Nathan Blewett[1], Megan Zais[1], Jingxiao Zhang[1], Jixin Deng[1], John DiCarlo[1], Jamie Hill[2], Christa Haldrup[2], Sarah Kingan[3], Deborah Moine[3], Camille Connor[3], Matthew Fosbrink[1], Jonathan Shaffer[1]

[1]QIAGEN Sciences LLC, 19300 Germantown Rd., Germantown, MD 20874, USA; [2]QIAGEN Aarhus A/S, Kalkværksvej 5, 11., 8000 Aarhus C, Denmark;
[3]Pacific Biosciences (PacBio), 1305 O'Brien Dr., Menlo Park, CA 94025, USA

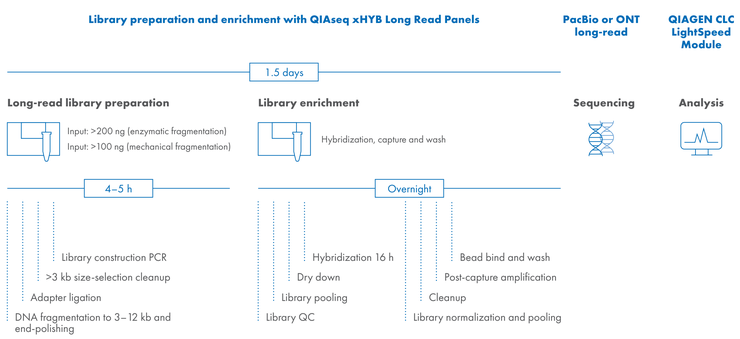
Large structural variant detection with the QIAseq xHYB Long Read Hereditary Cancer Panel

Introduction: Targeted long-read sequencing enables characterization of novel, complex mutations since long reads can span and be phased across large genomic regions. Long-read hybrid capture offers probe design flexibility for capturing targeted regions with unresolved structural variants. Here, we describe the development of QIAseq xHYB Long Read Hereditary Cancer Panel and chemistry for detection of large structural variants in 95 known cancer driver genes.

Methods: The QIAseq xHYB Long Read Hereditary Cancer Panel and analysis pipelines were used to identify large structural variants in genes involved in cancer progression that human reference DNA is known to harbor. Libraries were prepared using either enzymatic or mechanical fragmentation modules, after which targeted regions were captured with probes optimally designed for long DNA fragments. Captured DNA was amplified with chemistry developed for long, fast PCR and sequenced on both Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) platforms. The resulting long-read sequencing data was mapped and large structural variants detected with QIAGEN® CLC LightSpeed Module on the CLC Genomics Workbench.

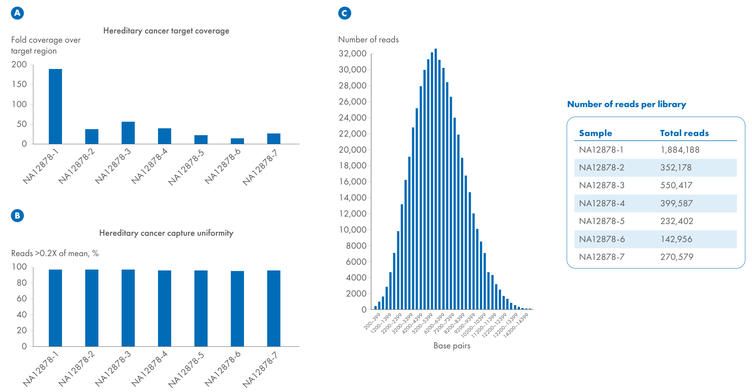
Results: Enzymatic and mechanical fragmentation produced libraries with average read lengths of 4.5 kb and 5.5 kb, respectively. Both fragmentation methods produce libraries with similar coverage and uniformity. Sequencing eight QIAseq xHYB Hereditary Cancer Panel libraries from one capture pool on a PacBio Revio® SMRT® Cell yields 30X average coverage, and uniformity greater than 95% of reads >0.2X of the mean. Sequencing the same libraries on the ONT MinION® platform for 24 hours produced 1 million reads, which resulted in 12X average coverage, and uniformity greater than 90% of reads >0.2X of the mean. Downstream analysis with the CLC LightSpeed Module clearly identified the expected large structural variants in different samples.

Flexible sample to sequencing workflow with two fragmentation methods



QIAseq xHYB Long Read capture workflow. The workflow begins with fragmentation via one of two methods – enzymatic or mechanical fragmentation. After fragmentation, libraries are end-repaired and ligated to a universal adapter and bead-based size-selection performed to remove DNA fragments smaller than 3 kb. PCR based sample indexing is performed during library construction, after which libraries undergo another round of size selection, and libraries are quality controlled for yield and size distribution. Indexed libraries are then pooled, concentrated, and hybridized with biotinylated capture probes overnight. After overnight hybridization, wash steps are performed to remove non-target DNA, post-capture amplification is then performed, followed by size selection and post-capture library QC for yield and size distribution. Captured libraries are now ready for attachment of long-read sequencing adapters.

QIAseq xHYB Long Read Hereditary Cancer Panels sequenced on PacBio Revio show excellent uniformity

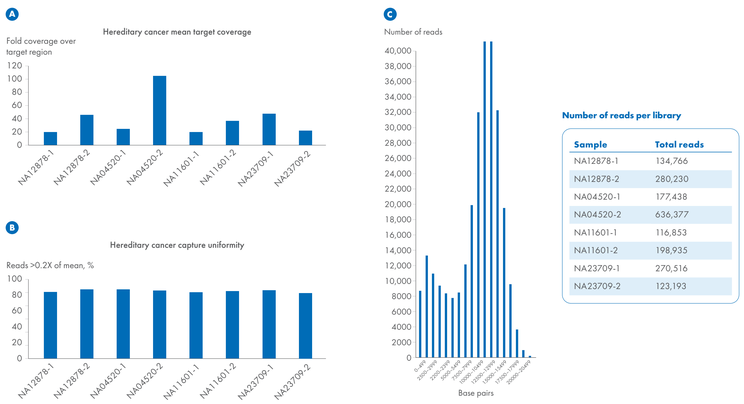


Number of reads per library

Sample	Total reads
NA12878-1	1,884,188
NA12878-2	352,178
NA12878-3	550,417
NA12878-4	399,587
NA12878-5	232,402
NA12878-6	142,956
NA12878-7	270,579

QIAseq xHYB Long Read Hereditary Cancer Panels sequenced at different depths on a PacBio Revio system. A Fold coverage over the targeted region for each library with the number of reads per library shown. **B** Library uniformity (percentage of reads covered at greater than 0.2X of the mean) for each QIAseq xHYB cancer capture library. **C** Histogram of SMRT Cell read-length distribution.

QIAseq xHYB Long Read Hereditary Cancer Panels sequenced on ONT MinION provide guidelines for coverage depth and uniformity



Number of reads per library

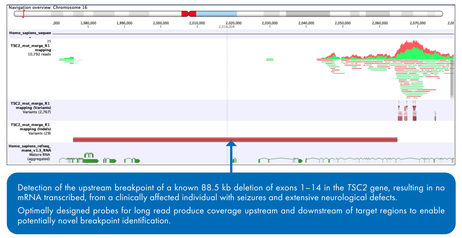
Sample	Total reads
NA12878-1	134,766
NA12878-2	280,230
NA04520-1	177,438
NA04520-2	636,377
NA11601-1	116,853
NA11601-2	198,935
NA23709-1	270,516
NA23709-2	123,193

QIAseq xHYB Long Read Hereditary Cancer Panels sequenced at different depths on an ONT MinION system. A Fold coverage over targeted region for each library with the number of reads per library shown. **B** Library uniformity (percentage of reads covered at greater than 0.2X of the mean) for each QIAseq xHYB cancer capture library. **C** Histogram of ONT MinION read-length distribution.

Detection of large structural and complex heterozygous variants

The QIAseq xHYB Long Read Hereditary Cancer Panel enables detection of large structural and complex heterozygous variants with QIAGEN CLC LightSpeed Module. Phasing with haplotagging using WhatsHap shows clear allelic separation. Two examples are shown below.

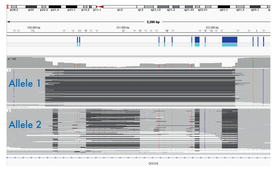
- *TSC2* which regulates mTOR signaling to control cell growth and division and plays an important role in neuronal homeostasis
- *DOCK8*, which regulates signal transduction in immune cells. *DOCK8* deficiency is associated with increased risk for lymphoma and squamous cell carcinoma



Detection of the upstream breakpoint of a known 88.5 kb deletion of exons 1–14 in the *TSC2* gene, resulting in no mRNA transcribed, from a clinically affected individual with seizures and extensive neurological defects.

Optimally designed probes for long read produce coverage upstream and downstream of target regions to enable potentially novel breakpoint identification.

QIAGEN CLC LightSpeed Module analysis of the *TSC2* gene in a clinically affected individual. A large deletion upstream of *TSC2* is shown. The 88.5 kb deletion was detected using QIAseq xHYB Long Read Hereditary Cancer Panel capture and CLC LightSpeed Module structural variant caller for long reads.



Integrative Genomics Viewer (IGV) report. IGV read stacks after phasing with WhatsHap, separated by alleles, from QIAseq xHYB Long Read Hereditary Cancer Panel capture using DNA from an individual with complex heterozygous structural variations in the known cancer driver gene *DOCK8*.

Conclusions

- The QIAseq xHYB Long Read Hereditary Cancer Panel produces even coverage through a robust workflow. It enables accurate detection of large structural variants in known cancer-driver genes with a user-friendly, long-read large structural variant calling pipeline using QIAGEN CLC LightSpeed Module on the CLC Genomics Workbench.
- QIAseq xHYB Long Read enzymatic and mechanical fragmentation options both produce long-read libraries that result in 4–6 kb average read lengths with robust uniformity and coverage that are compatible for sequencing on PacBio and Oxford Nanopore platforms.
  - After long-read sequencing, the known large structural variants were accurately identified. The probe design targets entire genes, including introns and UTR's, capturing extended regions of genes in an unbiased manner.
  - This allows for the detection of novel structural variants and breakpoints without prior knowledge.

QIAseq xHYB Long Read Panels are intended for molecular biology applications. These products are not intended for the diagnosis, prevention, or treatment of a disease.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit instructions for use or user operator manual. QIAGEN instructions for use and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services (or your local distributor).

Sample to Insight