QIAseq® xHYB Long Read HLA Capture: An Accurate, User-friendly, Scalable Long Read Solution for Class I and Class II HLA Haplotyping

Nathan Blewett[1], Megan Zais[1], Jingxiao Zhang[1], Jixin Deng[1], John DiCarlo[1], Sarah Kingan[2], Matthew Fosbrink[1], Jonathan Shaffer[1]

[1]QIAGEN Sciences LLC, 19300 Germantown Rd., Germantown, MD 20874, USA; [2]Pacific Biosciences (PacBio), 1305 O'Brien Dr., Menlo Park, CA 94025, USA

HLA haplotyping with QIAseq xHYB Long Read HLA targeted capture

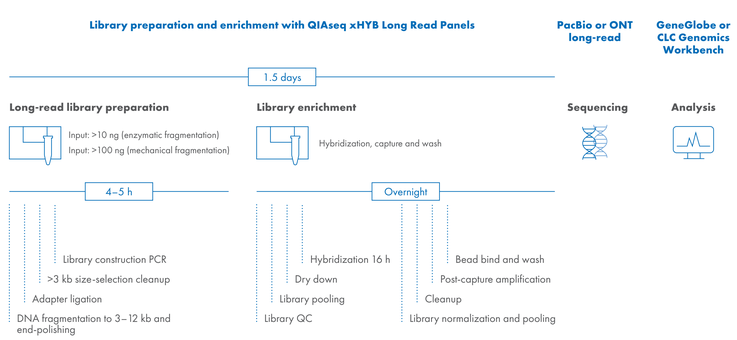
Accurate HLA haplotype phasing is important for predicting organ/stem cell transplant success, understanding autoimmunity and determining risk for several drug-induced adverse reactions. The highly polymorphic HLA locus encodes a group of cell-surface proteins that bind foreign antigens and modulate immune functions responsible for self vs. non-self immune regulation. Here, we detail the development of the QIAseq xHYB Long Read HLA Panel with indexed library construction and targeted HLA capture for accurate haplotype phasing and analysis. We present data on targeted HLA capture using PacBio® long-read HiFi sequencing on the Revio® system. The panel is also compatible with Oxford Nanopore platforms.

The optimized long-read amplification chemistry in our workflow achieves even coverage over high- and low-GC regions of the human genome. The Panel targets 19 class I and II HLA genes including introns, exons, and untranslated regions (UTRs), facilitating phasing and assembly of diploid HLA sequences and subsequent 4-field haplotype calling.

To assess workflow performance, we performed HLA capture on DNA from 58 different individuals (64 total libraries) with benchmarked, defined ground-truth haplotypes. Data showed that our HLA haplotyping approach is 97.7% accurate for HLA class I genes: HLA-A, HLA-B and HLA-C. Sequencing 64 individual HLA capture libraries on one PacBio Revio SMRT® Cell resulted in average read lengths of 4.5 kb – greater than 150X coverage for each of the 64 libraries – and excellent uniformity, with 99% of reads greater than 0.2X of the mean.

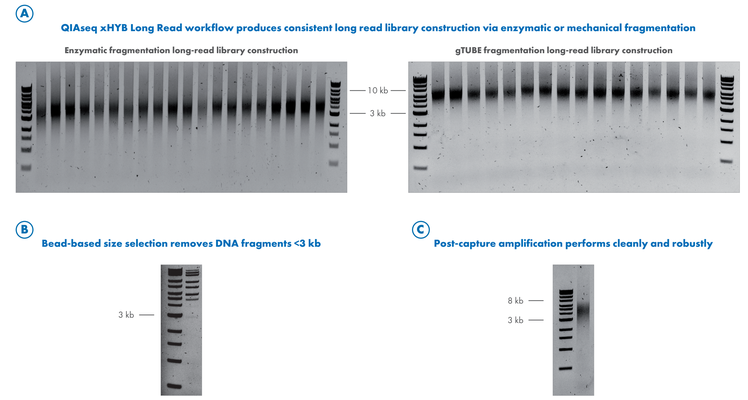
In summary, QIAseq xHYB Long Read HLA capture produces uniform data that only requires a total of 30,000 PacBio HiFi reads to achieve 30X coverage over the HLA target loci and 96 samples can be accurately haplotyped with one PacBio Revio SMRT Cell.

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 human capture workflow is robust



Consistent and robust human capture workflow. A 0.7% agarose gel analysis of typical library construction results from enzymatic fragmentation (left panel) and mechanical fragmentation (right panel). DNA ladder sizes indicated. **B** 0.7% agarose gel analysis of 48 kb DNA ladder that has undergone user-friendly bead-based size-selection without ethanol wash, DNA ladder sizes indicated. **C** Typical example of post-capture library size-distribution QC in 0.7% agarose gel, PacBio sequencing showed 4.5 kb N50 with reads ranging from 3 kb to 10 kb.

Long Read HLA panel targets 19 class I and class II HLA genes with 4-field resolution and even coverage over extended target regions

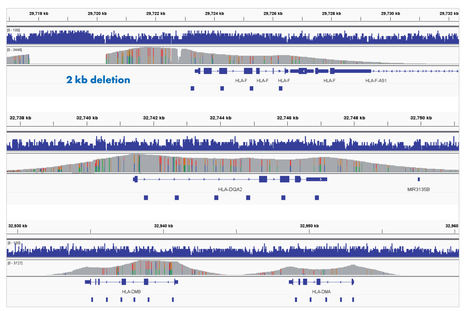
- Full HLA genes including introns, exons and UTRs are captured, providing 4-field haplotyping resolution.
- HLA genes are evenly covered with QIAseq xHYB long read capture, revealing clear allele separation and identification of unexpected structural variants.

HLA gene target and hg38 coordinates used for targeting probe design on chromosome 6

Class I	Start	End
HLA-A	29941260	29945884
HLA-B	31353875	31367067
HLA-C	31268749	31272092
HLA-E	30489509	30494194
HLA-F	29722775	29726526
HLA-G	29827825	29831021

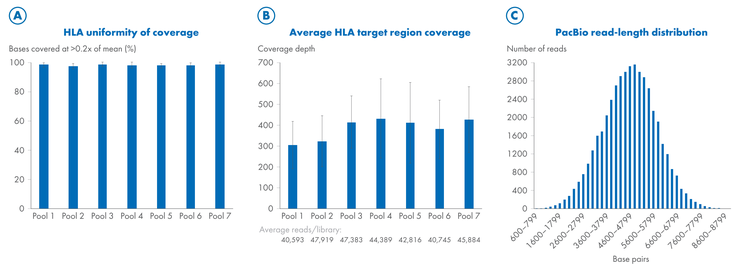
Class I	Start	End
HLA-DRA	32439887	32445046
HLA-DRB1	32578775	32589848
HLA-DRB5*	32517353	32530287
HLA-DQA1	32637406	32643684
HLA-DQA2	32741391	32747198
HLA-DQB1	32659467	32668383
HLA-DQB2	32756098	32763532
HLA-DOA	33004182	33009591
HLA-DOB	32812767	32820466
HLA-DMA	32948618	32953097
HLA-DMB	32934636	32941028
HLA-DPA1	33064569	33089696
HLA-DPB1	33064569	33089696

* HLA-DRB5 is not present in all individuals.



IGV viewing platform showing QIAseq xHYB Long Read HLA capture data coverage over selected HLA genes. Height of track denotes relative coverage, gray indicates reference base, color indicates variant base.

QIAseq xHYB Long Read HLA capture produces excellent uniformity and accurate HLA haplotyping



Excellent uniformity and accuracy. A Library uniformity (percentage of reads covered at greater than 0.2X of the mean) within each HLA hybridization pool (eight samples each). **B** Average coverage depth over the targeted HLA region for eight libraries within each hybridization pool. 64 total libraries were sequenced on one PacBio Revio SMRT Cell and the average number of reads per library is indicated for each hybridization pool. **C** QIAGEN® CLC Genomics Workbench mapping report histogram of HLA capture library size-distribution (N50: 4.5 kb) from PacBio sequencing on a Revio SMRT Cell.

Conclusions

- The QIAseq xHYB Long Read HLA Panel delivers even coverage through a robust workflow, and accurate haplotyping with a user-friendly downstream HLA haplotyping analysis solution.
- The panel targets 19 class I and class II HLA genes including introns, exons and UTRs, facilitating phasing and assembly of diploid HLA sequences and subsequent 4-field haplotype calling.
- We demonstrated that our HLA haplotyping approach is 97.7% accurate for HLA class I genes: HLA-A, HLA-B and HLA-C.
- Sequencing 64 individual HLA capture libraries on one PacBio Revio SMRT Cell delivered:
  - Average read lengths of 4.5 kb, greater than 150X coverage for each of the 64 libraries
  - Excellent uniformity, with 99% of reads greater than 0.2X of the mean
- Only 30,000 PacBio HiFi reads are required to achieve 30X coverage over the HLA target loci and 96 samples can be accurately haplotyped with one PacBio Revio SMRT Cell

HLA class I haplotyping accuracy

HLA gene	Accuracy (%)
A	99.2
B	97.7
C	96.1

QIAseq xHYB Long Read HLA capture was performed on DNA from 59 individuals with benchmarked haplotypes1. 64 HLA capture libraries were then sequenced on one Revio SMRT Cell. Samples were haplotyped with specHLA and 3-field accuracy determined relative to published, benchmarked haplotypes.

Reference

1. Bettinotti MP, et al. Characterization of 108 genomic DNA reference materials for 11 human leukocyte antigen loci: A GeT-RM collaborative project. *J Mol Diagn.* 2018;20(5):703-715.

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