Feasibility of long fragment hybrid capture for long-read sequencing: *Applications for HLA and beyond*

James Nhan, Chris Ventura*, Julio Avelar-Barragan, Ting Wang, JJ Chen

Transplant diagnostics business of Thermo Fisher Scientific Inc., West Hills, CA, USA

**Presenting Author*

Abstract

Purpose: Our aim was to synergize the best aspects of both long-read sequencing and hybrid capture technologies to overcome the current typing limitations relying on the use of long-range amplification methods or short-read hybrid capture methods. By combining hybrid capture with large DNA fragments, we can perform long-read sequencing on regions that have previous been unable to be sequenced in their entirety. These include Class II Loci of HLA and KIR which due to their length and novel mismatches in primer binding sites are challenging to sequence with current NGS methods.

Results: Various fragmentation methods were able to produce fragments of size >5kb in length. Long-fragment libraries generated by hybrid capture had an average size of 5kb. On-target analysis revealed that over 60% of captured fragments mapped to probe targets, average size of mapped fragments was >2kb. Genotypic analysis showed >99% concordance compared with reference typing.

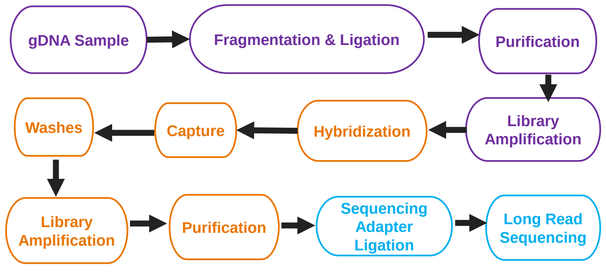
Introduction

Advancements in 3rd generation sequencing technology have made possible reliable sequencing of thousands to tens of thousands of base pairs. This long-read technology has enticed researchers and clinicians with the prospect of sequencing genes that are difficult to sequence and phase by short read sequencing, such as HLA. Current typing by long-read sequencing relies on an initial long-range amplification to enrich for full length HLA genes. Hybrid capture can enrich HLA targets of interest without the need for an initial amplification. Combining these methods allows for research and sequencing into longer stretches of HLA Loci and Chromosome 6.

Methods

Workflow Diagram

A. Long Fragment Library Prep and B. Target Enrichment and C. Long Read sequencing Library Prep



Testing Methods

Genomic DNA (gDNA) >50kb were enzymatically fragmented and long fragment sizes were measured using gel electrophoresis (Figure 1). Long fragments were subsequently barcoded, and adapter ligated. Samples were pooled and the Long fragment library amplified before proceeding through hybridization and capture. At each step fragment or library size was accessed using Tapestation (Figure 2 and Figure 3). Following 2nd Library amplification, the enriched library is taken through a long-read sequencing library preparation. After sequencing HLA typing was determined using TypeStream Visual (TSV) and coverage (Figure 4) concordance (Table 1) was accessed using custom software.

Data Analysis

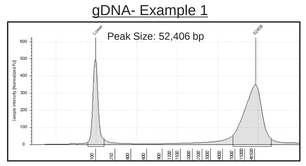
TypeStream Visual software was used to perform HLA typing and analysis. Custom software was used to determine sequencing metrics, coverage analysis and concordance.

Results

Figure 1. Electropherogram showing two examples of initial genomic DNA size

Two genomic DNA samples isolated from cell lines with known HLA typing. Figures showcase the large initial size pre-fragmentation. Good quality non-fragmented DNA is crucial to consistently large fragments via enzymatic fragmentation.

gDNA- Example 1



gDNA- Example 2

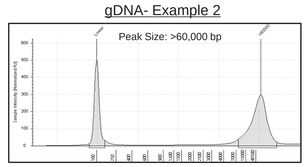
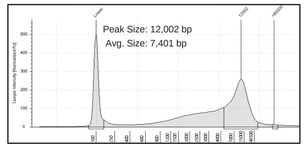


Figure 2. Electropherogram showing two examples of enzyme A fragmentation

Enzyme A was able to produce long fragments >10kb at sufficient yields to carry to hybrid capture. However, a large tail is present in both examples with sizes <3kb being carried forward to adapter ligation.

Enzyme A- Example 1



Enzyme A- Example 2

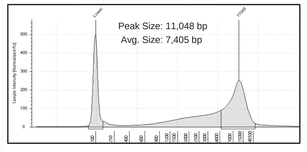
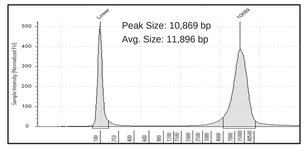


Figure 3. Electropherogram showing two examples of enzyme B fragmentation

Enzyme B was able to produce long fragments >10kb at insufficient yields to carry to hybrid capture. However, the large tail present in enzyme A was not observed with enzyme B. If we want to utilize enzyme B we have to overcome the challenge of low yields.

Enzyme B- Example 1



Enzyme B- Example 2

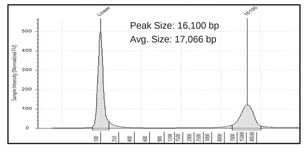


Figure 4. Electropherogram showing example of pooled long fragment library fragmented with enzyme A

Using 8 DNAs fragmented with enzyme A, a pooled library was prepared for hybrid capture. Pooling was necessary to provide sufficient input into the hybrid capture protocol. Post adapter ligation and bead clean-up, the peak and average size of the library dropped ~50%.

Pooled Pre-Hybrid Capture Library

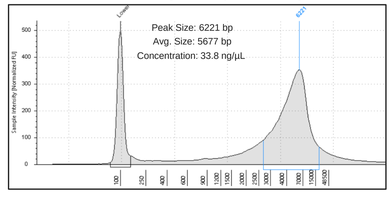


Figure 5. Electropherogram of enriched post-capture long fragment library

Following target enrichment via hybrid capture the library is amplified and purified with SPRI bead clean up. The library has less of a "tail" of smaller fragment sizes without a significant loss in library size from the pre-capture library shown in Figure 4. However, there was a drop in mean concentration.

Pooled Post-Hybrid Capture Library

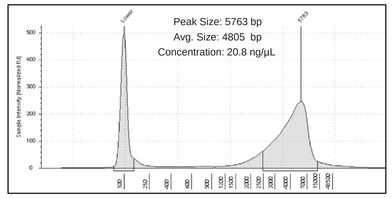


Figure 6. Electropherogram of final long read sequencing library

The enriched long fragment library was taken through a long-read sequencing workflow to create a final library for sequencing. Final library was further cleaned up and has fewer small fragments than in Figure 5, while peak and average size remain consistent. Slight increase in size was observed due to the sequencing adapter which could add to the length or slow migration on the gel. There was a significant drop in concentration of ~60% compared to Figure 5

Pooled Post-Hybrid Capture Library

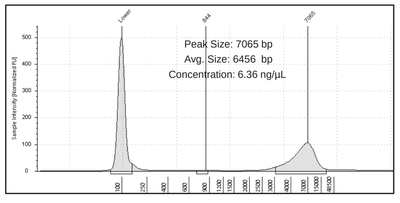


Figure 7. Sequencing read length from long range sequencing run

Once sequenced we observed that average size of sequencing adapted fragments was smaller (~5kb) compared with the post capture library (~6.5kb). This is likely due to limitations with the enzymes and workflow required for long read sequencing workflow.

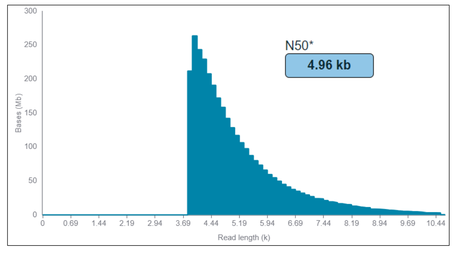


Table 1. HLA concordance table for long fragment sequencing library

# Samples = 8	A	B	C	DPA1	DPB1	DQA1	DQB1	DRB1	DRB345	Total
Perfect match	15	9	0	9	12	11	8	11	12	87
Ambiguous	1	7	16	7	4	5	8	5	3	56
Highly Ambiguous	0	0	0	0	0	0	0	0	0	0
Extremely Ambiguous	0	0	0	0	0	0	0	0	0	0
Extra Call (excluded)	0	0	0	0	0	0	0	0	0	0
No Call	0	0	0	0	0	0	0	0	0	0
Discordant	0	0	0	0	0	0	0	0	1	1
TOTALS	16	16	16	16	16	16	16	16	16	144
Concordance Total = 99.31% (143/144)	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	93.75%	99.31%
Concordant with Low Ambiguity Total = 99.31% (143/144)	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	93.75%	99.31%

Conclusions and Next Steps

We have demonstrated the feasibility to capture, sequence and type long DNA fragments.

1. Enzyme B was able to produce large fragments with small amounts of residual small fragments it did not produce sufficient yield to carry to forward to capture.
   - *NEXT STEPS:* Find ways to improve yield with Enzyme B by adding amplification or scaling up the enzymatic fragmentation reaction.
2. Enzyme A produced large fragments and sufficient yields to proceed to capture and sequencing.
3. We were able to successfully sequence long fragments with long read sequencing producing a final 3rd field concordance >99%.
   - *NEXT STEPS: Scale up testing with this method including blood extract samples to better reflect samples obtained in the field.*
4. Our current feasibility workflow has significant loss in concentration throughout and reduction in the mean fragment size where fragments >9kb are not sequenced efficiently.
   - *NEXT STEPS:* Optimize the workflow to allow for fragments >5kb to be sequenced. This would allow us to sequence longer more challenging targets such as KIR.
   - *NEXT STEPS:* Our analysis software had challenges effectively using long fragments for alignment. Moving forward we want to develop our analysis pipeline to further to better handle long fragments and provide more accurate and complete typing.

Trademarks/licensing

Learn more at thermofisher.com/transplant-solutions