Development of a High-Throughput Target Enrichment System

Kristin Butcher, Hutson Chilton, Richard Gantt, Leonardo Arbiza, Siyuan Chen, Yehudit Hasin-Brumshtein, Christina Thompson, Ramsey Zeitoun

1. Abstract

The potential to utilize target enrichment for large-scale genomic analysis has grown with the increased capacity and availability of next-generation sequencing instruments. Described here is the development of a systematic approach to better utilize the capacity of sequencing instruments by decreasing processing time and increasing the throughput of target enrichment while maintaining the high performance. This was accomplished by both developing a **robust multiplexing foundation** and harnessing it to a high-throughput fast hybridization process. Consisting of universal adapter blockers and a unique dual indexing system coupled to efficient probe design and production, the robust multiplexing foundation enables the capture of up to 16 samples in a single hybridization. This capability is further enhanced with a fast hybridization process that vastly reduces workflow time so libraries can be enriched and ready for sequencing in a single workday. As a demonstration of the power of these systems when they are coupled together, results from multiplex captures with hybridization times from 15 minutes to 4 hours are presented.

2. Barcode Independent Universal Blockers

A critical metric that limits performance of multiplex hybrid capture is the percentage of bases on-target. To address this, commercial hybrid capture systems typically employ two types of blocker systems to improve this metric for human samples: **(1)** Cot DNA, a product enriched for repetitive human gDNA and **(2)** adapter blockers, designed to limit adapter cross-hybridization. While sequence specific adapter blockers provide acceptable performance for small barcode sets, they hinder adoption of expanded index sets, restrict changing barcode length between experiments, and prohibit use of unique dual indices (UDI) or unique molecular identifiers (UMI).

A universal blocker system was developed to address these bottlenecks by increasing capture of on-target reads independent of barcode sequence length or design for all TruSeq compatible adapters (**Figure 2.1**). Functionality is also retained regardless of target panel size (**Figure 2.2**) and across singleplex and multiplex target enrichment workflows (see **Sections 3** & **4**). As a result, this universal adapter blocker technology drastically increases on- and near-bait capture, allows for index flexibility between experiments, simplifies design and execution of multiplex experiments, and directly reduces sequencing costs.

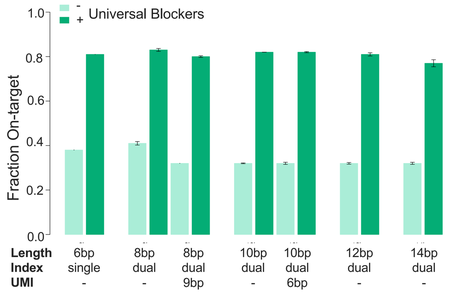

Figure 2.1: Performance is independent of barcode design. On-target performance of universal blockers (Twist Bioscience) across a variety of single and dual index TruSeq-compatible adapters. Individual libraries were generated from a single genomic source (NA12878; Coriell) and TruSeq-compatible adapters with barcode lengths ranging from 6 to 14 bp (UMI length not included). Hybrid capture was performed either in the absence or presence of universal blockers using an exome target enrichment panel (33.1 Mb; Twist Bioscience) using 500 ng of gDNA per individual library following the manufacturer's recommendations for 16-hour hybridization reactions. Cot DNA was present in all samples. Fraction of bases on-target is defined by the equation 1 - PCT_OFF_BAIT. Error bars denote one standard deviation or range of observations; N ≥ 2.

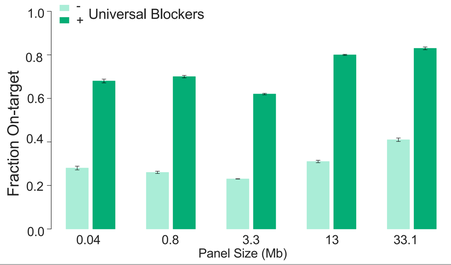

Figure 2.2: Performance is independent of panel size. On-target performance of Universal Blockers (Twist Bioscience) across a variety of panel sizes. Individual libraries were generated from a single genomic source (NA12878; Coriell) and 8 bp dual-indexed TruSeq-compatible adapters. Hybrid capture was performed either in the absence or presence of universal blockers with target enrichment panels of various sizes (0.04Mb to 33.1Mb; Twist Bioscience) and 500 ng of gDNA per individual library following the manufacturer's recommendations for 16-hour hybridization reactions. Cot DNA was present in all samples. Fraction of bases on-target is defined by the equation 1 - PCT_OFF_BAIT. Error bars denote range of observations; N = 2.

3. Improved Unique Dual Index Library Creation

With standard full length Y-adapters, NGS library preparation of challenging samples with 'T-A' overhang ligation often requires excessive PCR cycles, creating unwanted bias from amplification. In addition, the presence of adapter dimer at even low molar concentrations can overestimate the concentration of final libraries and lead to underloading when pooling on a mass basis for multiplexed hybrid capture applications. A 10bp Unique Dual Indexed (UDI) system was developed to alleviate these library preparation bottlenecks by increasing total recovered yield with a reduced number of PCR cycles (**Figure 3.1**) and eliminating adapter dimer formation under standard ligation conditions (gDNA:adapter ratio of 1:200; **Figure 3.2**).

Used in coordination with universal blockers, this adapter system not only improves numerous aspects of NGS library preparation, but lays the foundation for execution of high multiplex hybridization capture.

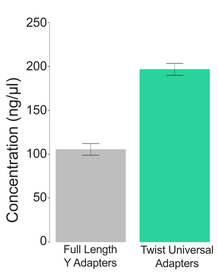

Figure 3.1: Improved Final Yield. Application of the 10bp UDI system for NGS library creation returns higher total yields when compared to standard full length Y-adapters with fewer PCR cycles. For all data, 50ng from a single genomic source (NA12878; Coriell) was processed with a commercially available NGS Library Preparation Kit (Twist Bioscience). Samples were generated with either commercially available standard TruSeq compatible Y-adapters and amplification primers with 10 PCR cycles (N = 12) or the 10bp UDI system with 8 PCR cycles (N = 12). gDNA:adapter ratios were held at a constant molar ratio of 1:200 for all reactions. Equal volumes of the final prepared libraries were evaluated with a Qubit dsDNA Broad Range Assay Kit (Invitrogen) for determination of final yield.

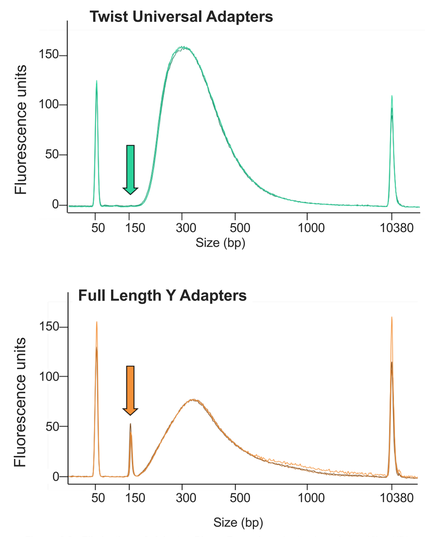

Figure 3.2: Elimination of Adapter Dimer Formation. Application of the 10bp UDI system eliminates formation of adapter dimer during gDNA library preparation. 50ng from a single genomic source (Coriell; NA12878) was processed with a commercially available NGS Library Preparation Kit. Samples were generated with either commercially available standard Full Length TruSeq compatible Y-adapters and amplification primers with 10 PCR cycles or the 10bp UDI system with 8 PCR cycles. gDNA:adapter ratios were held at a constant molar ratio of 1:200 for all reactions. Equal volumes of the final prepared libraries were evaluated with an BioAnalyzer 7500 Assay (Agilent). Overlays of representative electropherograms from both sample types presented (N = 3).

Materials & Methods

Unless otherwise noted, sequencing was performed with a NextSeq 500/550 High Output v2 kit to generate 2x76 paired end reads. Data was down-sampled to 150x of target size and analyzed using Picard Metrics with a mapping quality of 20.

4. Robust Multiplexing by Combining Blockers, Adapters and Probe Design

To support the described blocker and adapter systems, target enrichment probes must also be designed match the desired throughput of the system. Probes were designed to maximize the capture of unique molecules and minimize sequencing duplicates to deliver high multiplex performance. The adapter-blocker-probe system described was developed to maintain high-performance with up to a 16-plex capture. This is demonstrated on three panels of 800 kb, 3.3 Mb and a fixed Exome of 33.1 Mb. Consistent capture coverage at 30x is observed across all samples and multiplexing conditions (**Figure 4.1**). While an increase in duplicate rate is expected as the number of samples in a multiplex increases, the magnitude of this increase is minimal. For an 800kb panel duplication rate increases from 1.8% to 2.7% between 1-plex and 16-plex captures, respectively, and similar observations were made with larger panels. The minimal impact to performance is confirmed with consistent 30x coverage.

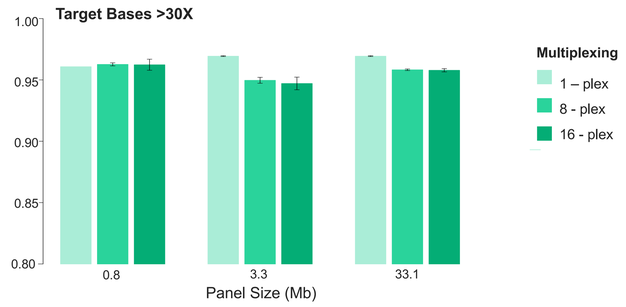

Figure 4.1: Multiplexing Performance. Multiplexing was performed for three panels at three degrees of multiplexing. Hybrid capture was performed using target enrichment panels of various sizes (800kb, 3.3Mb, or 33.1 Mb; Twist Bioscience) using a total of 1500 ng library (NA12878; Coriell) per multiplexed pool following manufacturer's recommendations. **a)** 800 kb cancer panel, **b)** 3.3 Mb cancer panel, and **c)** Exome panel. N = 2.

One strategy to increase multiplexing and capture efficiency is to increase sample mass. In designing this high-throughput system, probes were designed to support multiplexing without increasing sample mass. This was done for two reasons: increasing mass **(1)** requires increases in blocker loading that negate cost savings and **(2)** necessitates more library mass to be available pre-capture. This mass requirement can be challenging, especially when handling low-quality, degraded, or PCR-free sources of gDNA input where additional amplification cycles increase sample bias. This is demonstrated with a degradation of uniformity between 4 to 12 cycles of PCR (**Figure 4.2**). To demonstrate the invariance of this system to input mass, an 8-plex hybridization based capture with either 150 ng or 1,500 ng of total sample was performed. Consistent 30x coverage clearly demonstrates the capacity of this system to multiplex with reduced mass input without degradation to performance (**Figure 4.3**).

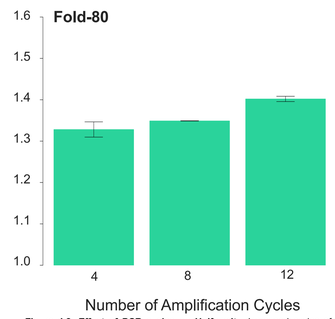

Figure 4.2: Effect of PCR cycles on Uniformity. Increased cycles of PCR during library preparation leads to more bias in the form of an increased Fold 80 base penalty. Hybrid capture was performed using an exome target enrichment panel (33.1 Mb; Twist Bioscience) using 500 ng of library (NA12878; Coriell) as a single-plex capture following manufacturer's recommendations. N = 2.

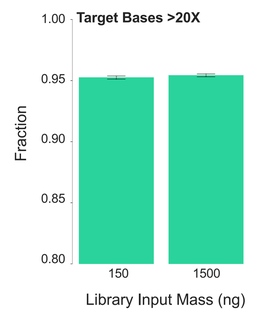

Figure 4.3: Reducing library Input Mass. The coverage and duplicate rate of a high mass capture or low mass capture. Hybrid capture was performed using an exome target enrichment panel (33.1 Mb; Twist Bioscience) using 150 ng (18.75 ng per library) or 1500 ng (187.5 ng per library) of library (NA12878; Coriell) per 8-plex pool following manufacturer's recommendations. Data was down-sampled to 100x of target size; N = 2.

5. Fast Hybridization Workflow for Further Throughput Improvements

While the development of the blocker-adapter-probe system provides a robust foundation for multiplexing, throughput continues to be limited by the length of hybridization. However, the quality of a capture is related to how close it is to equilibration and is time dependent. The target enrichment system described thus far is a 16 hour hybridization and shorter times enable higher throughput with the undesired cost of performance in target coverage and other critical metrics. To radically increase the throughput of samples that can be processed, the conditions for hybridization must be optimized such that the system comes to equilibrium on a shorter timescale.

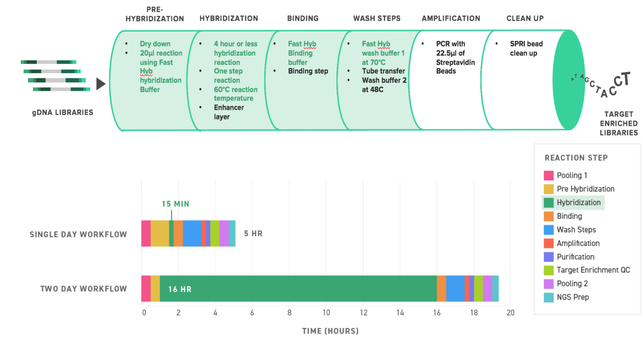

Figure 5.1: Fast Hybridization Workflow. a) General workflow present in the 'Fast Hybridization' system (Twist Bioscience). b) Timing of different steps are shown here and demonstrate the reduction of overall time to hybridization and sequencing. The vast majority of time savings come from the reduction of hybridization time from 16 to a maximum of 4 hours.

In addition, this system generates high-coverage capture results across a range of times with minimal performance degradation (**Figure 5.3**).

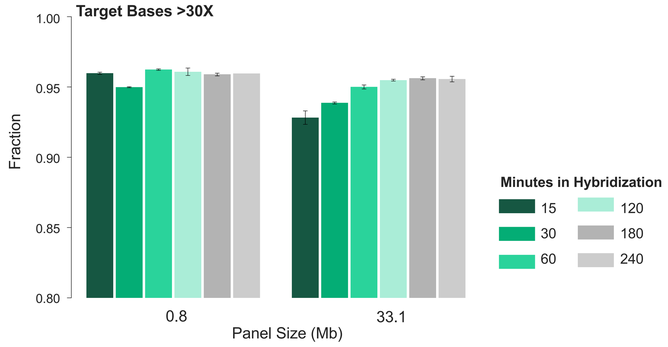

Figure 5.3: Fast Hybridization Time Course. Hybrid capture was performed using an exome target enrichment panel (33.1 Mb; Twist Bioscience) using 187.5 ng of library (NA12878; Coriell) and the 'Fast Hybridization' protocol at variable hybridization times. N = 2. Coverage at 30x is plotted. a) Exome panel (33.1 Mb), b) 800kb panel.

In order to address this, numerous step of the target enrichment buffer system were either optimized or re-engineered into a 'Fast Hybridization' system that arrives at equilibrium on a greatly reduced timescale and provides excellent sequencing performance in 4 hours or less (**Figure 5.1**). Capture performance was evaluated at 4 hours, showing expected uniformity, off target, and overall coverage are equivalent with panels designed for 16 hour capture (**Figure 5.2**).

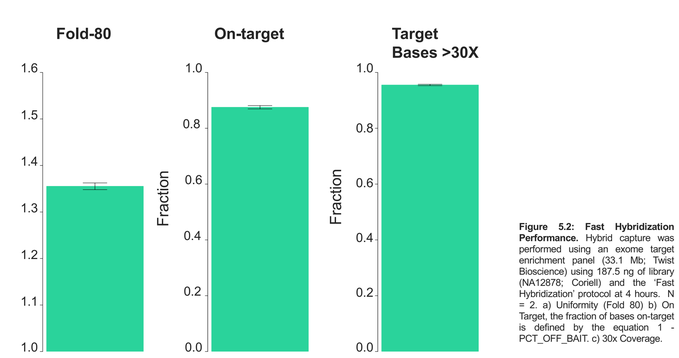

Figure 5.2: Fast Hybridization Performance. Hybrid capture was performed using an exome target enrichment panel (33.1 Mb; Twist Bioscience) using 187.5 ng of library (NA12878; Coriell) and the 'Fast Hybridization' protocol at 4 hours. N = 2. a) Uniformity (Fold 80) b) On Target, the fraction of bases on-target is defined by the equation 1 - PCT_OFF_BAIT. c) 30x Coverage.

Finally, this system is compatible for high-multiplexing applications using the adapter, blocker, and probe systems described above and enables a significant improvement for target enrichment throughput in regards to both samples and time (**Figure 5.4**).

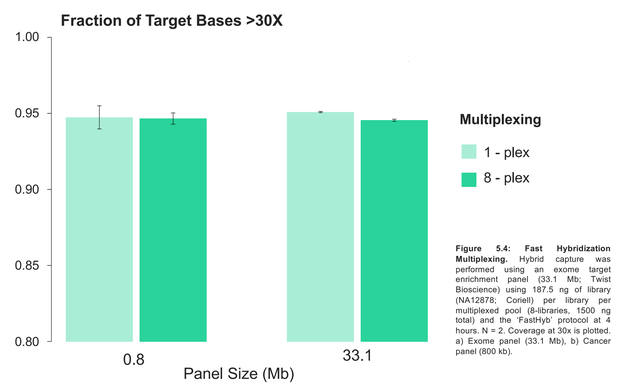

Figure 5.4: Fast Hybridization Multiplexing. Hybrid capture was performed using an exome target enrichment panel (33.1 Mb; Twist Bioscience) using 187.5 ng of library (NA12878; Coriell) per library per multiplexed pool (8-libraries, 1500 ng total) and the 'FastHyb' protocol at 4 hours. N = 2. Coverage at 30x is plotted. a) Exome panel (33.1 Mb), b) Cancer panel (800 kb).

Learn more about the new Twist Exome and Custom Target Enrichment solutions at twistbioscience.com/products/ngs