Improved Library Prep Offers Higher Percentage of On-target Reads and Better Coverage for SureSelect^{XT2} Target Capture

INTRODUCTION

Next-Generation Sequencing (NGS) has provided the scientific community with a powerful tool to sequence large and complex genomes in far less time than traditional Sanger sequencing methods. Even so, we have not yet reached a point in which sequencing and analyzing the 3.2 billion base human genome is either time or cost effective. Target capture strategies have been shown to provide a solution to this problem by reducing genome complexity to what is relevant to the question being asked. Target capture approaches require whole genome library preparation followed by systematic target selection through probe hybridization to genomic regions of interest followed by massively parallel sequencing. The size and complexity of the capture is user dependent but can range from large scale targeting of all protein coding regions (~50 Mb) to a subset of genes known to be involved in specific biological functions (<10 Mb). Pre-capture pooling of multiple samples prior to target capture is a cost effective method to maximize the amount of data obtained for multiple samples in a single sequencing run. Several target capture technologies are commercially available, all of which have their individual strengths and weaknesses. Agilent's SureSelect^{XT2} system offers a wide range of probe sets including All Exon options for many different model organisms, as well as a flexible custom option allowing users to design their own capture panels, all of which can be purchased independent of upstream and downstream reagents required to construct sequence-ready capture libraries.

Target capture performance is dependent on several variables; however, one important factor is the barcode blocking strategy employed. Barcode blocking oligos bind to platform binding adapter sequences to reduce non-specific probe hybridization. A problematic class of unwanted molecules arise during in-solution hybridization target capture from the annealing of unrelated DNA sequences at complimentary terminal adapter sequences during probe hybridization. Identical adapter terminal sequences flank each template molecule in the hybridization pool. Thus, this molecule and its complement are in very high concentration relative to other molecules in solution. The joining of non-complementary template molecules at either terminal adapter sequence can be as little as two molecules, or a chain of many different unrelated sequences. Consequently, the presence of a bait binding site on one of the linked molecules will pull down the entire linked chain, resulting in a larger number of contaminating molecules and increasing the number of off-target reads in target capture sequencing results.

Bioo Scientific now offers all reagents required upstream and downstream of probe hybridization (Table 1) to prepare sequence-ready, multiplexed capture libraries compatible with all SureSelect^{XT2} bait sets. NEXTflex barcode blocker technology allows for maximum binding efficiency through index-specific blocking oligos, greatly reducing the number of off-target reads encountered in small and large target capture assays. Furthermore, utilization of NEXTflex reagents with SureSelect^{XT2} bait set improves target capture efficiency through robust library preparation workflow as well as important hybridization efficiency improvements.

Table 1. Integration of the NEXTflex Pre- and Post- Capture Combo Kit with SureSelect^{XT2} Target Capture Baits

Workflow Stage	Reagents Needed	Bioo Supplied	Agilent Supplied
1. Library Preparation	NEXTflex Rapid DNA-Seq Library Prep Reagents	✓	
	NEXTflex DNA Barcodes	✓	
2. Hybridize Libraries to Probe	SureSelect XT2 Target Capture Baits		✓
	NEXTflex DNA Barcode Blockers	✓	
	NEXTflex Hybridization & Wash Buffers	✓	
3. Amplify Captured Libraries	NEXTflex Post-capture Amplification Reagents	✓	

METHODS

Library Preparation and Target Capture

Genomic DNA was isolated from human blood cells of a consented individual and sheared to an average size of 200 bp. Library prep was performed using 100 ng of the same input DNA for each capture using either Agilent's SureSelect^{XT2} Reagent Kit or Bioo Scientific's NEXTflex™ Pre- and Post-Capture Kit, and barcoded adapters from each, where stated. Target capture was performed according to the manufacturer's instructions using either all Agilent or all NEXTflex reagents, where stated. All baits sets used (SureSelect^{XT2} Human All Exon V5 and SureSelect^{XT2} Inherited Disease) were supplied from Agilent. Quality and quantity of capture libraries was determined using the Agilent 2100 Bioanalyzer. Normalized libraries were clustered on-board, and 150 bp paired-end sequencing was performed on the HiSeq 2500 across 4 lanes of 2 flow cells using rapid run mode.

RESULTS

NEXTflex Barcode Blocking Efficiency

Various barcode blocking strategies can be employed to reduce the presence of unwanted templates following target enrichment. Blocking strategies commonly used during in-solution hybridization assays bind the entirety of a designated strand, usually the common adapter, barcoded adapter, or both. Cost-effective configurations to block specific barcoded adapter sequences are randomized or deoxyinosine bases that barcode specific sequences. Additionally, manufactures may pool all blockers corresponding to the series of index sequences that they offer commercially, which would result in a substantially higher chance of index-specific blocking during hybridization than randomized or deoxyinosine bases. NEXTflex barcode blocker technology uses "index-specific" barcode blocking to most effectively block both adapter termini from annealing during hybridization while Agilent uses the "multiplexed blocker pool" strategy.

In this experiment (Figure 1) the efficiency of two new barcode blocking configurations were tested. The efficiency of an "index-specific" blocking strategy was tested by pre-capture pooling eight individually barcoded libraries and blocking with an oligo specific for the non-barcoded adapter (universal) and the respective index specific blockers. Pre-pooled libraries barcoded with indices 1-8 were blocked with a pool consisting of barcode blockers 1-8 and the universal blocker. This "index-specific" blocking configuration was compared to a "multiplexed blocker pool" strategy in which eight individually barcoded libraries were pre-capture pooled and blocked with a pool consisting of the universal blocker and 96 different blockers, eight of which were the exact complement to the eight barcoded libraries pooled prior to hybridization to Agilent's SureSelect^{XT2} Human All Exome Target Enrichment V5 bait set (Figure 1).
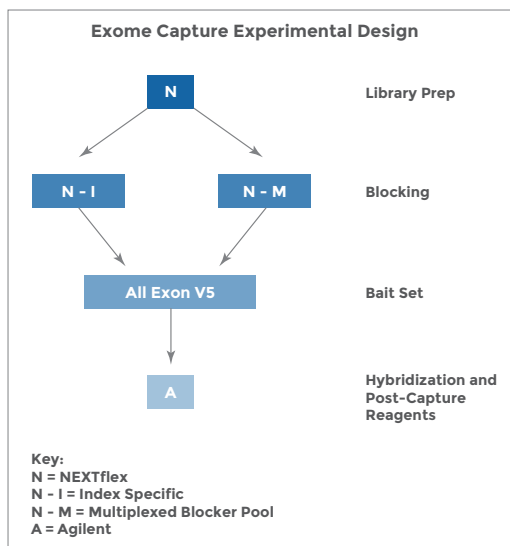

Figure 1. Experimental design flowchart for NEXTflex libraries using different barcode blocking strategies prior to SureSelect^{XT2} all exon V5 target capture.

Of the 50.18 Mb of sequence targeted in the exome capture, both 8-sample captures had roughly the same amount of raw reads with ~170 million reads mapping to the hg19 genome and ~4% of all mapped reads calculated to be duplicates (Table 2).

	Index Specific	Multiplexed Blocker Pool
Targeted Bases	50,180,000	50,180,000
Reads Mapped	170,789,288	167,675,673
% Unique / Duplication	95.7 / 4.3	95.3 / 4.7
% On Target	69.3	65.6
Mean Coverage	14.3	13.2

Table 2. Mapping and coverage statistics for NEXTflex libraries using different barcode blocking strategies prior to SureSelect[XT2] all exon V5 target capture. The NEXTflex Pre- & Post- Capture Combo Kit incorporates the index specific blocking stratagy.

Of the two blocking strategies, the index-specific blocking performed better than the multiplexed blocker pool with reads mapping to target percentages of 69.3% and 65.5%, respectively. Index-specific exome capture produced a mean coverage of 14X while the multiplexed blocker pool approach produced 13X mean coverage of the 50.18 Mb targeted exome sequence. Index-specific blocking showed less variation in blocking efficiency of all 8 capture libraries than the multiplexed blocker pool, resulting in more uniform coverage of the target genome than alternative blocking configurations. The high percentage of total reads mapping to the target sequence demonstrates the effectiveness of index-specific blocking, yielding a higher proportion of on-target reads and increased coverage of genomic regions of interest with less sequencing.

Comparison of Agilent SureSelect[XT2] Reagent Kit to NEXTflex Pre- & Post- Capture Kit

A head to head comparison between all Agilent and all NEXTflex reagents was also performed upstream and downstream of probe hybridization with the SureSelect[XT2] bait set (Figure 2).

Inherited Disease Experimental Design

N	A	Library Prep
N - I	A	Blocking
Inherited Disease		Bait Set
N	A	Hybridization and Post-Capture Reagents

Key:
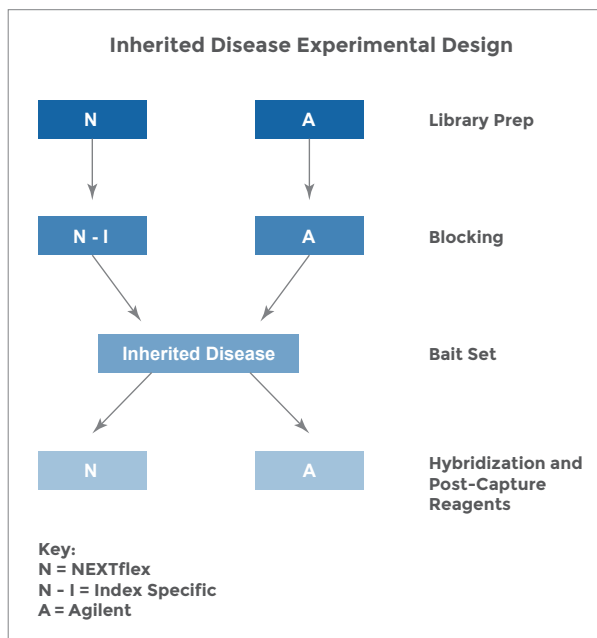**N = NEXTflex**
N - I = Index Specific
A = Agilent

Figure 2. Experimental design flowchart for both all Agilent and all NEXTflex pre and post-capture reagents using the SureSelect[XT2] Inherited disease bait set.

Based on the improved blocking performance demonstrated in the exome capture, index specific blocking was used for the NEXTflex capture and the supplied SureSelectXT2 blocking mix was used for the Agilent captures. Both Agilent and NEXTflex captures produced roughly the same amount of raw reads and both shared ~56 million reads mapping to the hg19 reference genome and 95% of the mapped reads were unique (Table 3).

	NEXTflex	Agilent
Targeted Bases	10,180,000	10,180,000
Reads Mapped	57,608,301	55,621,420
% Unique / Duplication	95.5 / 4.5	95.7 / 4.3
% On Target	83.3	78.5
Mean Coverage	151.2	137.8

Table 3. Mapping and coverage statistics for both all Agilent and all NEXTflex pre and post-capture reagents using the SureSelect^{XT2} Inherited disease bait set.

Importantly, the capture library prepared using all NEXTflex reagents showed better performance, with 83% of reads mapping to the targeted sequence compared to 78% on-target for the Agilent control. Furthermore, the NEXTflex capture pool produced a mean coverage of the target sequence at 151X compared to 137X attained by the Agilent control capture.

CONCLUSION

The seamless compatibility between the NEXTflex Pre- and Post- Capture library preparation kit and all reagents required for target capture using Agilent's SureSelect^{XT2} bait set was demonstrated. The advantages of using NEXTflex reagents for SureSelect^{XT2} target capture experiments begin with the minimal hands on time and enhanced ligation technology offered by our library prep kit, which delivers more unique reads and higher mean coverage per sample compared to competitors. Importantly, the index specific barcode blocking strategy offered by Bioo Scientific outperforms alternate blocking configurations, resulting in better overall target capture performance with an increased number of bases on target. Agilent's SureSelect^{XT2} probe sets can be purchased individually and all other components required for target capture experiments can be purchased from Bioo Scientific in convenient 16 or 96 reaction packages offering greater multiplexing flexibility. Currently, Bioo Scientific offers 96 8-nucleotide barcoded adapters and barcode blockers to match.

BIOO SCIENTIFIC®