An Integrated Solution to Simplify Library Preparation and Multiplexing for NimbleGen Sequence Capture

Authors:
Suchitra Ramani, Marianna Goldrick, Dawn Obermoeller and Masoud Toloue
Bioo Scientific, Austin, Texas

Contributions:
The libraries were constructed by Bioo Scientific and sent along with NEXTflex blocking oligonucleotides to Roche NimbleGen for capture, sequencing and analysis.

Abstract

Targeted sequencing is an important tool in analyzing disease or exome mutations. In this study, we describe how Bioo Scientific's NEXTflex™ DNA Pre-Capture Combo Library construction kits were used in conjunction with NimbleGen Sequence Capture technology to obtain high coverage comparative genomic data from a panel of human HapMap DNA samples. We identified 96-98% of known HapMap SNPs in the NimbleGen SeqCap EZ Exome v3.0 (64 Mb) and SeqCap EZ Design – Comprehensive Cancer Design (3.9 Mb) probes with a high percentage of reads mapping to the targeted regions. This study illustrates both the ability of Bioo Scientific's NEXTflex library construction kit to produce high-quality material for NGS, and the robust performance of the NimbleGen Sequence Capture technology.

Introduction

Recent advances in high throughput sequencing have enabled the ability to rapidly sequence entire human genomes. The ability to compare genomic information between individuals can be maximized by focusing on genomic regions of high interest, since the costs in materials and bioinformatic resources to sequence entire human genomes impose a practical limit on the number of individuals that can be compared in a single study. More specifically, sequencing-ready genomic DNA shotgun libraries prepared from many individuals are created and processed in such a way as to allow them to be mixed, enriched and used as input for efficient Next Generation Sequencing (NGS) while retaining accurate sample identification. Accurate interpretation of the resulting large dataset requires efficient amplification and conversion of primary genomic DNA samples into DNA libraries prior to target enrichment. It is important that minimal bias is introduced during this process so that the types and proportions of inter-individual sequence differences observed faithfully reflect the differences in the unmanipulated genomes. The NEXTflex™ line of kits and protocols developed at Bioo Scientific were designed to meet these demanding needs. The studies described herein demonstrate the capability of Bioo Scientific's novel NEXTflex™ DNA Pre-Capture Combo Library construction kits used in conjunction with the NimbleGen Sequence Capture technology to obtain high coverage comparative genomic data from a panel of human HapMap DNA samples.

NimbleGen Sequence Capture products are designed for one-step parallel enrichment of selected target regions of interest from human genomic DNA samples. Compared to PCR-only based targeted enrichment methods, the NimbleGen approach allows single-tube sequence capture, avoiding the need to set up thousands of PCR reactions. Using NimbleGen technology, large panels of genomic regions of interest can be efficiently enriched in a single experiment. NimbleGen Sequence Capture products include both fixed-content panels and fully customized capture options, providing researchers the flexibility to choose the best option for their own experiments, rather than being locked in to predetermined panels.

Bioo Scientific's NEXTflex™ Pre-Capture Combo Kit is ideally suited for use with the NimbleGen SeqCap EZ Library, and consists of reagents for NGS library preparation, multiplexing, and barcoded adapter blocking within the targeted enrichment workflow. The

NEXTflex Pre-Capture Combo Kit can be used to produce single- or paired-end multiplexed genomic DNA libraries for target capture and Illumina compatible sequencing. The kit features "Enhanced Adapter Ligation Technology", resulting in library preps with a larger number of unique sequencing reads. Improvements to the ligation enzymatic mix give researchers the ability to perform ligations with longer adapters, leading to better flow cell binding efficiencies. In addition, the NEXTflex Pre-Capture Combo Kit simplifies the NGS workflow by using pre-mixed reagents and magnetic bead based cleanup, reducing pipetting steps and streamlining the library preparation workflow.

Multiplexing and Blocking Options

The NEXTflex™ DNA Barcodes are dsDNA library adapters containing index sequences that facilitate an improved multiplexing workflow and flexible setup. The blockers are compatible with the NEXTflex DNA Barcodes, and allow pre-capture or post-capture pooling with Nimblegen SeqCap EZ Library target enrichment systems. The barcodes are supplied in sets of 6, 12, 24, and 96 unique adapters with matching blocking oligos. The blockers are designed to prevent non-specific capture of non-targeted regions during the downstream target enirchment step. This flexibility increases the ability to pool samples most efficiently according to the number of samples available and the project needs, and significantly decreases hands on time while providing robust data quality.

To illustrate the benefits of interfacing the NEXTflex Pre-Capture Combo Kit with the NimbleGen SeqCap EZ Library, a study was carried out using genomic DNA from several individuals recruited for the International HapMap project. The HapMap project has the lofty goal of developing a haplotype map of the human genome to identify patterns of genetic variation that distinguish populations of humans selected from around the globe. These data are expected to allow researchers to find genomic patterns associated with the likelihood of developing pathological conditions such as cancer, heart disease, and diabetes, and to predict the response to treatment with therapeutic agents. Because most HapMap samples have already been deeply analyzed using multiple methods, existing data on sequence variants in individual samples can be used as a standard to evaluate the efficacy and sensitivity of targeted enrichment techniques. The results presented below were obtained using the NimbleGen SeqCap EZ Exome v3.0 (64 Mb) and SeqCap EZ Design – Comprehensive Cancer Design (3.9 Mb) probes to enrich target regions from NEXTflex libraries prepared from four different HapMap DNA samples. The NEXTflex library kit, NEXTflex Barcodes and blocking oligonucleotides were provided by Bioo Scientific. The target capture, sequencing and analysis were performed by NimbleGen R&D.

EZ Exome v3.0 targets enriched from four multiplexed libraries demonstrate the importance and specificity of NEXTflex blocking oligos

Four NEXTflex libraries were prepared from 1 µg of NA19238, NA19239, NA19240, or NA12762 genomic DNA with NEXTflex index adapters 1, 2, 3 and 4, respectively, using standard NEXTflex protocols. Then, 250 ng of each individually amplified library was pooled before capture (for a total of 1 µg) along with 0.5 nmole of blocking oligos specific for each of the barcoded adapters in samples 1 and 2, plus 1 nmole of blocking oligo specific for the non-barcoded (i.e. universal) adapter. Adapter blocking oligos for samples 3 and 4 were intentionally omitted to demonstrate the specificity of the performance boost provided by the blocking oligos included in the NEXTflex Pre-Capture Combo Kit (Table 1).

Table 1. Pre-capture multiplexing with selective blocking of barcoded adapters

SeqCap EZ Exome V3 (64Mb)	
NEXTflex Libraries (pooled before capture)	Pooled NEXTflex Blocking Oligos
250 ng amplified Library 1	0.500 nmol Blocker 1
250 ng amplified Library 2	0.500 nmol Blocker 2
250 ng amplified Library 3	intentionally omitted
250 ng amplified Library 4	intentionally omitted

Capture of the pooled amplified libraries was performed in triplicate, with two of the three replicates sequenced using an Illumina HiSeq2000 2x76 bp paired end read run. Approximately equal numbers of raw reads were obtained for replicate samples between experiments, but the number of raw reads was different within an experiment depending on whether specific barcoded adapter blocking oligos were included in the capture or not (~75 million raw reads were generated for samples 1 and 2 matched with blocking oligos and ~110 million reads for samples 3 and 4 not matched with blocking oligos). Unfiltered reads were mapped to the hg19 reference genome using SOAPv2. Reads that could not be unambiguously assigned to a single location in the genome were not counted as mapped. SNP calling was performed using SOAPv2. Sequencing read coverage and SNP calling results for two of the four samples within the same multiplex capture are shown in Table 2.

Table 2. Mapping and coverage statistics for two of the four samples from the same multiplex capture experiment demonstrate the performance advantage with the use of NEXTflex adapter blocking oligos.

	Sample 2 (with blocking oligos)	Sample 3 (blocking oligos omitted)
Targeted Bases	63,564,965	63,564,965
Raw Reads	76,154,408	124,457,186
Gb of Sequence	11.6	18.9
% Raw Reads Mapped	80.0	82.7
% Unique Reads Mapped	75.5	77.8
% Mapped Reads on Target	75.6	47.2
Average Coverage	43.2	45.6
Sensitivity in Detecting Known SNPs	96.8	96.9
Specificity in Detecting Known SNPs	99.2	99.1

Regardless of whether or not the sample library was matched with a specific blocking oligo included in the capture experiment, approximately 80-83% of raw reads were mapped to the hg19 genome. Of these, approximately 6% were presumed to be duplicate fragments resulting from the PCR amplification steps. Of the unique reads which mapped to the genome, there was a significant difference between samples in the fraction of reads that aligned to the capture target (i.e. capture specificity) depending on whether the sample was paired with NEXTflex adapter blocking oligos or not. For sample 2 (with blocking oligo) 75.5% of mapped unique reads aligned to the target, while for sample 3 (blocking oligo omitted) only 47.2% of mapped unique reads aligned to the target, demonstrating the specificity of the NEXTflex blocking oligo design and the importance of including them in the capture. The mean coverage depths over the targeted bases for samples 2 and 3 were quite similar (43-46%), but reaching this coverage depth required 39% less sequencing for sample 2 due to the higher specificity of capture for this sample. These data suggest that the increase in the number of raw reads obtained for the sample libraries whose barcoded adapters were not specifically blocked consists predominately of off-target reads that were generated at the expense of on-target reads obtained from the blocked samples. SNP calls were compared to ~30,000 known HapMap SNPs in the targeted regions. We were able to detect 97% of these SNPs for both sample 2 and 3, due to the comparable mean coverage for these samples, with a specificity of >99% in each case. The capture data for sample 1 and 4 (the other two samples in the multiplex capture) were very similar to the data for samples 2 and 3 (not shown).

When we examined the results from a pre-capture multiplexing experiment where all four libraries were captured in the presence of specific barcoded adapter blocking oligos (Table 3), we saw much more equivalent results among the individual samples, with 77.0-77.5% of mapped unique reads aligned to the target for all four samples. Again, we detected approximately 97% of known HapMap SNPs in the targeted region, with a specificity of 99%.

Table 3. Pre-capture pooling with NEXTflex Blocking Oligos

SeqCap EZ Exome V3 (64Mb)	
NEXTflex Libraries (pooled before capture)	Pooled NEXTflex Blocking Oligos
250 ng amplified Library 1	0.250 nmol Blocker 1
250 ng amplified Library 2	0.250 nmol Blocker 2
250 ng amplified Library 3	0.250 nmol Blocker 3
250 ng amplified Library 4	0.250 nmol Blocker 4

Multiplexed pre-capture pooling is an efficient and scalable solution for variant discovery in candidate gene panels

We also performed targeted resequencing of candidate gene panels using the NEXTflex Pre-Capture Combo kit and the SeqCap EZ Design – Comprehensive Cancer Design. For these capture experiments, twelve libraries were prepared with NEXTflex library reagents, each with distinct barcoded adapters and blocking oligos (all components of the NEXTflex Pre-Capture Combo Kit). Amplified libraries were pooled prior to capture with the SeqCap EZ Design – Comprehensive Cancer Design (Table 4).

Table 4. Pre-capture pooling using NEXTflex and SeqCap EZ Design – Comprehensive Cancer Design

SeqCap EZ Design – Comprehensive Cancer Design	
NEXTflex Libraries (pooled before capture)	Pooled NEXTflex Blocking Oligos
83 ng amplified Library 1	0.083 nmol Blocker 1
83 ng amplified Library 2	0.083 nmol Blocker 2
83 ng amplified Library 3	0.083 nmol Blocker 3
83 ng amplified Library 4	0.083 nmol Blocker 4
83 ng amplified Library 5	0.083 nmol Blocker 5
83 ng amplified Library 6	0.083 nmol Blocker 6
83 ng amplified Library 7	0.083 nmol Blocker 7
83 ng amplified Library 8	0.083 nmol Blocker 8
83 ng amplified Library 9	0.083 nmol Blocker 9
83 ng amplified Library 10	0.083 nmol Blocker 10
83 ng amplified Library 11	0.083 nmol Blocker 11
83 ng amplified Library 12	0.083 nmol Blocker 12

The raw sequences from the twelve pre-capture multiplexed samples shown in Table 4 were mapped and analyzed. The results are summarized in Table 5. Sequence reads were distributed uniformly among samples, an indication that a large set of samples can be effectively

multiplexed and pooled prior to capture. The percentage of targeted bases with at least 10X coverage depth was >98% for all twelve samples. The sensitivity for SNP detection was 97-98% for known SNPs (from publicly available HapMap Project genotype data) in the Comprehensive Cancer Design target for each of the twelve samples in the multiplexed captures. The average coverage depth for the twelve pooled sample libraries was 215X, which is more than sufficient to call variants with high confidence using stringent sequence depth and quality filters. Clearly, multiplexing even more samples to achieve the same goal would be possible.

Table 5. Mapping and coverage statistics for twelve sample libraries demonstrates the efficacy of pre-capture multiplexing using NEXTflex library and blocking oligo reagents.

Samples	NEXTflex Index	Targeted Bases	Raw Reads	Gb of Sequence	% Raw Reads Mapped	% Unique Reads Mapped	% Mapped Reads on Target	Unique Gb of Sequence on Target	Average Coverage	% Sensitivity in Detecting Known SNPs	% Specificity in Detecting Known SNPs
1	CGATGT	3,898,910	26,548,162	4.0	90.4	78.8	63.4	2.6	220.4	97.70	98.70
2	TGACCA	3,898,910	25,218,056	3.8	90.3	79.2	63.6	2.4	211.1	98.10	98.90
3	ACAGTG	3,898,910	27,026,818	4.1	90.7	79.0	64.2	2.6	227.9	97.90	99.00
4	GCCAAT	3,898,910	24,589,892	3.7	90.2	78.8	63.5	2.4	204.7	97.70	98.60
5	CAGATC	3,898,910	26,077,566	4.0	90.5	79.3	63.9	2.5	219.4	97.80	98.70
6	CTTGTA	3,898,910	24,412,142	3.7	90.4	80.1	63.4	2.4	206.2	98.00	98.90
7	ATCACG	3,898,910	29,317,270	4.5	90.7	78.2	64.2	2.9	244.6	97.80	98.90
8	TTAGGC	3,898,910	20,994,732	3.2	90.5	80.0	64.0	2.0	178.6	97.70	98.40
9	ACTTGA	3,898,910	25,344,596	3.9	90.5	79.8	63.4	2.4	213.3	97.70	98.60
10	CATCAG	3,898,910	25,371,206	3.9	90.4	79.5	63.9	2.5	214.5	98.00	98.90
11	TAGCTT	3,898,910	27,844,310	4.2	90.6	79.3	63.9	2.7	234.6	97.80	99.00
12	GGCTAC	3,898,910	23,987,332	3.6	90.4	79.3	64.0	2.3	202.2	97.70	98.70

Conclusions

Multiplexing of NEXTflex DNA libraries for sequence capture with Nimblegen SeqCap EZ Libraries is an efficient solution for improving sample throughput and reducing project costs. When attempting to determine the number of samples that can be multiplexed together, the user should consider the size of the capture target, the amount of sequencing that will be done, the percentage of sequence reads that map to the target for that design, the uniformity of the coverage obtained for that design, and the minimum coverage depth needed for calling SNPs. We were able to achieve enriched libraries of uniform high quality and target coverage, with 96-97% of all known single-nucleotide polymorphisms (SNPs) detected in the exome capture target for each of the four samples in a pre-capture multiplex experiment, sequenced on a single lane of an Illumina HiSeq2000 instrument. We identified 97-98% of known HapMap SNPs in the Comprehensive Cancer Design target for each of the twelve sample libraries prepared using multiplexed captures, with a high percentage of reads mapping to the targeted regions. These results illustrate both the ability of Bioo Scientific's NEXTflex library construction kit to produce high-quality material for NGS, and the robust performance of the NimbleGen Sequence Capture technology. These studies also demonstrate the compatibility of the NEXTflex Pre-Capture Combo Kit used in conjunction with the NimbleGen SeqCap EZ Library to enable the efficient acquisition of high quality genomic data while simplifying and reducing the costs of the upfront library preparation needed for target capture analysis.

Bioo Scientific Corporation
3913 Todd Lane, Suite 312
Austin, Texas 78744

BIOO SCIENTIFIC®

www.biooscientific.com
info@biooscientific.com
1.512.707.8993