Multiplex Sequence Capture for Targeted Resequencing of Candidate Gene Panels in Cancer

A Protocol to Increase Sample Throughput and Reduce Redundant Sequencing

Dan Burgess, PhD
Senior Scientist
Roche NimbleGen, Inc.

Additional Contributors:
Jennifer Wendt, Dawn Green, Michael Brockman, PhD,
Xinmin Zhang, PhD, and Thomas Albert, PhD
Roche NimbleGen, Inc.

Kulvinder Kaur, DPhil, Richard Copley, DPhil, Stefano Lise,
PhD, Enric Domingo, PhD, I. Tomlinson, PhD, BM BCh,
and Jenny Taylor, DPhil
Wellcome Trust Centre for Human Genetics
University of Oxford, Oxford, UK

Introduction

Complex, progressive, multigenic, somatic mutation of the genome is now widely accepted as the primary driving force in the evolution of tumor initiation, growth, metastasis and pharmacoresistance for many types of cancer. This fact was first recognized by cytogeneticists decades ago, upon observing that tumor cell karyotypes frequently exhibited gross chromosomal rearrangements, including aneuploidy, polyploidy, translocations, deletions, duplications and inversions. The more recent development and application of techniques for interrogating tumor genomes with higher resolution confirms that a similarly diverse assortment of genomic variants is equally pervasive with decreasing scale, from submicroscopic rearrangements involving dozens of genes down to isolated mutations of single DNA base pairs. Different types of tumors, at different stages of progression, are associated with mutations in distinct sets of genes, and there is a pressing need for a sequencing-based method of analyzing panels of candidate genes that can differ for each tumor type. The method must have adequate sensitivity to detect somatic variants at sub-Mendelian molecular frequencies due to the variable tumor load and cellularity among different tissue samples. Finally, the ideal method should have high throughput and low cost. Sequence Capture, a method of using long oligonucleotide DNA probes to isolate selected genomic DNA targets from solution by hybridization in preparation for advanced sequencing, meets these criteria. Here, we demonstrate the application of multiplex Sequence Capture techniques using custom gene panels for the analysis of somatic genome variation in colorectal tumors.

For life science research only. Not for use in diagnostic procedures.

Results

Multiplex Sequence Capture of custom gene panels from matched tumor-normal DNA samples

Two SeqCap EZ library designs were developed to enrich for the coding regions of cancer pathway genes in preparation for high-throughput sequencing using the Illumina HiSeq Instrument. The first design targeted a set of 141 genes comprising 899Kb of genomic space. The second design targeted a subset of 17 genes of the 141 genes, comprising 85Kb of genomic space. Standard Roche NimbleGen SeqCap EZ Library design algorithms were used to select capture probes for the targets. Greater than 97% of the targeted bases for each design were within the predicted range of selected capture probes (Appendix 1, Section A1). Eight DNA samples representing four tumor-normal matched pairs were isolated from colorectal tumor samples and pathologically normal surrounding tissue. DNA libraries were constructed using Illumina TruSeq DNA Sample Preparation Kits with eight different indexed (barcoded) adapters, AD001 – AD008 (Appendix 1, Section A2). The peak of the library fragment size distribution for each library was approximately 300 bp in size, indicating genomic DNA inserts of around 200 bp (data not shown).

Sequence Capture from multiplexed Illumina TruSeq DNA Libraries followed by HiSeq sequencing yields a balanced distribution of indexed reads

The eight indexed DNA sample libraries were amplified separately via pre-capture ligation-mediated PCR (LM-PCR), quantified and pooled in equal amounts. DNA from this single pool was split into four different Sequence Capture hybridizations (two replicate captures for the 85Kb design, and two for the 899Kb design). One µg of pooled DNA (125ng of each component DNA sample library) was used in each of the hybridizations. The hybridization enhancing (HE) oligos used to suppress secondary capture were matched in length and sequence to each of the indexed library adapters. Following washing and post-capture LM-PCR amplification, each captured multiplex DNA sample was sequenced with one lane of Illumina HiSeq sequencing (2x100bp paired-end). The number of read pairs containing different adapter index is shown in Figure 1. The similar rank order of indices among the four captures suggests that neither the capture process nor the capture target was the primary source of variation in sample-to-sample read representation, but rather that the primary influence on this relationship was the accuracy of initial library pooling prior to capture or factors intrinsic to the adaptors themselves.
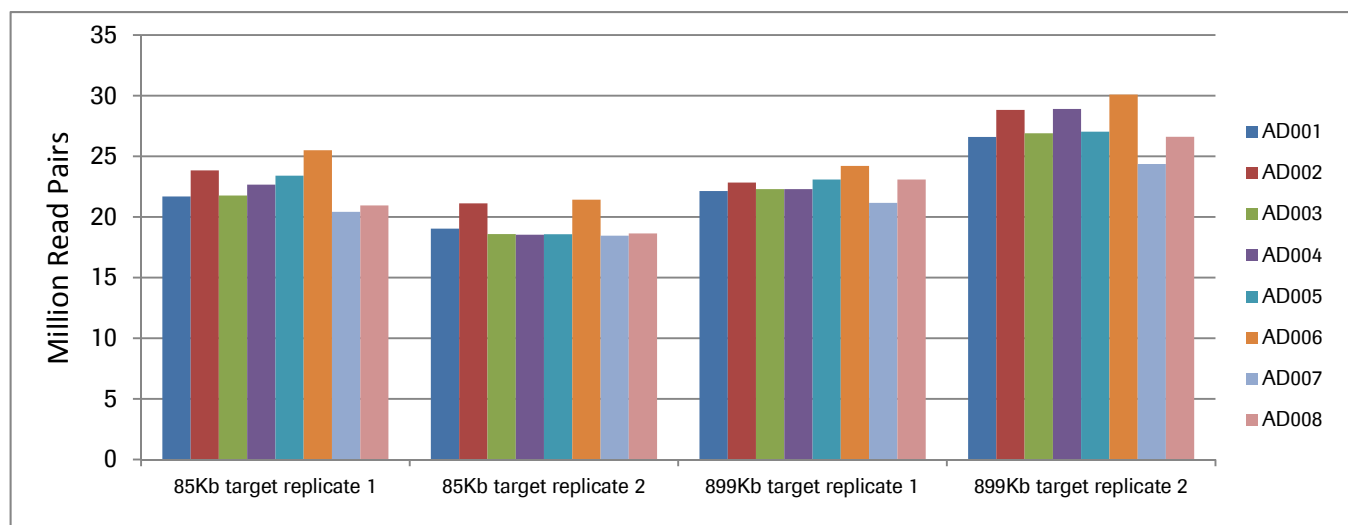

Figure 1. Distribution of Adapter Indices in Reads from Multiplexed Sequence Capture. Adapter indices were counted among raw reads prior to mapping or duplicate removal. The difference between the most- and least-represented indices was less than 20% ((High - Low) / High) in each capture. The Y-axis indicates the number of read pairs, in millions, representing each Illumina TruSeq DNA library adapter index. The identity of each specific index is shown at the right side of the chart.

Multiplexed Sequence Capture is an efficient and scalable solution for variant discovery in candidate gene panels

The raw sequences from the four multiplexed captures shown in Figure 1 were mapped and analyzed. The results are summarized in Figure 2. The smaller (85Kb) target was captured with a lower percentage of mapped unique reads on target compared to the larger (899Kb) target (top-center), but due to its smaller size exhibited a greater average coverage depth over the targeted bases (top-right). Median coverage depth, averaged over all captures for both designs was only 2.1% ± 0.1% (value ± standard deviation; data not shown) different from the average coverage depth, suggesting that the coverage over the capture target was very uniform. Sequence coverage uniformity is a strong determinant of how many samples can be effectively multiplexed assuming sequencing yield is the rate-limiting factor. The percentage of targeted bases with at least 10X coverage depth was > 99% for the 85Kb target and > 97% for the 899Kb target (bottom-left). The percent of targeted bases with ≥ 1X coverage depth for the 85Kb Rep1, 85Kb Rep2, 899Kb Rep1 and 899Kb Rep2 captures was only slightly greater, 99.8 ± 0.1, 99.7 ± 0.1, 99.0 ± 0.1, 98.9 ± 0.1 (value ± standard deviation), respectively.

 The number of filtered SNPs called from on target reads between the two designs was roughly proportional to the target size, and the number of filtered SNPs called from

tumor DNA samples was not significantly different from the number called from the matched normal samples (experimental data not shown). The fraction of called SNPs that matched known SNPs in the dbSNP database was near 90% for tumor and normal samples and both capture targets. A more accurate estimate of actual true positive SNP discovery rates would require capture from a previously characterized DNA sample and/or thorough follow-up validation of called SNPs.

Bioinformatic analysis parameters are important in multiplex capture experimental design

 The average coverage depth for the eight sample libraries that were pooled, captured together in a single reaction with the 85Kb design probes and sequenced using one lane of a HiSeq instrument, was 4,759.9 ± 292.2 (value ± standard deviation), while the lowest average coverage depth for an individual sample library was 4,358. These values, which were determined after PCR duplicate removal, are more than sufficient to call homozygous and heterozygous variants with high confidence using stringent sequence depth and quality filters. Clearly, for this particular design, multiplexing more than eight sample libraries to meet that same goal would have been possible.
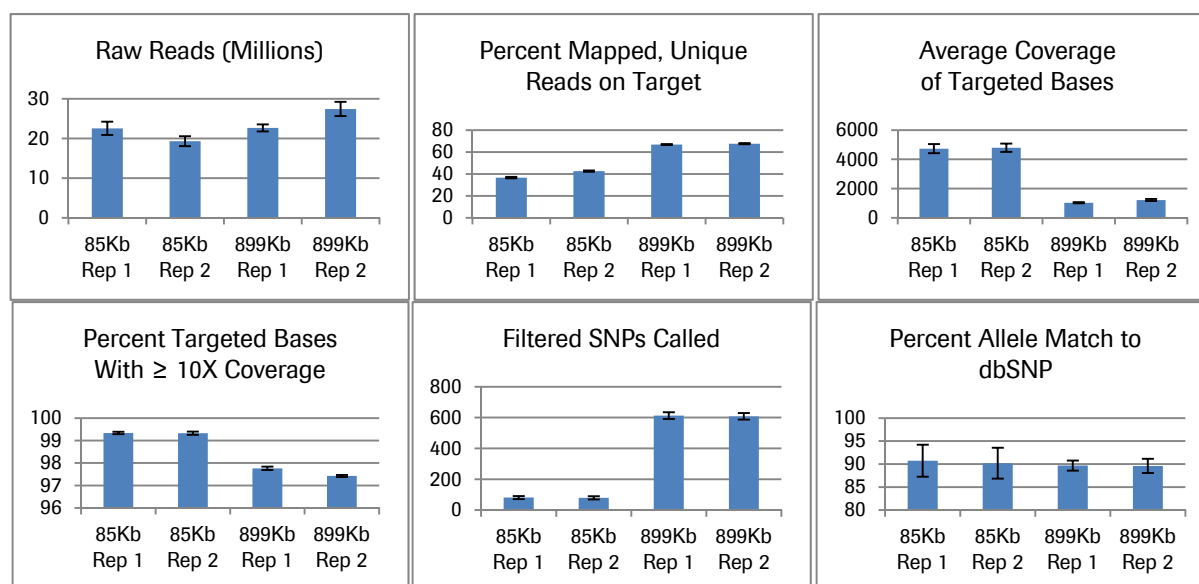

Figure 2. Mapping, coverage and SNV discovery rates in two candidate gene panels. Raw reads were mapped against the hg19 reference genome. Following duplicate removal, reads were used to determine coverage and call single nucleotide polymorphisms (SNPs). Duplicates were defined by identical start coordinates for both reads of a read-pair; for duplicate removal, only the highest quality pair was retained. Read "on-target" was defined as ≥ 1 base pair overlap with the target. Mapping and SNP calling were performed using the SOAP2 software package (BGI). SNPs were called with depth ≥ 4 plus quality ≥ Q20 filters. Heterozygotes were defined using MAF thresholds of 15%, 85%. Build131 (hg19) of dbSNP was used. Error bars are ± 1 standard deviation. The bottom-left and bottom-right graphs are not zero-scaled.

For example, if the desired average coverage depth were 100X for each individual sample library in the pool, the data suggests >300 samples could be multiplexed in one capture reaction and the same amount of sequencing performed. Cumulative variability in the process, such as in DNA quantification steps and library preparation efficiency, could make this high of a level of consolidation impractical, however, it is obvious that significant improvements in sample throughput and cost reduction can be facilitated by multiplexing Sequence Capture experiments. If the goal of an experiment is to identify somatic mutations in tumors, where the variants may be present in only a small fraction of cells in the tissue sample, then additional considerations become important. Specifically, the expected frequency of somatic variants becomes the primary factor in determining the desired coverage depth, and consequently, in the number of samples that can be effectively multiplexed in one capture reaction. Bioinformatic analysis and follow-up validation of candidate germline and somatic mutations identified in this study were ongoing at the time of this report.

Conclusions

Multiplexing of DNA sample libraries for Sequence Capture, as demonstrated here with indexed Illumina TruSeq libraries generated from matched tumor-normal samples, is a viable and efficient solution for improving sample throughput and reducing project costs. Determination of the number of research samples that can be multiplexed depends primarily on: (1) size of the capture target, (2) accuracy in processing steps such as quantification and mixing of samples, (3) specificity of capture (i.e. the percentage of sequence reads that map to the target), (4) uniformity of sequence coverage over the target, (5) sequencing allowances per capture, and (6) bioinformatics requirements or filters such as minimum coverage depth needed for calling SNPs. Multiplex capture of targets from matched tumor-normal samples provides the additional important benefit that any sequence variation observed between them is less likely due to variability in the capture process, and more likely to be intrinsic to the sample genomes. These same benefits are also relevant to the analyses of trios for Mendelian disease, and for larger-scale population studies where multiplexed Sequence Capture enables studies that would otherwise be impractical or unaffordable.

Appendix 1: Materials and Methods

A1: Design specifications

SeqCap EZ probe design utilized standard Roche NimbleGen algorithms employing up to 2.1 million distinct biotinylated long DNA oligonucleotide probes per design. Small target regions (e.g. small exons) were padded to a minimum of 100bp in size to facilitate effective probe selection. Target bases covered with 0bp offset are determined by counting target bases directly represented in probe sequences. Target bases covered with 100bp offset are determined by counting all target bases within 100bp of a probe, which is a more accurate indicator of actual coverage obtained from the sequencing of captured library fragments.

Designs:	Design 1 (899Kb)		Design 2 (85Kb)	
Offset in bases	0	100	0	100
Consolidated/Padded Regions	2,315	2,315	264	264
Target Bases	898,633	898,633	84,771	84,771
Target bases covered	858,303	883,975	80,304	82,838
Target bases covered (%)	95.5	98.4	94.7	97.7
Target bases not covered	40,330	14,658	4,467	1,933
Target bases not covered (%)	4.5	1.6	5.3	2.3

A2: Research sample profiles

Matched Tumor-Normal DNA Samples Representing Four Colorectal Tumors with the Corresponding Illumina TruSeq DNA Library Indices.

Gender	Age at Surgery	Site	Differentiation	Follow-up	Sample Method	Sample ID N=Normal T=Tumor	Illumina TruSeq DNA Library Index
M	75 yrs	Sigmoid colon	Well/Moderate	Liver met.	Fresh Frozen	286N	AD001
						280T	AD002
M	66 yrs	Rectum	Well/Moderate	Liver met.	Fresh Frozen	398N	AD003
						408T	AD004
M	52 yrs	Sigmoid colon	Moderate	Liver met.	Fresh Frozen	490N	AD005
						485T	AD006
F	66 yrs	Ascend. colon	Moderate	Liver met.	Fresh Frozen	552N	AD007
						557T	AD008

Appendix 2: Instructions for using multiplexed Illumina TruSeq DNA Libraries with SeqCap EZ Libraries

Description

Note that the performance of targeted enrichment methods is variable due to population variability among genomes, DNA sample input quality, and other properties intrinsic to the targeted region(s). When using a previously untested capture design and protocol, it is recommended to perform a small-scale experiment to determine the capture characteristics before beginning a larger study.

Use this protocol in conjunction with the *NimbleGen SeqCap EZ Library SR User's Guide v3.0* (*http://www.nimblegen.com/seqcapez*), which provides instructions for performing Sequence Capture from individual libraries prepared with the Illumina TruSeq DNA Sample Preparation Kit. The instructions provided below allow the user to adapt the *NimbleGen SeqCap EZ Library SR User's Guide v3.0* to perform Sequence Capture from multiplexed libraries prepared with the Illumina TruSeq DNA Sample Preparation Kits (revision C, June 2011). This includes important changes to the Hybridization Enhancing (HE) Oligo design and the Hybridization step.

This table guides you through the SeqCap EZ multiplex protocol. It identifies when to follow instructions in the *User's Guide* and when to follow instructions in this document.

Protocol Step	Refer to the following:
1. **Before You Begin**	Chapter 1 in the User's Guide* <u>except</u> for the section entitled Custom Oligonucleotides Purchased from IDT or Another Vendor, which is <u>replaced</u> by the section entitled Custom Oligonucleotide Consumables for Multiplex Sequence Capture in this Application Note.
2. **Storage of the SeqCap EZ Library**	Chapter 2 in the *User's Guide*.*
3. **Sample Library Preparation**	Chapter 3 in the *User's Guide** <u>and</u> the section entitled *Sample Library Preparation for Multiplex Sequence Capture* in this Application Note.
4. **Sample Library Amplification Using LM–PCR**	Chapter 4 in the User's Guide* <u>and</u> the section entitled *Sample Library Amplification Using LM-PCR for Multiplex Sequence Capture* in this Application Note.
5. **Hybridization of Sample and EZ Libraries**	"*Hybridization of Multiplexed Samples and EZ Libraries*" section in this Application Note, which <u>replaces</u> Chapter 5 in the *User's Guide*.*
6. **Washing and Recovery of Captured DNA**	Chapter 6 in the *User's Guide*.*
7. **Captured DNA Amplification Using LM–PCR**	Chapter 7 in the *User's Guide*.*
8. **Measurement of Enrichment Using qPCR**	Chapter 8 in the *User's Guide*.*

* User's Guide = NimbleGen SeqCap EZ Library SR User's Guide v3.0 or later

Custom oligonucleotide consumables for multiplex Sequence Capture

Important Note: This section replaces the section entitled Custom Oligonucleotides Purchased from IDT or Another Vendor in Chapter 1 of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0*. NimbleGen SeqCap EZ enrichment experiments utilizing Illumina TruSeq libraries require the use of at least two Hybridization Enhancing (HE) Oligos. TS-HE Universal Oligo 1 is a universal blocking oligo used for all experiments regardless of which TruSeq index sequence is present in the library adapter. One or more TS-INV-HE Index Oligos (depending on the multiplexing strategy) should be selected from among the TS-INV-HE Index Oligos listed in the table below, and should correspond to the specific TruSeq indexed adapters used in the experiment. For example:

- To perform SeqCap EZ enrichment using a single Illumina TruSeq library constructed with adapter AD001, you will need to use TS-HE Oligo 1 and TS-INV-HE Index 1 Oligo.

- To perform SeqCap EZ enrichment from a mixture of four Illumina TruSeq libraries constructed with adapters AD001, AD002, AD003 and AD004 (i.e. multiplex sequence capture), you will need to use TS-HE Universal Oligo 1, TS-INV-HE Index 1 Oligo, TS-INV-HE Index 2 Oligo, TS-INV-HE Index 3 Oligo and TS-INV-HE Index 4 Oligo.

Read the footnotes at the end of the table (page 8) which explain the purposes of the Oligos listed. The following custom oligonucleotides can be purchased from IDT or another vendor.

Component	Concentration	Sequence	Note(s)
TS-PCR Oligo 1[A]	100 µM	5'- AAT GAT ACG GCG ACC ACC GAG A - 3'	These oligos are used for both Pre-Capture and Post-Capture LM-PCR.
TS-PCR Oligo 2[A]	100 µM	5' - CAA GCA GAA GAC GGC ATA CGA G - 3'	
TS-HE Universal Oligo 1[B, C]	1000 µM	5' - AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T - 3'	These oligos must be resuspended in molecular grade water.
TS-INV-HE Index 1 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	These oligos are used in the hybridization step to increase capture specificity.
TS-INV-HE Index 2 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT ACA TCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 3 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT GCC TAA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	Order only those TS-INV-HE Index oligos that correspond to the indexed Illumina TruSeq DNA sample libraries to be used in capture experiments.
TS-INV-HE Index 4 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT TGG TCA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 5 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT CAC TGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 6 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT ATT GGC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 7 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT GAT CTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 8 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT TCA AGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	TS-INV-HE index oligos must be full-length and efficiently terminated (refer to footnote D for this Table).
TS-INV-HE Index 9 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT CTG ATC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 10 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT AAG CTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 11 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT GTA GCC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 12 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT TAC AAG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T /term/ - 3'	
TS-INV-HE Index 13 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT TGT TGA CTG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ - 3'	
TS-INV-HE Index 14 Oligo[B, D]	1000 µM	5' - CAA GCA GAA GAC GGC ATA CGA GAT ACG GAA CTG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ - 3'	

TS-INV-HE Index 15 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT TCT GAC ATG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 16 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT CGG GAC GGG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 18 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT GTG CGG ACG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 19 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT CGT TTC ACG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 20 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT AAG GCC ACG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 21 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT TCC GAA ACG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 22 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT TAC GTA CGG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 23 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT ATC CAC TCG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 25 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT ATA TCA GTG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
TS-INV-HE Index 27 Oligo[B, D]	1000 µM	5' – CAA GCA GAA GAC GGC ATA CGA GAT AAA GGA ATG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT /term/ – 3'	
qPCR NSC-0237, forward, Oligo	2 µM	5' – CGC ATT CCT CAT CCC AGT ATG – 3'	These oligos are used in qPCR analysis of enrichment.
qPCR NSC-0237, reverse, Oligo	2 µM	5' – AAA GGA CTT GGT GCA GAG TTC AG – 3'	
qPCR NSC-0247, forward, Oligo	2 µM	5' – CCC ACC GCC TTC GAC AT – 3'	
qPCR NSC-0247, reverse, Oligo	2 µM	5' – CCT GCT TAC TGT GGG CTC TTG – 3'	
qPCR NSC-0268, forward, Oligo	2 µM	5' – CTC GCT TAA CCA GAC TCA TCT ACT GT – 3'	
qPCR NSC-0268, reverse, Oligo	2 µM	5' – ACT TGG CTC AGC TGT ATG AAG GT – 3'	
qPCR NSC-0272, forward, Oligo	2 µM	5' – CAG CCC CAG CTC AGG TAC AG – 3'	
qPCR NSC-0272, reverse, Oligo	2 µM	5' – ATG ATG CGA GTG CTG ATG ATG – 3'	

Oligonucleotides can be resuspended in PCR grade water or TE buffer unless otherwise noted.

[A] TS-PCR Oligo 1 and TS-PCR Oligo 2 are the same as the PE-POST 1 and PE-POST 2 oligos described in the SeqCap EZ Human Exome SR (version2.2) and SeqCap EZ Choice SR (version 1.0) User Guides.

[B] HE = "Hybridization Enhancing."

[C] TS-HE Universal Oligo 1 is designed to block the universal segment of TruSeq DNA library adapters during the sequence capture hybridization step. TS-HE Universal Oligo 1 is the same as the PE-HE1 oligo described in the SeqCap EZ Human Exome SR (version2.2) and SeqCap EZ Choice SR (version 1.0) User Guides.

[D] TS-INV-HE Index Oligos are designed to block the indexed segment of TruSeq DNA library adapters during the sequence capture hybridization step. TS-INV-HE Index Oligos 2, 4, 5, 6, 7 and 12 correspond to the TruSeq indexed adapters AD002, AD004, AD005, AD006, AD007 and AD012 available in the Illumina TruSeq DNA Sample Preparation Kit, FC-121-1001 (v1, Set A). The other TS-INV-HE Index Oligos in the table correspond to like-numbered TruSeq indexed adapters in other Illumina TruSeq DNA Sample Preparation Kits, including FC-121-2001 (v2, Set A), FC-121-1002 (v1, Set B) and FC-121-2002 (v2, Set B). Nucleotides shown in red boldface text correspond to the variable (barcode) nucleotides of the TruSeq DNA Adapters. Note that the barcode region of TruSeq adapters numbered 1-12 is *six* nucleotides in length, while the barcode of TruSeq adapters numbered 13 and higher is *eight* nucleotides in length. The notation '/term/' indicates the addition of a 3' terminator modification (*e.g.* dideoxy-C, inverted-dT, or C3 spacer) to block extension from residual TS-INV-HE oligos in the subsequent LM-PCR step. *It is very important that HE oligos used in Sequence Capture experiments be* <u>*full-length*</u> *and effectively* <u>*terminated*</u>*. HPLC purification of these oligos is recommended to achieve the best performance.*

Sample library preparation for multiplex Sequence Capture

Important Note: This section should be read before following the instructions in *Chapter 3 - Sample Library Preparation of the NimbleGen SeqCap EZ Library SR User's Guide v3.0.*

NimbleGen SeqCap EZ Library SR User's Guide v3.0 provides instructions for preparing a single DNA Sample Library using the Illumina TruSeq DNA Sample Preparation Kit for use in Sequence Capture. In preparation for Sequence Capture from multiplexed libraries, it is essential that each library that will be included in the experiment be constructed using a different DNA Adapter Index from the Illumina TruSeq DNA Sample Preparation Kits.

Sample library amplification using LM-PCR for multiplex Sequence Capture

Important Note: This section should be read before following the instructions in *Chapter 4 - Sample Library Amplification Using LM-PCR* of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0.*

NimbleGen SeqCap EZ Library SR User's Guide v3.0 describes how to amplify an individual sample library (prepared in Chapter 3) using LM-PCR (Ligation Mediated-PCR) in preparation for hybridization to the SeqCap EZ Library probes. When performing Sequence Capture from multiplexed libraries, it is essential that each library that will be included in the multiplexed library pool is separately amplified, cleaned-up, and analyzed for quality. Pooling of different sample libraries for multiplexed Sequence Capture does not occur until immediately prior to hybridization (refer to the section entitled *Hybridization of Multiplexed Samples and SeqCap EZ Libraries* in this Application Note).

Hybridization of multiplexed samples and SeqCap EZ Libraries

Important Note: This section replaces *Chapter 5 - Hybridization of Sample and SeqCap EZ Libraries* of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0.*

This section describes the Roche NimbleGen protocol for hybridization of the amplified sample library and the SeqCap EZ Library. Be aware that the hybridization protocol requires a thermocycler capable of maintaining 47°C for 64 - 72 hours.

Step 1. Prepare the heat block for hybridization

1. Turn on a heat block to 95°C and let it equilibrate to the set temperature of 95°C.

Step 2. Prepare the Multiplex DNA Sample Library Pool

1. Thaw on ice each of the differently-indexed amplified DNA sample libraries that will be included in the multiplex capture experiment (*these were generated in Chapters 3 and 4 of the NimbleGen SeqCap EZ Library SR User's Guide v3.0*)
2. Mix together equal amounts by mass of each of these amplified DNA sample libraries to obtain a single pool with a combined mass of at least 1.1 µg. This mixture will subsequently be referred to as the "Multiplex DNA Sample Library Pool." One µg of the multiplex DNA sample library pool will be used in the sequence capture hybridization step, and 60 ng will be used for *Measurement of Enrichment Using qPCR* (*Chapter 8* of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0*). **Note:** To obtain equal numbers of sequencing reads from each component library in the Multiplex DNA Sample Library Pool upon completion of the experiment, it is very important to combine identical amounts of each independently amplified DNA sample library at this step. Accurate quantification and pipetting are critical.

Step 3. Prepare the Multiplex Hybridization Enhancing Oligo Pool

1. Thaw on ice the TS-HE Universal Oligo 1 (1,000 μM) and each TS-INV-HE Index oligo (1,000 μM) that matches a DNA Adapter Index included in the Multiplex DNA Sample Library Pool from Step 2 of this section.

2. Mix together the HE oligos so that the resulting Multiplex Hybridization Enhancing Oligo Pool contains, by mass, 50% TS-HE Universal Oligo 1 and 50% of a mixture of the appropriate TS-INV-HE Index oligos. The total combined mass of the Multiplex Hybridization Enhancing Oligo Pool should be 2,000 pmol, which is the amount required for a single Sequence Capture experiment.

Example: If a Multiplex DNA Sample Library Pool contains four DNA sample libraries prepared with Illumina TruSeq indexed adapters 2, 4, 6 and 8, respectively, then the Multiplex Hybridization Enhancing Oligo Pool would contain the following:

Component	Amount
TS-HE Universal Oligo 1	1,000 pmol (1 μl of 1,000 μM)
TS-INV-HE Index 2 Oligo	250 pmol (0.25 μl of 1,000 μM)
TS-INV-HE Index 4 Oligo	250 pmol (0.25 μl of 1,000 μM)
TS-INV-HE Index 6 Oligo	250 pmol (0.25 μl of 1,000 μM)
TS-INV-HE Index 8 Oligo	250 pmol (0.25 μl of 1,000 μM)
Total	2,000 pmol (2 μl of 1,000 μM)

Note: Due to the difficulty in accurately pipetting small volumes, it will be helpful where possible to prepare a larger volume Multiplex Hybridization Enhancing Oligo Pool and then to dispense this into individual single-use aliquots that can be stored at -15 to -25°C until needed. For optimal results, it is important that the individual TS-INV-HE oligos contained in a Multiplex Hybridization Enhancing Oligo Pool are precisely matched with the adapter indexes present in the Multiplex DNA Sample Library Pool in a multiplexed Sequence Capture experiment.

Step 4. Prepare the SeqCap EZ Library for hybridization

1. Remove a single 4.5 μl EZ Library aliquot (1 aliquot per multiplex capture experiment) from the -15 to -25°C freezer and allow it to thaw on ice.

Step 5. Prepare the hybridization sample

1. Add 5 μl of 1 mg/ml COT DNA and 1 μg of Multiplex DNA Sample Library Pool to a new 1.5 ml tube.

2. Add 2 μl of Multiplex Hybridization Enhancing Oligo Pool (2,000 pmol total mass) to the Multiplex DNA Sample Library Pool plus COT DNA.

3. Close the tube's lid and make a hole in the top of the tube's cap with an 18 - 20 gauge or smaller needle. The closed lid with a hole in the top of the tube's cap is a precaution to suppress contamination in the DNA vacuum concentrator.

4. Dry the Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool in a DNA vacuum concentrator on high heat (60°C). Denaturation of the DNA with high heat is not problematic after linker ligation because the hybridization utilizes single-stranded DNA.

5. To each dried-down Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool, add:

- 7.5 μl of 2X Hybridization Buffer (vial 5)

- 3 μl of Hybridization Component A (vial 6)

The tube with the Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool should now contain the following components:

Component	Solution Capture
COT DNA	5 µg
Multiplex DNA Sample Library Pool	1 µg
Multiplex Hybridization Enhancing Oligo Pool	2,000 pmol*
2X Hybridization Buffer (vial 5)	7.5 µl
Hybridization Component A (vial 6)	3 µl
Total	10.5 µl

* Composed of 50% (1,000 pmol) TS-HE Universal Oligo 1 and 50% (1,000 pmol) of a *mixture* of the appropriate TS-INV-HE Index oligos.

6. Cover the hole in the tube's cap with a sticker or small piece of laboratory tape.

7. Vortex the Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool plus Hybridization Cocktail (2X Hybridization Buffer + Hybridization Component A) for 10 seconds and centrifuge at maximum speed for 10 seconds.

8. Place the Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool/Hybridization Cocktail in a 95°C heat block for 10 minutes to denature the DNA.

9. Centrifuge the Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool/Hybridization Cocktail at maximum speed for 10 seconds at room temperature.

10. Transfer the Multiplex DNA Sample Library Pool/COT DNA/Multiplex Hybridization Enhancing Oligo Pool/Hybridization Cocktail to the 4.5 µl aliquot of EZ Library in a 0.2 ml PCR tube prepared in Chapter 2 of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0.*

11. Vortex for 3 seconds and centrifuge at maximum speed for 10 seconds.

The hybridization sample should now contain the following components:

Component	Solution Capture
COT DNA	5 µg
Multiplex DNA Sample Library Pool	1 µg
Multiplex Hybridization Enhancing Oligo Pool	2,000 pmol*
2X Hybridization Buffer (vial 5)	7.5 µl
Hybridization Component A (vial 6)	3 µl
EZ Library	4.5 µl
Total	15 µl

* Composed of 50% (1,000 pmol) TS-HE Universal Oligo 1 and 50% (1,000 pmol) of a *mixture* of the appropriate TS-INV-HE Index oligos.

12. Incubate in a thermocycler at 47°C for 64 - 72 hours. The thermocycler's heated lid should be turned on and set to maintain 57°C (+10°C above the hybridization temperature).

13. Proceed to Washing and Recovery of Captured DNA (Chapter 6 of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0*), and continue through to the end of that User's Guide.

Note: After proceeding back to Chapter 6 of the *NimbleGen SeqCap EZ Library SR User's Guide v3.0* to complete the Sequence Capture workflow, the term "captured DNA" can be considered to refer to the "captured Multiplex DNA Sample Library Pool" from this Application Note protocol.

NimbleGen

For life science research only.
Not for use in diagnostic procedures.

NIMBLEGEN and SEQCAP are trademarks of Roche.
All other product names and trademarks are the property of their respective owners.

06629725001 ❶ 12/11