illumina®

TruSeq® Custom Enrichment Kit

Multiplexed, targeted sample capture with high coverage uniformity and enrichment rates for flexibile study design with the easiest workflow.

Highlights

- **Proven Next-Generation Sequencing Technology and Custom Oligo Quality:**
  Most widely-adopted sequencing technology combined with oligo design expertise for confidence in your experiments.
- **Flexible Assay for Total Design Freedom:**
  Use DesignStudio to design and order probes for any human target regions with high coverage and enrichment rates.
- **Cost-Effective, Scalable Resequencing:**
  Pre-enrichment sample pooling of up to 12 reactions increases throughput with minimal hands-on time.
- **Integrated Solution:**
  Complete, supported workflow for design, sample preparation, sequencing, and data analysis.

Introduction

Targeted resequencing allows researchers to analyze a specific subset of the genome to discover and validate novel variants, examine specific genes in pathways, or as a follow-up to GWAS data. The TruSeq Custom Enrichment Kit is an in-solution capture assay for isolating customized human genomic regions, featuring master-mixed reagents supporting 12-plex enrichment.

With a simple and scalable workflow that includes custom probe design, gel-free sample prep, target enrichment, sequencing, and data analysis, the TruSeq Custom Enrichment Kit delivers high enrichment efficiency and coverage uniformity (Figure 1). Confidently pursue any human targeted resequencing study with a combination of assay design expertise and the most widely-adopted next-generation sequencing technology.

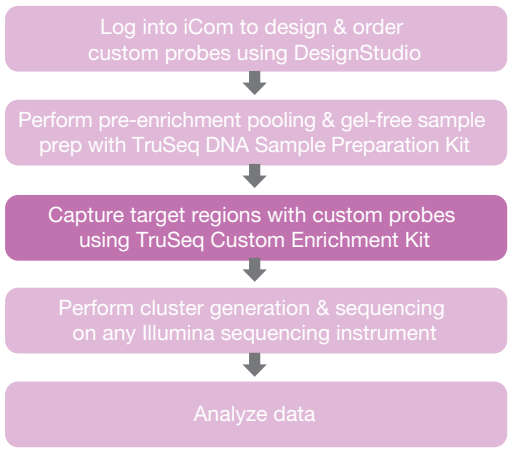
Custom Probe Design with DesignStudio

Prior to the enrichment assay, custom-selected oligonucleotide capture probes are designed and ordered online using DesignStudio. DesignStudio is a free, simple interface for custom enrichment probe design, providing dynamic feedback to optimize coverage and estimate total project pricing. After logging on to your personalized account, naming the project, and choosing the reference genome build of choice, you can select coordinates or gene names (individually or batch uploaded), and the probe design option of exons or full genomic regions. Capture probe design is automatically performed using an algorithm that considers a range of factors, including GC content, specificity, spacing, and coverage. Candidate probes are then visualized and assessed using estimated success scores (Figure 2). Probes can be filtered with user-defined tags, added to, or removed from the design. After visualization and QC, the custom probe library is added to the final design, and can then be ordered along with the recommended TruSeq kits needed for the project, including DNA Sample Preparation, Cluster Generation, and Sequencing by Synthesis Kits. All TruSeq Custom Enrichment Project information can be saved, reordered, or copied to a new project. To use DesignStudio, a MyIllumina account is required. Accounts can be requested at https://icom.illumina.com/Account/Register.

TruSeq DNA Sample Preparation Kits

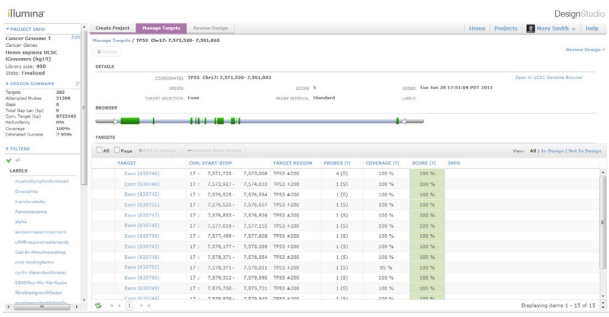
TruSeq Custom Enrichment is used with TruSeq DNA Sample Preparation Kits to create sample libraries for enrichment and subsequent sequencing. The Sample Preparation Kits provide automation-friendly master-mixed reagents, optimized index adapters, and a flexible workflow for preparing multiplexed samples. Multi-sample pooling and gel-free size selection dramatically reduces hands-on time, making large, high-throughput studies feasible and economical. For multiplexed sequencing, TruSeq DNA Sample Prep Kits come with the option of ordering two sets of 6 index adapters for a total of 12 unique indexes. The kits supply sufficient reagents to prepare 48 total DNA samples. TruSeq Sample Preparation enables processing of 96 samples per flow cell (12 samples x 8 lanes), or a maximum of 192 samples in two flow cells on a HiSeq® 2000 system. For more information, refer to the TruSeq Sample Preparation data sheet.[1]

Figure 1: Overview of Custom Sequence Capture as Part of Illumina Sequencing



The Illumina TruSeq Custom Enrichment Kit is an integral part of a complete and fully supported solution for targeted resequencing.

Figure 2: DesignStudio for Custom Probe Design



Manage Targets screen showing Target Region view in DesignStudio. Easily visualize genomic target regions and probes, and assess coverage and score. Design Summary metrics for the entire project are located on the left sidebar, along with project information and user defined labels for convenient data sorting during the design phase.

TruSeq Custom Enrichment Workflow

The custom enrichment workflow begins with pooled, indexed libraries of up to 12 samples that are denatured into single-stranded DNA (Figure 3A) and then hybridized to biotin-labeled custom oligonucleotide capture probes specific to the targeted region (Figure 3B). Streptavidin beads are added to bind to the biotinylated probes (Figure 3C). Biotinylated DNA fragments bound to the streptavidin beads are magnetically pulled down from the solution (Figure 4D). The enriched DNA fragments are then eluted from the beads and re-hybridized for a second enrichment reaction. After amplification of the enriched regions, the targeted library is ready for cluster generation and subsequent sequencing.

Data Analysis

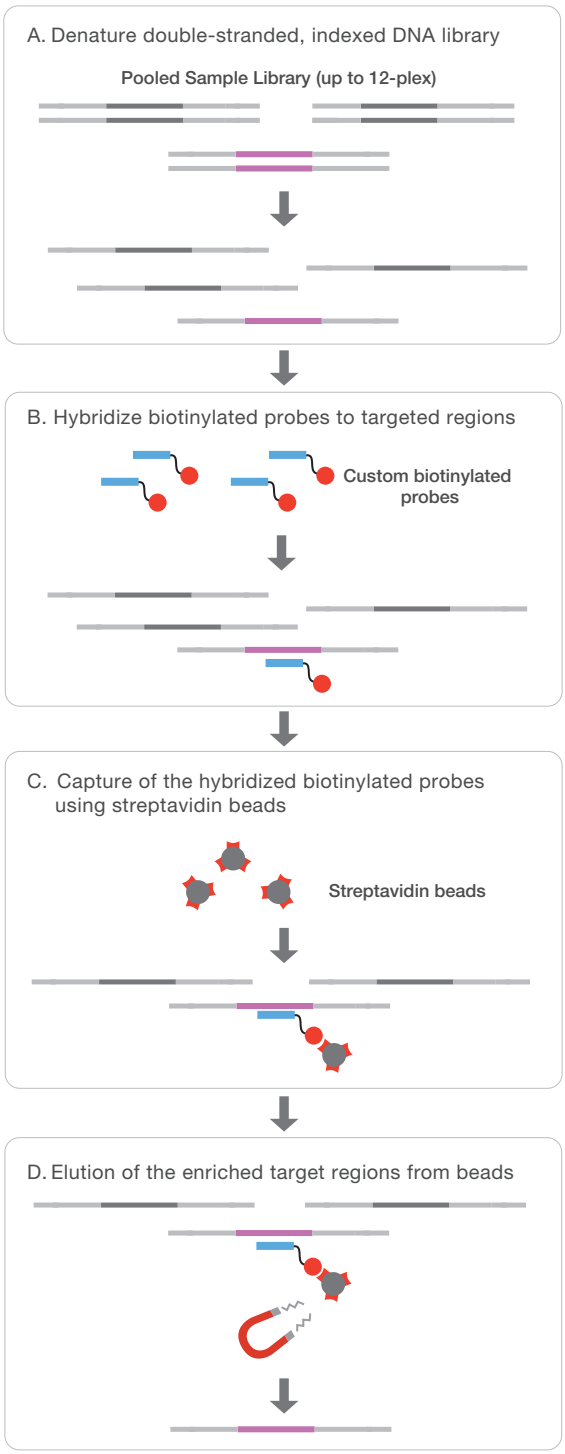
Sequence data generated from custom enrichment samples are analyzed using the TruSeq Enrichment Analysis Script to generate two sets of statistics: post-alignment and post-CASAVA (Consensus Assessment of Sequence and Variation) analysis. Post-alignment analysis counts the number of reads that overlap any targeted region and defines whether a read falls within a target. Post-CASAVA analysis calculates the coverage at each base within a region. Data are visualized using GenomeStudio® Data Analysis Software to examine the on-target and off-target coverage in a sample.

Optimizing Targeted Resequencing

To maximize the efficiency of targeted resequencing studies and ensure that sufficient coverage is obtained for highly sensitive variant calling, three key factors should be taken into account:

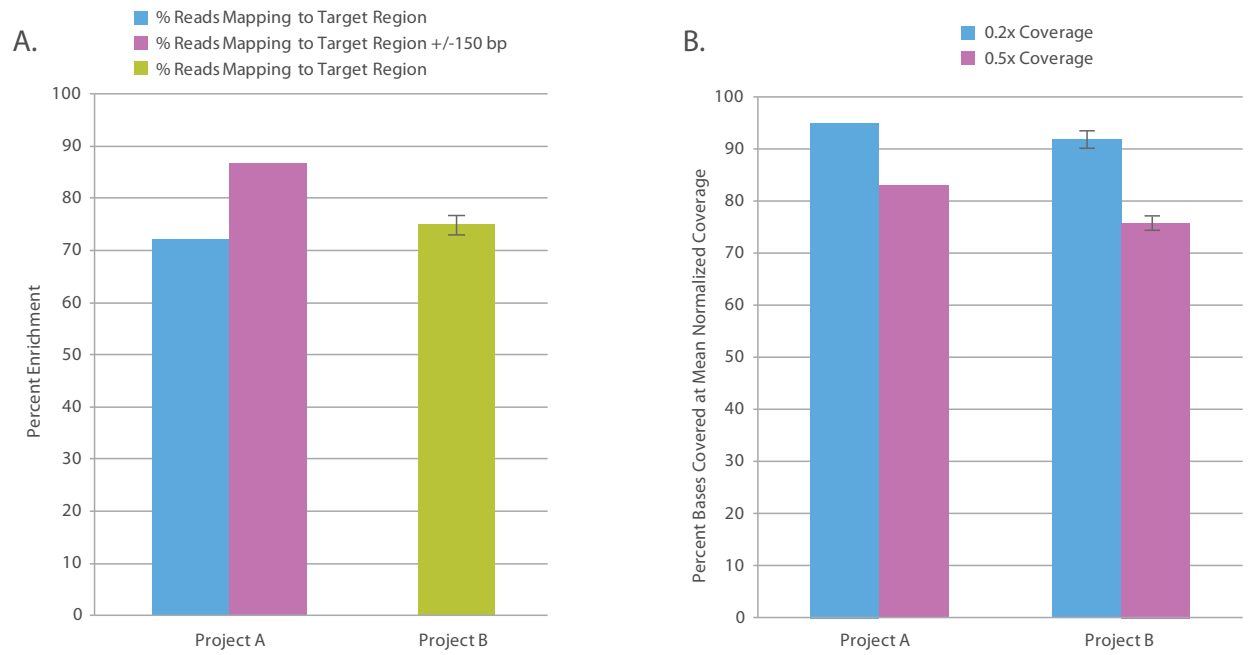
1. Sum length of targeted regions, equaling the total amount of targeted genomics sequence (500 kb–25 Mb)
2. Enrichment efficiency (percentage of reads passing filter and mapping to targeted regions)
3. Distribution of coverage depth for targeted regions

Figure 3: TruSeq Custom Enrichment Workflow



The TruSeq Custom Enrichment Kit provides a simple and streamlined in-solution method for isolating and enriching targeted regions of interest. Note that there are two successive rounds of enrichment (B, C, D) in the TruSeq Custom Enrichment workflow.

Figure 4: High Target Specificity and Coverage Uniformity



A. Percent enrichment, defined as reads mapping to target regions out of total reads per run, is shown for two example custom enrichment projects. For single-plex Project A, > 70% enrichment is achieved for reads mapping exactly to target regions (blue). An increase to > 85% is observed when the regions are expanded to +/- 150 bp surrounding the targeted coordinates (purple). For 12-sample multiplexed Project B, the mean percent enrichment averages ~75%, (green) are shown. B. Mean normalized coverage plots for the same two example projects show that > 90% of bases are covered at 0.2× of the mean coverage (blue bars), and > 75% of bases are covered at 0.5× of the mean coverage in both projects (purple bars).

These key parameters and a method for precalculating the amount of sequencing and mean coverage required to fully optimize any targeted sequencing study is described in greater detail in the Optimizing Coverage for Targeted Resequencing Technical Note.[2]

Data Examples

Two different TruSeq Custom Enrichment experiments were performed following the workflow described in Figure 1. Each project included different target regions, plexities, library sizes, target region sizes, probe interval spacing, and coverage depths (Table 1). Representative enrichment and coverage data are shown in Figure 4. Project A employed single-plex targeting of ~2 Mb total sequence with 20,000 attempted probes. Project B used a 12-plex strategy to target ~1.0 Mb of total sequence with 6,200 attempted probes. Both projects used gel-free TruSeq DNA Sample Preparation Kits prior to enrichment, and were sequenced using a Genome Analyzer$_{IIx}$. In both single- and multiplexed projects, high percent enrichment in targeted and padded regions was achieved, shown in Figure 4. For both projects, mean normalized coverage plots show that > 90% of bases are covered at 0.2× of the mean coverage, and > 75% of the bases are covered at 0.5× of the mean coverage.

Table 1: TruSeq Custom Enrichment Project Details

Detail	Project A	Project B
Unique Bases Targeted	~2.2 Mb	~1.0 Mb
Multiplex Level	1	12
Library Size	350 bp	400 bp
Full Region/Exon	Full Region	Exons
Probe Interval Spacing	Dense	Dense
Reference Sequence	UCSC hg19	UCSC hg19
Total Probes	~20K	~6.2K
Percent Enrichment*	72/87	~75
Percent Bases Covered**	95/83	91/75
Avg. Sequencing Depth†	47×	100×

*Percent enrichment shown as mapped only to target regions/mapped to target regions +/-150 bp (Project A), and to exons (Project B).

**Percent bases covered, shown as mean normalized coverage plots.

†Sequenced on the Genome Analyzer IIx.

TruSeq Custom Enrichment Details

Enrichment Efficiency*	> 55-60% (on target) > 60-65% (+/-150 bp)
Coverage Uniformity (0.2× mean)**	> 80%
Content Range	500 kb–25 Mb
Number of Oligos	2,500–67,200
Samples in Pre-Enrichment Pooling	Up to 12
Library Prep Input	1 µg
Library Size	250–600 bp

*Using gel-free sample preparation.
**Target values will vary due to custom designs.

Ordering Information

Kit*	Reactions/ Samples**	Catalog No.
TruSeq Custom Oligo Set	Variable	FC-121-0200
TruSeq Custom Enrichment Kit	4/48	FC-123-1004
	8/96	FC-123-1008
	24/288	FC-123-1024
	96/1152	FC-123-1096
TruSeq DNA Sample Preparation Kit	48 samples/ 6 indexes per set	FC-121-1001
		FC-121-1002

*Use the DesignStudio Project Calculator to determine the correct type and number of cluster generation and sequencing reagent kits needed for your instrument type, desired read length, and sequencing depth.

Summary

By harnessing the power of Illumina sequencing, the TruSeq Custom Enrichment Kit provides multiplexed, targeted sample capture with the highest coverage uniformity and enrichment rates, enabling flexible study design with the easiest workflow.

Learn More

To learn more about complete solutions for targeted resequencing, visit www.illumina.com/applications/sequencing/targeted_resequencing.ilmn.

References

1. TruSeq DNA Sample Preparation Kit Data Sheet.

2. Optimizing Coverage for Targeted Resequencing Technical Note.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2012 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2011-007 Current as of 19 Mar 2012

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA