illumina®

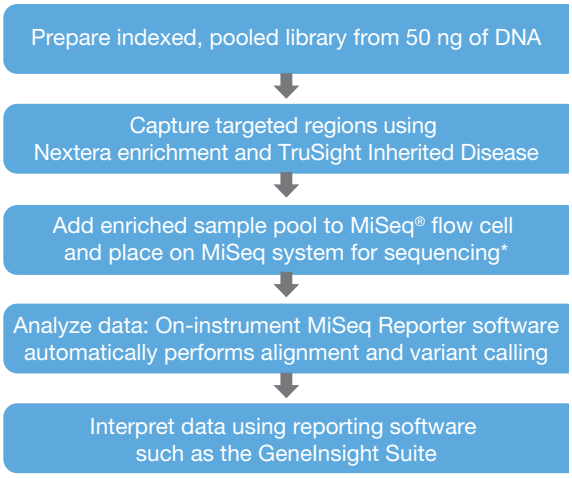
Nextera® Exome Enrichment Kit

Delivering the most comprehensive exome coverage with the fastest workflow and lowest sample input.

Highlights

- **Fastest and Simplest Enrichment Workflow**
  Reduces total sample preparation and enrichment time to just 2.5 days
- **Lowest Sample Input Expands Studies**
  Excellent data quality with 50 ng input to preserve precious samples
- **Most Comprehensive Exome Coverage**
  Comprehensive coverage of 62 Mb exome sequence with both RefSeq coding and UTR content
- **Integrated Solution for Simple, Scalable Workflow**
  All-in-one optimized kit offers a streamlined, automation-friendly workflow

Figure 1: Fastest Exome Enrichment Workflow



The Nextera Exome Enrichment Kit integrates library sample preparation and enrichment steps to offer a fast, streamlined, and optimized workflow, delivering fully enriched libraries for up to 96 samples in just 2.5 days.

Introduction

Identifying variants linked with disease is now simpler and faster with the Nextera Exome Enrichment Kit, which offers a single, integrated sample preparation and enrichment workflow that can be completed in just 2.5 days (Figure 1). Delivering excellent data quality from low sample input (50 ng), Nextera Exome Enrichment enables researchers to access precious samples, while retaining sufficient material for future analyses. Developed, tested and optimized for Illumina sequencing technology, the kit offers the most integrated, end-to-end exome resequencing solution on the market.

Integrated Library Preparation and Exome Enrichment Workflow

Leveraging the speed of Nextera sample library preparation technology, no mechanical DNA fragmentation requirement, and unique 12-plex pre-enrichment sample pooling, the Nextera Exome Enrichment method reduces hands-on time for a cost-effective, high-throughput workflow that saves at least one full day over all other currently available enrichment workflows (Figure 2). Master-mixed reagents are coupled with plate-based processing for up to eight enrichment reactions (96 total samples), and volumes are optimized for liquid handlers to make the process automation-friendly for even higher throughput.

The workflow begins with breakthrough Nextera technology that enables sample libraries to be prepared in less than 3 hours, with no need for mechanical DNA shearing. A single "tagmentation" reaction simultaneously fragments and tags DNA. Sequencing adaptors and indices are then added by PCR (Figure 2A). Using the proven TruSeq® Exome Enrichment technology, sample libraries are then denatured into single-stranded DNA (Figure 2B) and hybridized to biotin-labeled probes specific to the targeted region (Figure 2C). The pool is enriched for the desired regions by adding streptavidin beads that bind to the biotinylated probes (Figure 2D). Biotinylated DNA fragments bound to the streptavidin beads are magnetically pulled down from the solution (Figure 2E). The enriched DNA fragments are then eluted from the beads and hybridized for a second enrichment reaction. After amplification, a targeted library is ready for cluster generation and subsequent sequencing. This unique, integrated sample preparation and enrichment workflow delivers fully enriched libraries for up to 96 samples in just 2.5 days.

Superior Exome Coverage and More

The Nextera Exome Enrichment Kit supports the discovery of the greatest number of variants, featuring a highly optimized probe set that delivers comprehensive coverage of exomic sequence, starting from only 50 ng of DNA input. The kit includes > 340,000 95mer probes, each constructed against the human NCBI37/hg19 reference genome. The probe set was designed to enrich > 200,000 exons, spanning 20,794 genes of interest (Table 1).

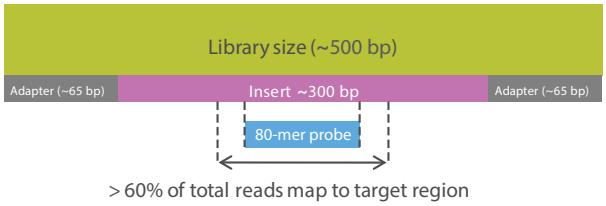
While the sum length of these probes is 32 Mb, the kit actually targets 62 Mb of the human genome (117.5 Mb if the 150 bp regions captured upstream and downstream of target are also considered). Each 95mer probe targets libraries of 300–350 bp (insert size of 170–220 bp), enriching 265–465 bases centered symmetrically around the midpoint

of the probe (Figure 3)[2]. This means that, in addition to comprehensive coverage of the major exon databases (Table 2), the kit also provides broad coverage of non-coding DNA in exon-flanking regions (promoters and UTRs). Genome-wide association studies suggest that > 80% of disease-associated variants fall outside coding regions[3]. Analysis of these regions enables researchers to discover variants that effect gene function, at a more affordable price than whole-genome sequencing[4].

Table 1: Coverage Details

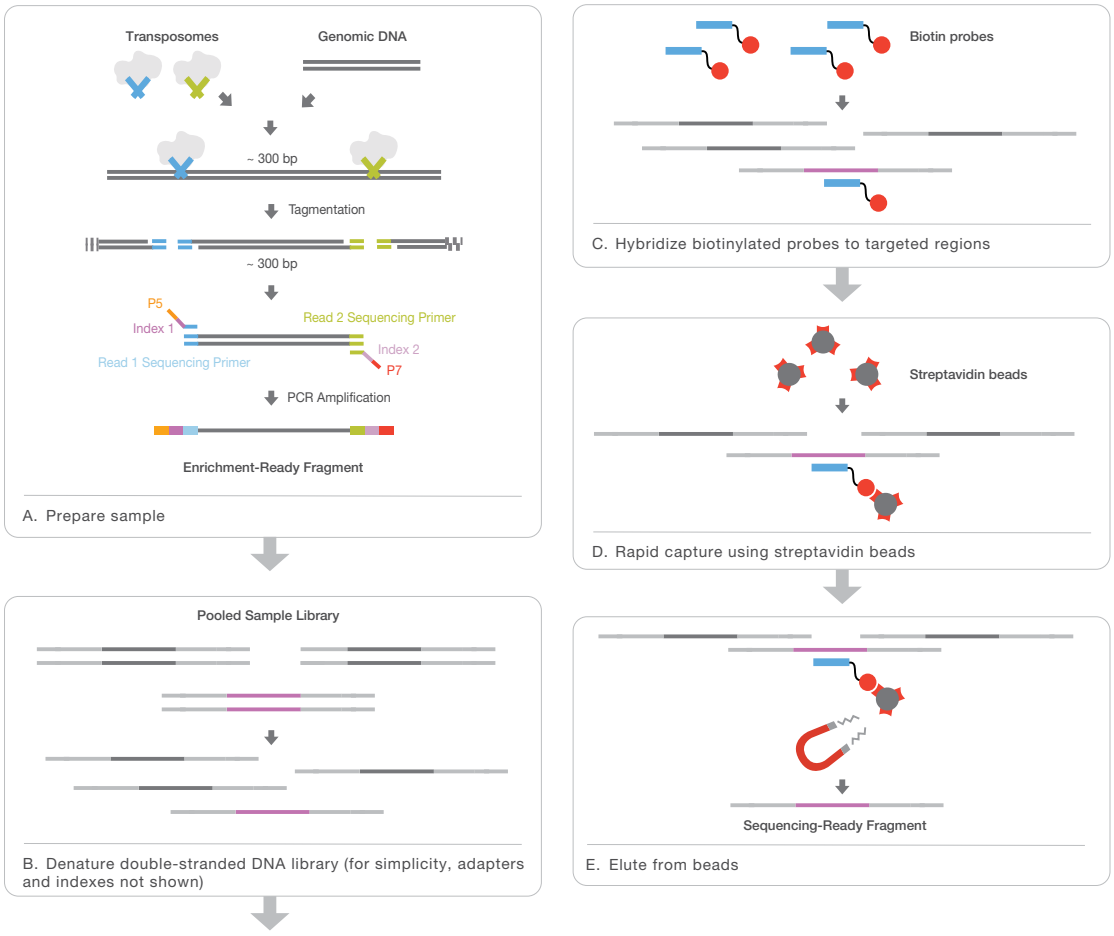
Target region size	62 Mb
Number of target genes	20,794
Number of target exons	201,121
Probe size	95-mer
Number of probes	340,427
Recommended library size	300–350 bp
Percent bases covered at 0.2x mean coverage	> 80%

Figure 3: Probe Footprint



With a 300–350 bp DNA library (mean insert size = 230 bp), the probe will enrich 365 bp (2 × Insert – Probe) centered around its midpoint.

Figure 2: Integrated Nextera Exome Enrichment Workflow



The Nextera Exome Enrichment Kit provides a fast, simple method for isolating the human exome. The streamlined, automation-friendly workflow combines library preparation and exome enrichment steps, and can be easily completed in 2.5 days with minimum hands-on time.

Table 2: Databases Covered by the Nextera Exome Enrichment Kit

Database	% Database Covered	Description	Web Address
CCDS coding exons (31.3 Mb; hg19)	97.2%	Core set of human protein coding regions that are consistently annotated and of high quality	http://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi
RefSeq (regGene) coding exons (33.2 Mb; hg19)	96.4%	Known protein-coding genes taken from the NCBI RNA reference collection	http://www.ncbi.nlm.nih.gov/RefSeq/
RefSeq (regGene) exons plus (67.8 Mb; hg19)*	88.3%	Known protein-coding genes taken from the NCBI RNA reference collection along with non-coding DNA	http://www.ncbi.nlm.nih.gov/RefSeq/
Encode/Gencode coding exons (Encyclopedia of DNA Elements) (25.6 Mb; hg19)†	93.2%	Project to identify all functional elements in the human genome	http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=183763205&c=chr13&g=wgEncodeGencode
Predicted microRNA targets (9.0 Mb, hg19) ‡	77.6%	Includes predicted microRNA targets	http://www.microrna.org/microrna/get-Downloads.do

* Includes coding exons, 5' UTR, 3' UTR, microRNA, and other non coding RNA
† Manual V4
‡ mirbase 15 targets predicted by www.microrna.org

Highest Efficiency Protocol

For targeted resequencing, high enrichment efficiency and coverage uniformity ensure that all targeted regions are sequenced and minimize the required sequencing depth to accurately determine variants without bias. The Nextera Exome Enrichment Kit is designed and optimized to deliver high enrichment rates and on-target specificity, while ensuring the highest coverage uniformity and reproducibility (Figures 4–5). Greater than 60% of reads that pass filter and map to the reference genome will align to the targeted region, and > 70% will align within 150 bases of the targeted region. The kit produces excellent uniformity of coverage for larger libraries, effectively capturing variance across library sizes. This not only increases the uniformity of coverage for smaller exons (< 150 bp), but also across long coding exons, UTRs, and non-coding RNA targets.

With the high-throughput processing power of Illumina sequencing systems, multiple exomes can be sequenced in a single run, reducing cost and minimizing hands-on time.

Data Assessment

Sequence data generated from exome enrichment samples are analyzed using a 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 can be visualized to examine the on-target and off-target coverage in a sample using GenomeStudio® Data Analysis Software.

Summary

With a fast and easy workflow, and low sample DNA input, the Nextera Exome Enrichment Kit offers researchers a highly efficient exome resequencing solution that preserves precious samples. Providing high enrichment rates and exceptional coverage uniformity, it allows for the identification of the greatest number of variants to accelerate disease research.

Learn more, visit:
www.illumina.com/applications/sequencing/targeted_resequencing.ilmn

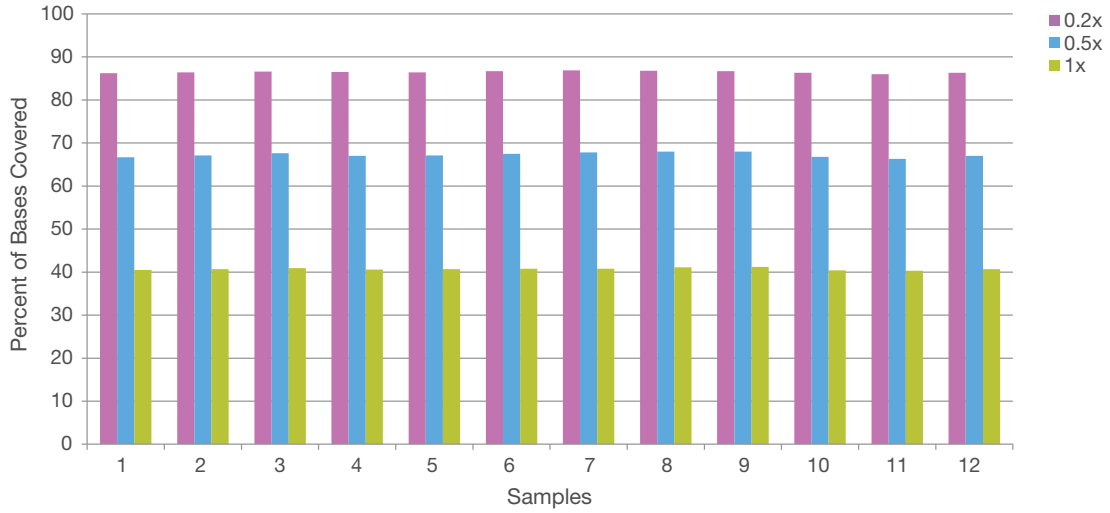
References

1. Clark MJ, Chen R, Lam NY, Karczewski KJ, Chen R, et al. (2011) Performance comparison of exome DNA sequencing technologies. Nat Biotechnology 29:908–914.
2. Optimizing Coverage for Targeted Resequencing Technical Note. www.illumina.com/documents/products/technotes/technote_optimizing_coverage_for_targeted_resequencing.pdf
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex disease. Nature 4618: 747-753.
4. Bainbridge, MN, Wang M, Wu YQ, Newsham I, Muzny DM, et al. (2011) Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biology 12(7):R68.

Ordering Information

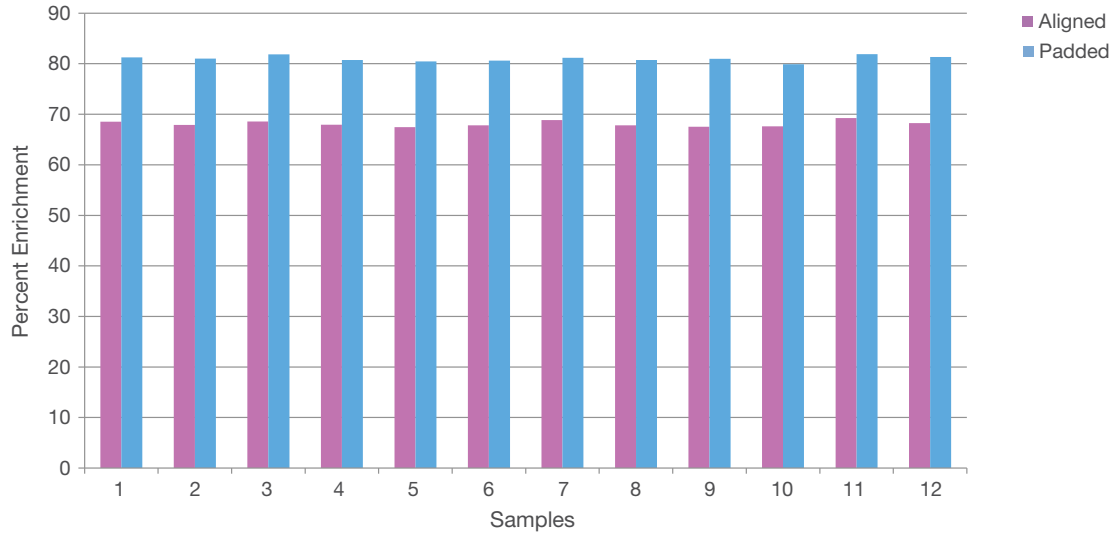
Catalog No.	Kit	Description
FC-121-1204	Nextera Exome Enrichment Kit (48 samples)	Contains reagents for preparing and enriching up to 48 samples at 12-plex. Includes 14 dual-indexes supporting 24 index combinations.
FC-121-1208	Nextera Exome Enrichment Kit (96 samples)	Contains reagents for preparing and enriching up to 96 samples at 12-plex. Includes 14 dual-indexes supporting 24 index combinations.

Figure 4: Highest Coverage Uniformity



Coverage uniformity is given for 12 samples with respect to the percentage of bases covered across the 62 Mb target region at varying mean normalized read depths. The 12 samples were prepared and then simultaneously enriched using the Nextera Exome Enrichment Kit. The pooled samples were sequenced across four lanes of a HiSeq® flow cell, generating mean read depths of 24–29× (varying for each sample). Over 80% of bases were covered at 0.2× mean coverage.

Figure 5: High-Target Specificity



Percent Enrichment is defined as the number of reads mapping to the targeted regions out of the total reads produced in a sequencing run (on a per-sample basis). The 12 samples shown here were prepared and then simultaneously enriched using the Nextera Exome Enrichment Kit. A 68% enrichment (purple bars) was achieved when considering only the reads that mapped to bases in the regions targeted by the capture probes. An increase to > 80% enrichment (blue bars) is observed when the assessed reads are expanded to include those that map to regions +/- 150 bp flanking the probe-targeted region. This occurs because each probe is designed to capture more sequence than just the absolute targeted region.

Illumina, Inc. • 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, NuPCR, 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. 770-2012-028 Current as of 18 July 2012

illumina®

AATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTA
AACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAG
ACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGT
GTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACA
AATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTA
TTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACGAA
TATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTAAC