Whole-Exome Sequencing with Nextera™ Flex for Enrichment

Library preparation and unique dual indexing combined with hybrid capture of exome content from Illumina or third-party oligo vendors.

Introduction

Although the human exome is only 1-2% of the entire human genome, it consists of coding regions of expressed genes and represents the most functionally relevant part of the human genome. Exome sequencing has gained recognition in the scientific community as a powerful method for discovering potential causative variants for genetically driven diseases.[1-3] Focusing sequencing efforts on this small portion of the genome has proven to be a time- and cost-effective method for identifying variants linked with diseases and other health conditions. The Nextera DNA Flex Pre-Enrichment Library Prep and Enrichment Reagents Kit (Illumina, Cat. No. 20025524) with the Illumina Exome Panel (Illumina, Cat. No. 20020183) delivers highly accurate human exome sequencing data using a workflow that combines the rapid and innovative Nextera Flex library preparation chemistry with streptavidin bead-based exome enrichment (Figure 1). Nextera Flex for Enrichment human exome libraries are compatible with all Illumina sequencing systems, but higher throughput instruments are recommended.

A key component of the Nextera Flex chemistry is On-Bead Tagmentation, which uses bead-linked transposomes to mediate a uniform tagmentation reaction (tagmentation combines DNA fragmentation and adapter ligation into a single reaction). On-Bead Tagmentation offers several significant advantages such as high tolerance to varying DNA sample input amounts and superior coverage uniformity compared to solution based tagmentation methods.[4] In addition to On-Bead Tagmentation chemistry, Nextera Flex for Enrichment offers another unique benefit–it enables whole-exome sequencing from Illumina and third-party oligo vendors. This flexibility supports a wide range of panel types including custom, fixed, and whole-exome sequencing panels. Here we describe the successful library preparation and sequencing of six exome libraries using Nextera Flex for Enrichment with the Illumina Exome Panel and three additional third-party exome panels.

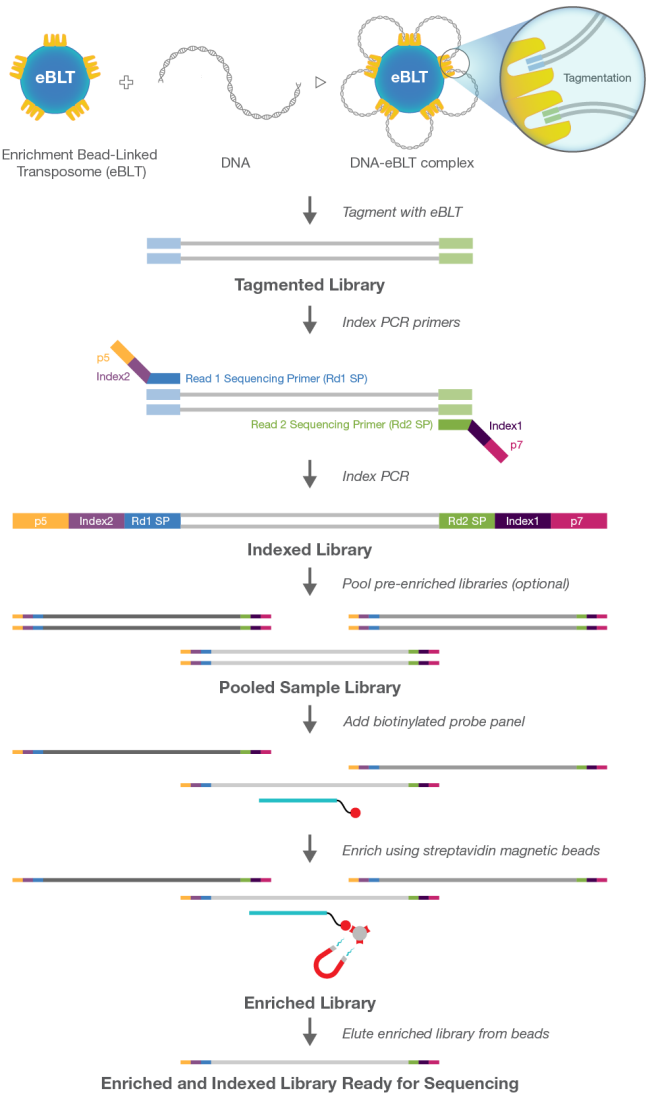

Figure 1: Nextera Flex for Enrichment assay chemistry–On-Bead Tagmentation followed by a single hybridization reaction enables a fast and flexible workflow.

Methods

Library preparation and sequencing

Six sequencing libraries were prepared from human NA12878 gDNA (Coriell Institute) using the Nextera DNA Flex Pre-Enrichment Library Prep and Enrichment Reagents Kit (Illumina, Cat. No. 20025524), the Illumina Exome Panel (Illumina, Cat. No. 20020183), and three different third-party exome panels (Table 1). Enrichment reactions were performed in 12-plex with IDT for Illumina Nextera DNA UD Indexes (Illumina, Cat. No. 20027213). Company D exome libraries were prepared with three different enrichment hybridization times: 1.5, 4.0, and 16 hours. All libraries except Company A exome libraries were

Table 1: Exome panel specifications

	Illumina Exome Panel	Company A	Company T	Company D
Exome panel features				
Panel size	45.2 Mb	36 Mb	33 Mb	39 Mb
Probe size	80 bp	N/A	120 bp	120 bp
Probe type	ssDNA	RNA	dsDNA	ssDNA
Enrichment hyb time	1.5 hr	16 hr	1.5 hr	1.5, 4.0, and 16 hr
Databases used for exome panel design				
RefSeq	99.83	99.88	99.08	99.45
GENCODE	98.02	97.29	96.01	96.82
CCDS	99.99	99.91	99.76	99.67
UCSC Known Genes	99.89	98.72	97.63	98.13
clinVar	84.95	73.41	72.56	72.9

a. RefSeq - NCBI Reference Sequence Database. Accessed June 15, 2019.

b. CCDS - Consensus CDS (CCDS) Database. Accessed June 15, 2019.

c. ENSEMBL - Ensembl Genome Browser. Accessed June 15, 2019.

d. GENCODE - GENCODE Project: Encyclopedia of genes and gene variants. Accessed June 15, 2019.

e. clinVar - NCBI Clinical Variance database. Accessed June 15, 2019.

f. **Panel size** - the total length of sequence in the target regions, **Probe size** - length of enrichment pull-down probe, **Probe type** - probe oligonucleotides can be RNA, DNA, single stranded (ss), or double stranded (ds).

sequenced on a NovaSeq™ 6000 System using NovaSeq 6000 S2 Reagent Kits (Illumina, Cat. No. 20012861) with a run configuration of 2 × 101 bp. Company A exome libraries were sequenced on the NextSeq™ 550 System using NextSeq 500/550 High Output Kit v2.5 Reagent Kits (Illumina, Cat. No. 20024908) with a run configuration of 2 × 125 bp, as recommended by Company A.

Data analysis

Analysis was performed using the BaseSpace™ Enrichment App.[5] The Enrichment App is a rapid alignment and variant detection tool designed for targeted sequencing data. The analysis workflow maps reads, performs small, structural, and copy number variant calling, annotates variants, and calculates enrichment metrics. Small variant precision and recall metrics were calculated by extraction of shared regions among the Illumina, Company A, T, and D panels. The shared regions (32.5 Mb) were then analyzed with the Variant Calling Access Tool[6] in BaseSpace Sequence Hub. All data sets in this application note are publicly available in the BaseSpace Sequencing Hub under the Public Data tab.

Table 2: Enrichment performance and sequencing coverage performance*

	Illumina Exome Panel (1.5 hr hyb)	Company A (16 hr hyb)	Company T (1.5 hr hyb)	Company D (1.5 hr hyb)	Company D (4.0 hr hyb)	Company D (16 hr hyb)
Mean target coverage depth	60.7x	66.4x	80.9x	51.5x	64x	69.7x
Autosome callability	92.6%	92.9%	96.9%	94.6%	95.4%	96.0%
Total aligned reads	47.8 million	48.1 million	48.2 million	48.9 million	49.2 million	49.4 million
Targeted aligned reads	33.7 million	32.6 million	35.6 million	26.6 million	33.1 million	36.3 million
Duplicate aligned reads	10.5%	6.6%	10.2%	33.5%	15.5%	10.0%
Read enrichment	78.6%	72.5%	82.4%	81.8%	79.6%	81.5%
Padded read enrichment	84.6%	82.6%	91.2%	91.5%	90.9%	92.4%
Uniformity of Coverage (Pct > 0.2*mean)	94.0%	94.4%	97.4%	96.2%	95.9%	96.2%

*All data sets were down sampled to 50 million reads. Data represents example comparison data. Actual performance specifications may vary.

Mean target coverage depth - The average sequencing depth for targeted bases.

Autosome callability - The percent of non-N reference positions in targeted regions in autosomal chromosomes with a passing genotype call.

Total aligned reads - The number of pass-filter reads that are aligned and are not flagged as duplicates.

Targeted aligned reads - The number of reads that align to the target regions (excluding non-PF, unmapped, or duplicate reads).

Duplicate aligned reads - The percentage of paired-end reads that are flagged as duplicates.

Read enrichment - The percentage of reads that align to the target regions (excluding non-PF, unmapped, or duplicate reads).

Padded read enrichment - The percentage of padded reads that align to the target regions (excluding non-PF, unmapped, or duplicate reads).

Uniformity of coverage (Pct > 0.2*mean) - The percentage of targeted base positions in which the read depth is greater than 0.2 times the mean region target coverage depth.

Table 3: Small variant analysis*

	Illumina Exome Panel (1.5 hr hyb)	Company A (16 hr hyb)	Company T (1.5 hr hyb)	Company D (1.5 hr hyb)	Company D (4.0 hr hyb)	Company D (16 hr hyb)
Total variants passing filter						
SNVs	406,744	655,862	172,146	170,621	198,447	177,989
Insertions	4678	4031	3701	3609	4379	4328
Deletions	5531	6230	4193	4146	4940	4880
Variants in dbSNP						
SNVs	94.0	93.9	94.6	94.6	94.7	94.8
Insertions	87.1	87.3	83.0	87.8	87.6	86.2
Deletions	92.9	89.9	93.7	91.6	91.9	92.3
Variants in genes						
SNVs	29,951	22,171	20,686	25,605	25,616	25,707
Insertions	987	329	287	602	622	633
Deletions	1009	360	263	604	630	640
Variants in UTR regions						
SNVs	1238	309	62	1334	1336	1345
Insertions	90	12	4	113	122	124
Deletions	76	21	2	101	108	108
Non-synonymous variants						
SNVs	9699	9452	9461	9665	9634	9685
Insertions	128	118	135	134	140	141
Deletions	76	21	2	101	108	108

*All data sets were down sampled to 50 million reads. Data represents example comparison data. Actual performance specifications may vary.
Definitions: **SNVs** = single nucleotide variants, **Genes** = exons, introns and UTR regions, **UTR regions** = 5'and 3' UTR regions.

Table 4: Small variant precision and recall analysis*

	Illumina Exome Panel (1.5 hr hyb)	Company A (16 hr hyb)	Company T (1.5 hr hyb)	Company D (1.5 hr hyb)	Company D (4.0 hr hyb)	Company D (16 hr hyb)
SNV recall	94.3%	93.0%	95.1%	81.0%	94.8%	95.1%
SNV precision	99.4%	99.0%	99.4%	85.7%	99.6%	99.6%
Indel recall	85.4%	85.9%	87.2%	74.0%	88.3%	87.8%
Indel precision	83.0%	79.8%	81.7%	71.1%	82.5%	81.2%

*Precision and recall metrics were calculated using only the shared regions between the four exome panels (32.5 Mb). Data represents example comparison data. Actual performance specifications may vary.
Definitions: **Precision (accuracy)** = calculated as the ratio of [# of True Positive Calls/(# of True Positive Calls + # of False Positive Calls)], and **Recall (sensitivity)** = calculated as the ratio of [# of True Positive Calls/(# of True Positive Calls + # of False Negative Calls)].

Results

Enrichment and sequencing coverage performance

A key benefit of Nextera Flex for Enrichment chemistry is that On-Bead Tagmentation enables a highly uniform bead-based normalization. After the bead-linked transposomes are saturated with DNA, no additional tagmentation can occur delivering highly uniform saturation-based normalization. This saturation-based normalization is the basis for uniform and consistent insert sizes, tolerance to DNA input variability, consistent library yields, and extremely high coverage uniformity for exome enriched libraries.[4]

Nextera Flex for Enrichment delivers excellent enrichment and coverage uniformity with the Illumina Exome Panel as well as all third-party exome panels. Across all six exome libraries (with padding sizes of 150 bp for all libraries), padded read enrichment values ranged from 79.94% to 91.47%, with the Illumina Exome Panel delivering 84.81% padded read enrichment (Table 2). In addition to comparable enrichment values, the six libraries demonstrated high uniformity of coverage and high genotype-calling values. Coverage uniformity ranged from 94.40% to 97.45% while autosomal callability values ranged from 93.06% to 96.52%.

Small variant analysis

Nextera Flex for Enrichment enables accurate identification of single nucleotide variants (SNVs) and insertions/deletions (indels).[4] The Nextera Flex for Enrichment exome libraries show high percentages of variants found in dbSNP for SNVs and indels (Table 3). The data also shows comparable values across all six exomes. Other metrics that depend on the total size of the content

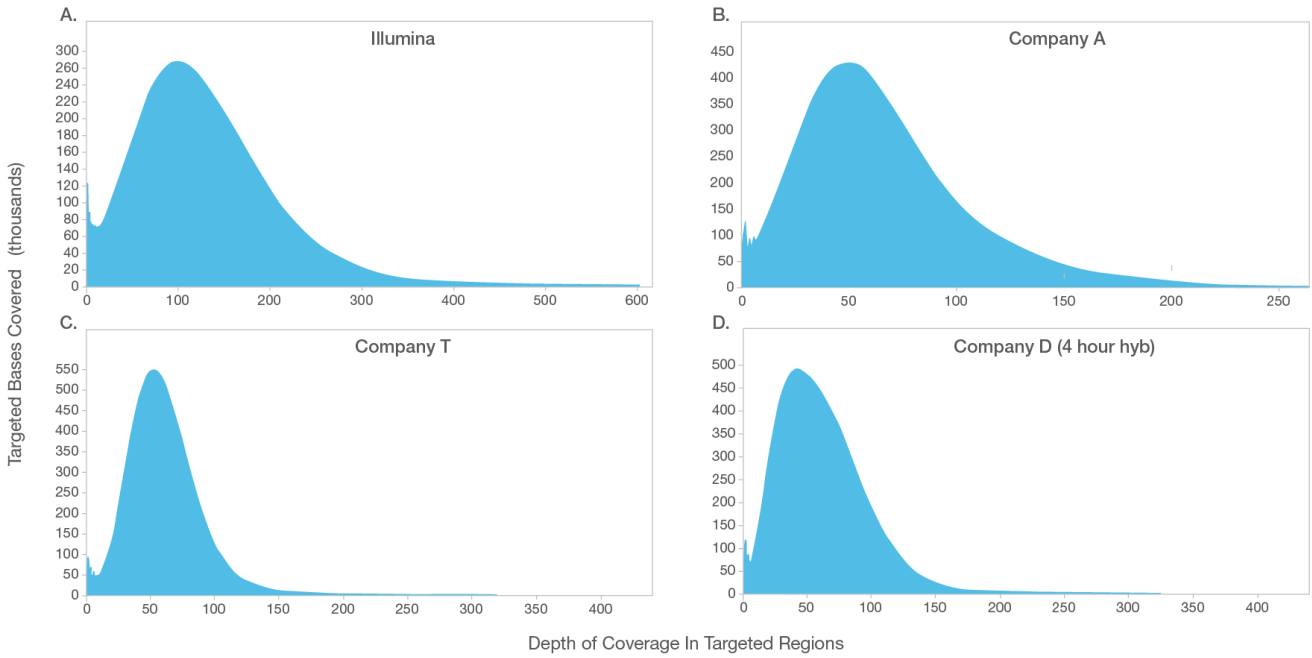

Figure 2: Nextera Flex for Enrichment coverage depth–Coverage histograms for the Illumina Exome Panel, Company A, Company T, and Company D show a Poisson -like distribution, which indicates even, uniform coverage across the target region.

panel may show greater variation from exome to exome. The Illumina Exome Panel has the largest panel size and includes content from UTR regions. These differences are reflected in metrics that depend on total panel size and UTR content such as total variants passing filter, variants in genes, and variants in UTR regions.

Precision and recall metrics show equivalency across 5 exome panels (excluding the Company D 1.5 hour hyb panel) indicating high accuracy and sensitivity for small variant detection of SNVs and indels (Table 4). The Company D 1.5 hour hyb panel demonstrated slightly lower performance metrics for precision and recall.

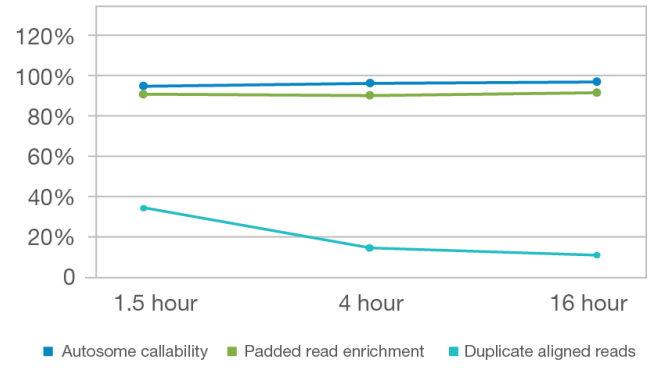

Figure 3: Different hybridization times affect sequencing and variant calling metrics–Company D panel libraries were prepared with a range of enrichment hybridization times from 1.5-16 hours. Percent duplicate aligned reads, percent autosome callability, and percent read enrichment results are shown.

Coverage depth and uniformity

Nextera Flex for Enrichment with On-Bead Tagmentation chemistry delivers excellent coverage uniformity due to the production of consistent target insert sizes and uniform, saturation-based normalization.[7] While mean coverage is a function of experimental design, coverage uniformity is a function of library preparation and sequencing quality. Coverage histograms with a smooth, Poisson-like distribution indicate excellent coverage uniformity with no gaps in coverage across the target regions. The Illumina exome panel as well as the third-party exome panels all display even, uniform coverage distribution (Figure 2).

Effects of enrichment hybridization time on sequencing data

Three libraries were prepared with Nextera Flex for Enrichment using the Company D exome panel. The enrichment hybridization reactions were carried out for 1.5, 4.0, and 16 hours. A comparison of hybridization times to enrichment metrics show that percent read enrichment and percent autosome callability were not significantly affected by increasing hybridization time. However, percent duplicate aligned reads did show a significant drop in response to increasing enrichment hybridization times (Figure 3). In addition, the 1.5 hour hyb panel demonstrated slightly lower performance metrics for precision and recall compared to the 4.0 and 16 hour results.

Summary

Nextera Flex for Enrichment with the Illumina Exome Panel leverages the advantages of targeted exome sequencing with the speed and consistency of On-Bead Tagmentation. Furthermore, Nextera Flex for Enrichment is compatible with third-party enrichment probes/panels, which supports flexible experimental design and content portability. With exome panels that cut across multiple providers, probe types, and panel sizes, Nextera Flex for Enrichment shows equivalent performance across key metrics, such as read enrichment, uniformity of coverage, and small variant callability.

Ordering information

Order Nextera Flex products online at www.illumina.com.

Product	Catalog No.
Nextera DNA Flex Pre-Enrichment Library Prep and Enrichment Reagents 96 samples (8, 12-plex enrichment reactions)	20025524
Nextera DNA Flex Pre-Enrichment Library Prep and Enrichment Reagents 16 samples (16, 1-plex enrichment reactions)	20025523
Nextera DNA Flex Pre-Enrichment Library Prep Reagents (96 samples)	20025520
Nextera DNA Flex Pre-Enrichment Library Prep Reagents (16 samples)	20025519
Illumina Exome Panel (8 enrichment reactions)	20020183
IDT for Illumina - Nextera DNA Unique Dual Indexes - Set A (96 indexes, 96 samples)	20027213
IDT for Illumina - Nextera DNA Unique Dual Indexes - Set B (96 indexes, 96 samples)	20027214
IDT for Illumina - Nextera DNA Unique Dual Indexes - Set C (96 indexes, 96 samples)	20027215
IDT for Illumina - Nextera DNA Unique Dual Indexes - Set D (96 indexes, 96 samples)	20027216
IDT for Illumina - Nextera DNA Unique Dual Indexes - Sets A-D (384 indexes, 384 samples)	20027217

Learn more

To learn more about the visit Nextera Flex for Enrichment, visit www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-flex-enrichment.html?langsel=/us/

To view the data sets in this application note, sign in to MyIllumina, then select:

- Illumina Exome Panel
- Company A
- Company T
- Company D (1.5 hr hyb)
- Company D (4.0 hr hyb)
- Company D (16 hr hyb)

References

1. Litchfield K, Summersgill B, Yost S, et al. Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nat Commun.* 2015;6:5973.
2. Srivastava S, Cohen JS, Vernon H, et al. Clinical whole exome sequencing in child neurology practice. *Ann Neurol.* 2014;76:473-483.
3. Worthey EA, Mayer AN, Syverson GD, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med.* 2011;13:255-262.
4. Illumina (2017). Nextera DNA Flex Library Preparation Kit Data Sheet. Accessed June 15, 2019.
5. Illumina BaseSpace Sequence Hub. BaseSpace Enrichment App. Accessed June 15, 2019.
6. Illumina BaseSpace Sequence Hub. BaseSpace Variant Calling Assessment Tool. Accessed July 25,2019.
7. Illumina (2018). Nextera Flex for Enrichment Data Sheet. Accessed June 15, 2019.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com