Comparison of TruSeq® Library Preparation Kits

Find the best library prep solution for your research.

Introduction

The TruSeq family of library preparation solutions includes the original TruSeq DNA method, the TruSeq DNA PCR-Free kit for superior genomic coverage, and the low-input TruSeq Nano DNA kits. All three kits are based on the widely adopted TruSeq workflow and can be used for virtually any sequencing application, including whole-genome sequencing, resequencing, *de novo* assembly, and metagenomics studies. However, the TruSeq Nano DNA and TruSeq DNA PCR-Free kits offer improvements in coverage and uniformity while reducing bias. Due to these improvements, the original TruSeq DNA kits will be discontinued after December 2013. This technical note presents a comparison of data quality and features of the TruSeq DNA kits so researchers can identify the ideal solution for each sequencing project.

Data Comparison

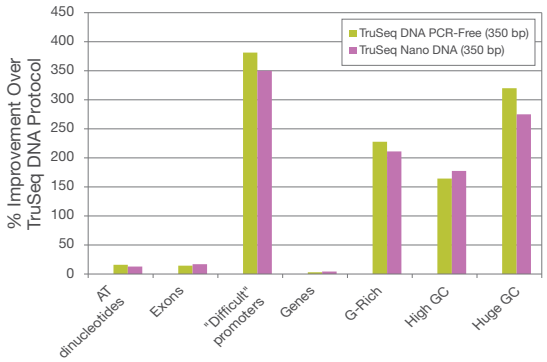
Genome Coverage Quality

Coverage describes the average number of reads that align to known reference bases. The coverage often determines the confidence level at which variant discovery can be made at any base position. At higher levels of coverage, a greater number of aligned reads corresponds to each base, so variant calls can be made with a higher degree of confidence.

To assess coverage performance, libraries were prepared from each of the three TruSeq DNA kits. A TruSeq DNA PCR-Free library was prepared using 1 µg of Coriell Human-1 DNA (NA18507) following the 350 bp insert workflow. A TruSeq Nano DNA library was prepared using 100 ng of the same Coriell Human-1 DNA following the 350 bp insert workflow. A TruSeq DNA library was prepared using 1 µg of Coriell Human-1 DNA. Each sample was analyzed using the Isaac™ aligner and variant caller from four lanes of HiSeq® 2500 rapid-run data. Coverage depth averaged 39× for all samples, and coverage data can be accessed through the BaseSpace® analysis environment[1].

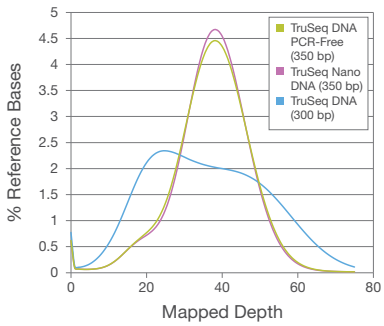
For challenging genomic content, including "difficult" promoters and GC-rich regions, TruSeq Nano DNA and TruSeq DNA PCR-Free libraries show more than 150% improvement in coverage compared to the original TruSeq DNA method (Figure 1). "Difficult" promoters denote the set of 100 promoter regions that are insufficiently covered, which are empirically defined by the Broad Institute of MIT and Harvard[2]. G-Rich regions are defined as 30 bases with ≥ 80% G. High GC regions are defined as 100 bases with ≥ 75% GC content. Huge GC regions denote 100 bases with ≥ 85% GC content. TruSeq Nano DNA and TruSeq DNA PCR-Free libraries also demonstrate approximately 10% performance improvement over exons, genes, and regions with repeated AT dinucleotides when compared to TruSeq DNA libraries. Known human coding and non-coding exons and genes are defined in the RefSeq Genes track in the UCSC Genome Browser[3]. AT dinucleotides indicate 30 bases of repeated AT dinucleotides. This increase in coverage provides a more comprehensive view of the genome.

Figure 1: Improved Coverage of Challenging Regions



TruSeq DNA PCR-Free and TruSeq Nano DNA libraries demonstrate improved coverage of challenging genomic content.

Figure 2: Greater Coverage Uniformity



TruSeq Nano DNA and TruSeq DNA PCR-Free libraries provide greater coverage uniformity across the genome when compared to those generated using the TruSeq DNA protocol.

Coverage Uniformity

Coverage uniformity is used to measure data comprehensiveness for a sequencing run. When a sequencing run achieves highly uniform coverage, the shape of the plot will resemble a Poisson-like distribution with a small standard deviation, indicating a narrower distribution of mapped depth. Mapped depth refers to the total number of bases sequenced and aligned at a given reference base position. This distribution is valid under the assumption that reads are randomly distributed across the genome and that the ability to detect true overlaps between reads is consistent within a sequencing run.

As shown in Figure 2, TruSeq Nano DNA and TruSeq DNA PCR-Free libraries demonstrate greater coverage uniformity across the genome when compared to TruSeq DNA libraries. Figure 2 demonstrates that

reads consistently achieve an average mapped depth of approximately 40×. The high coverage uniformity indicates that these two runs enable more accurate calling of variants that are distant from the mean depth. Highly uniform coverage also signifies that a sequencing run achieved a high percentage of bases at ≥ 30× coverage, while a wider distribution would require additional sequencing.

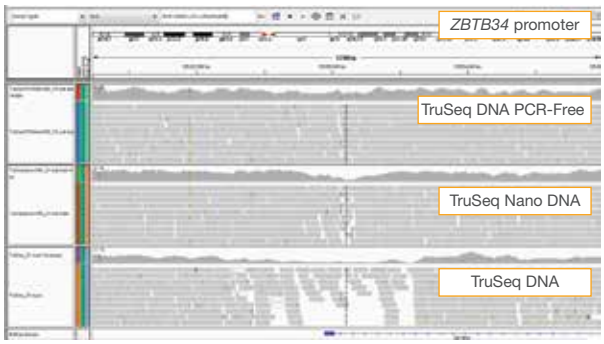
Reduced Coverage Gaps

The percentage of bases at low read depth quantifies the number of gaps in a data set. A gap exists if a base or several consecutive bases are either not sequenced at all or sequenced with a read depth below the specified cutoff. In such instances, there will be few or no reads to support variant calls (SNPs or indels) in the gap region, resulting in an increased rate of missed calls. Deep coverage of these traditionally difficult-to-sequence regions enables researchers to have more confidence in the variant calls made. Increased coverage can be especially significant when analyzing somatic mutations in the context of certain phenotypic traits or diseases.

Figure 3: Reduced Gaps in Coverage

A

B



The increased genomic coverage provided by TruSeq Nano DNA and TruSeq DNA PCR-Free libraries results in fewer coverage gaps. The examples shown here demonstrate enhanced coverage of the GC-rich coding regions of the *RNPEPL1* promoter (A) and the *ZBTB34* promoter (B). These data were generated using the Integrative Genomics Viewer developed by the Broad Institute[4].

Both TruSeq DNA PCR-Free and TruSeq Nano DNA libraries show an approximate 70% decrease in the number of gaps compared to libraries created using the original TruSeq DNA method. This drastic reduction in coverage gaps stems from the reduced bias provided by TruSeq Nano DNA kits and the elimination of PCR-induced bias by the PCR-Free kits. TruSeq DNA PCR-Free and Nano DNA libraries also demonstrate almost 50% reduction in the total size of gaps. Examples of decreased gaps in coverage are shown in Figure 3, using the human *RNPEPL1* (arginyl aminopeptidase-like 1) and the *ZBTB34* (zinc finger and BTB domain-containing 34) promoters.

Diversity

Diversity refers to the number of unique fragments generated in a sequencing run. High diversity and few duplicate reads provide greater coverage of the genome, increasing the confidence of variant calls as each base is covered by a greater number of unique fragments.

All three TruSeq DNA kits demonstrate similar depths of coverage and produce low percentages of duplicates (Table 1). Diversity levels, reported in billions of fragments, are dependent on the complexity of a given genome. The data in Table 1 demonstrate that the reduced input requirement offered by TruSeq Nano DNA kits and the elimination of PCR by TruSeq DNA PCR-Free kits continue to provide high-quality data. Coverage depth and diversity are consistent across all three kits, and the percentages of duplicates remain low (below 2.5%). The diversity levels provided by all TruSeq DNA kits (a minimum of 2 billion) are ideal for sequencing a human genome at ≥ 30× coverage without increasing the number of duplicates.

Table 1: TruSeq Kits Provide High Diversity Levels

Sample	Coverage Depth	Duplicates	Diversity (billions)
TruSeq DNA Kit (300 bp)	38.64	1.88%	13.22
TruSeq DNA PCR-Free Kit (350 bp)	39.76	1.49%	17.19
TruSeq Nano DNA Kit (350 bp)	39.80	2.43%	10.52

All TruSeq DNA kits generate enough unique molecules to sequence a human genome at ≥ 30× coverage. These data were generated from a 2 × 101 bp sequencing run using four lanes of a HiSeq 2500 instrument in rapid-run mode.

Table 2: Comparison of TruSeq DNA Library Preparation Kits

Specification	TruSeq Nano DNA	TruSeq DNA PCR-Free	TruSeq DNA
Input quantity	100–200 ng	1–2 µg	1 µg
Includes PCR	Yes	No	Yes
Assay time	~6 hours	~5 hours	1–2 days
Hands-on time	~5 hours	~4 hours	~8 hours
Target insert size	350 bp or 550 bp	350 bp or 550 bp	300 bp
Gel-free protocol	Yes	Yes	No
Number of samples supported	24 (LT) or 96 (HT) samples	24 (LT) or 96 (HT) samples	48 (LT) or 96 (HT) samples
Supports enrichment	No*	No*	Yes
Size-selection beads	Included	Included	Not included
Catalog number	FC-121-4001 (LT, Set A)	FC-121-3001 (LT, Set A)	FC-121-2001 (LT, Set A)**
	FC-121-4002 (LT, Set B)	FC-121-3002 (LT, Set B)	FC-121-2002 (LT, Set B)**
	FC-121-4003 (HT)	FC-121-3003 (HT)	FC-121-2003 (HT)**

*Nextera Rapid Capture products support a variety of enrichment applications. For more information, visit www.illumina.com/NRC.

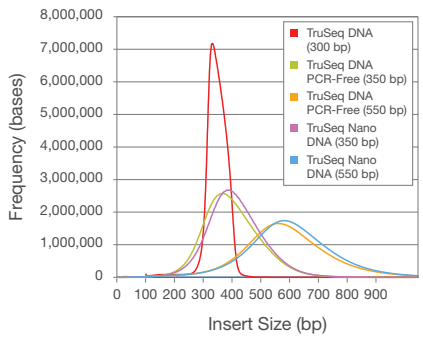
**TruSeq DNA kits will be discontinued after December 2013.

Size Selection

The TruSeq Nano DNA and TruSeq DNA PCR-Free workflows feature automation-friendly bead-based size selection. Although bead-based selection provides a broader size distribution than the traditional gel-based selection (Figure 4), it maintains high data quality while reducing hands-on time. In addition to accelerating the workflow, simple bead-based size selection avoids typical sample loss associated with gel selection. TruSeq Nano DNA and TruSeq DNA PCR-Free workflows offer two insert sizes, 350 bp and 550 bp, to support various sequencing applications.

The TruSeq portfolio of library preparation kits provides a flexible set of solutions that can be tailored to meet the needs of each sequencing project (Table 2). Compatible with a range of study designs, all TruSeq kits support high- and low-throughput studies while offering unique benefits depending on your research needs.

Figure 4: Broad Size Distribution Maintains High Data Quality



The TruSeq Nano DNA and TruSeq DNA PCR-Free workflows feature bead-based size selection, which yields broader size distributions while maintaining high data quality.

Summary

The TruSeq DNA portfolio provides inclusive and reliable library preparation solutions, ideal for virtually any sequencing application. The innovative TruSeq DNA PCR-Free and TruSeq Nano DNA workflows reduce bias, improving genomic coverage and uniformity. Learn more about the newest additions to the TruSeq DNA portfolio at www.illumina.com/nano and www.illumina.com/pcrfree.

References

1. BaseSpace data sets are available for the following libraries: TruSeq DNA, TruSeq DNA PCR-Free (350 bp), TruSeq DNA PCR-Free (550 bp), TruSeq Nano DNA (350 bp), and TruSeq Nano DNA (550 bp) libraries
2. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. (2013) Characterizing and measuring bias in sequence data. Genome Biol 14: R51.
3. genome.ucsc.edu
4. www.broadinstitute.org/igv

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

FOR RESEARCH USE ONLY

Pub. No. 770-2013-029 Current as of 20 November 2014

illumina®