illumina®

High-performance whole-genome sequencing with Illumina DNA PCR-Free Prep, Tagmentation

Fast and flexible library preparation that provides exceptional variant calling performance.

Introduction

Whole-genome sequencing (WGS) is the most comprehensive and unbiased method for interrogating the 3.2 billion bases of the human genome.[1,2] The rapid drop in sequencing costs and the ability of WGS to produce large volumes of data quickly make it a powerful tool for use in many human genomics research applications. When choosing a library preparation solution for WGS, it is important to consider the performance, cost, and scalability of the workflow. Illumina DNA PCR-Free Prep, Tagmentation is an advanced solution that uses On-Bead Tagmentation as part of an integrated, scalable, and rapid WGS workflow (Figure 1).

PCR-free library preparation is recognized as the gold standard for WGS, showing more uniform coverage and improved performance in calling many different variant types, as compared to PCR-based methods.[3] This is because PCR bias can lead to uneven coverage across difficult regions of the genome, such as promoter regions with extreme base composition or repetitive regions (Figure 2). However, in the past PCR-free library preparation represented a burdensome and time-consuming assay. Illumina DNA PCR-Free Prep, Tagmentation is fast and flexible, while still delivering the high-quality variant calling performance expected from PCR-free assays.

The Illumina DNA PCR-Free Prep workflow

Illumina DNA PCR-Free Prep uses On-Bead Tagmentation to enable a simple, rapid library preparation workflow.[4] Tagmentation is an enzymatic reaction that combines adapter ligation (ie, “tagging”) and DNA fragmentation in a single, rapid step without the costly equipment and consumables associated with mechanical shearing. For DNA inputs ≥ 300 ng, On-Bead Tagmentation results in saturation of the bead-bound transposomes, enabling self-normalization and library pooling using equal volumes of each library.[5] The pool is then quantified and diluted for sequencing using a single Qubit quantification (~10 minutes), reducing the time and cost associated with qPCR normalization.

The workflow is rapid, with library prep and pooling of 16 samples completed in ~ 90 minutes, and is scalable, available in 24- and 96-sample kit configurations, with automation-friendly volumes. Due to the fact that there are few plate transfers and no need for manual intervention, mechanical shearing, or qPCR quantitation, there is potential for an entirely walk-away automation solution. Moreover, Illumina DNA PCR-Free Prep, Tagmentation consistently provides enough yield for two standard runs on NovaSeq™ S2 or S4 flow cells, or one standard and one DFL run on a NovaSeq S1 flow cell, when using the saturating workflow. This option to requeue sequencing runs provides additional flexibility and robustness to laboratory testing processes. This application note demonstrates the exceptional performance of Illumina DNA PCR-Free Prep, Tagmentation as part of a comprehensive workflow for WGS applications.

Methods

Library preparation

Library preparation was performed using 300 ng human reference DNA (Coriell, Catalog no. NA12878) and the Illumina DNA PCR-Free Prep, Tagmentation workflow. Resulting libraries were pooled by volume, and loaded onto the NovaSeq 6000 System for sequencing.

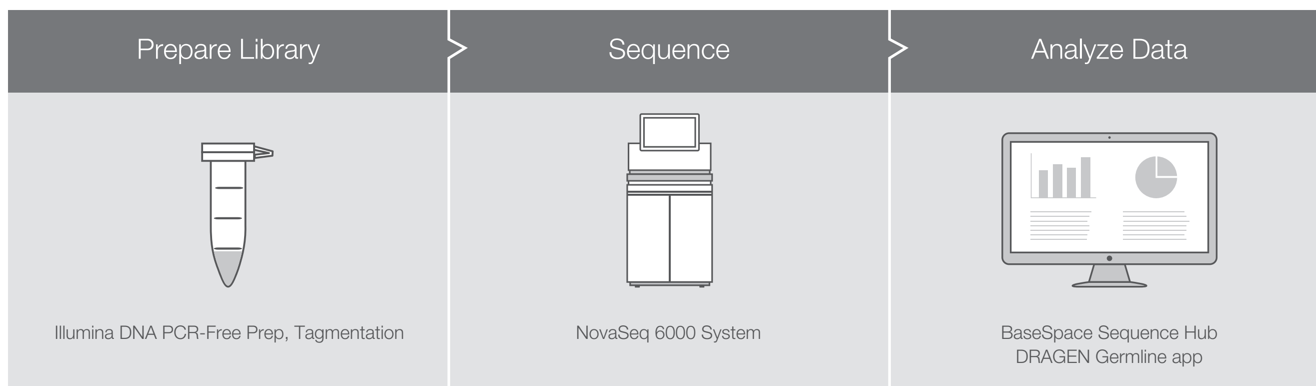

Figure 1: Illumina DNA PCR-Free Prep workflow—Illumina DNA PCR-Free Prep, Tagmentation may be used as part of a comprehensive, streamlined workflow for human WGS that includes library preparation, sequencing, and data analysis.

For Research Use Only. Not for use in diagnostic procedures.

Sequencing

Prepared library pools were loaded onto an S4 flow cell on the NovaSeq 6000 System at plexities shown (Table 1) and achieved ≥ 40× coverage for > 99% of samples (163/164) delivering mean run yields routinely above NovaSeq 6000 specifications. Higher plexities can be achieved by optimizing index pooling or targeting lower coverage.

Figure 2: Comparison of read coverage across GC-rich regions—Illumina DNA PCR-Free Prep provides superior read coverage across the GC-rich promoter region of the human *RNPEPL1* gene, as compared to TruSeq™ DNA PCR-Free and PCR-based library prep. Read maps were visualized with the Integrative Genomics Viewer (IGV) App, available in BaseSpace Sequence Hub.

Table 1: Coverage metrics

Metric	S4 flow cell	S2 flow cell	S1 flow cell
No. of sample libraries	16	6	3
No. of runs	8	4	4
Mean yield per sample[a]	200.7 Gbp	199.1 Gbp	182.2 Gbp
Mean no. of reads per sample	669.2M	663.6M	607.3M
Mean yield per run[a]	3643 Gbp	1333 Gbp	606 Gbp
Mean autosomal coverage	45.19	46.26	46.76
Unique dual index pairs (UDP, plate A)	UDP0037-44 + UDP0065-72 UDP0081-96	UDP0081-86 UDP0089-94	UDP0089-91 UDP0081-83
Index CV, Median % (range)	13.9% (8.4-26)	5.0% (3.3-7.0)	3.7% (1.3-15.3)

a. Mean run and sample yields were above NovaSeq 6000 specifications.

Data analysis

After sequencing was complete, data was streamed directly from the instrument into the cloud ecosystem for push-button analysis using DRAGEN™ apps available through BaseSpace™ Sequence Hub. The DRAGEN Germline app v3.5.7 was used for alignment against the hg38 reference genome assembly. Evaluation of single nucleotide variant (SNV) and insertion/deletion (indel) calling was done against the Platinum Genomes v2017.1 truthset using the Variant Calling Assessment Tool v4.0.1 BaseSpace app. Evaluation of structural variant (SV) calling (indels in the 50 bp–10 Kbp size range) was performed based on comparison of variant calls made in each sample against an internal truthset generated using long-read technology for that cell line using the witty.er tool v0.3.2.

Results

To evaluate the performance of Illumina DNA PCR-Free Prep, Tagmentation for WGS, results were compared to data generated from llibraries using a PCR-based library preparation kit.

High-quality variant calling

Improvements in performance compared to the PCR-based assay were observed in nearly every metric evaluated in this study (Table 2). These include precision and recall for both SNVs and indels, as well as an evaluation of the fraction of genomic positions where a passing genotype call can be made (callability).

Uniform genomic coverage

The evaluation of CNV performance was limited in this study. There are few events observed in the NA12878 cell line (~60 CNVs), and there is a lack of a complete and reliable truthset for evaluation of variant calls. In that limited evaluation, no significant difference in performance was observed between Illumina DNA PCR-Free Prep and the PCR-based library prep. However, significantly more even coverage, often a good predictor of CNV calling performance, was achieved with Illumina DNA PCR-Free Prep, as compared to the PCR-based assay (Figure 3).

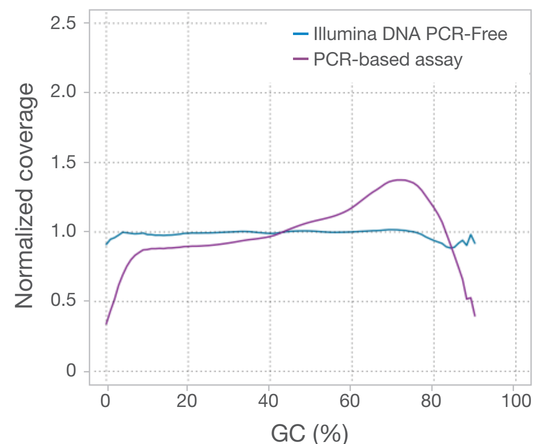

Figure 3: Illumina DNA PCR-Free coverage uniformity—Illumina DNA PCR-Free provides improved uniformity of coverage across a range of GC content in the human genome. Normalized coverage data was calculated using the DRAGEN Germline app v3.5.7.

Highly accurate STR detection

Comprehensive detection of GC-rich STR expansions is a known limitation of PCR-based assays, with nearly all reads from those genomic segments being absent from such data sets. The performance of Illumina DNA PCR-Free Prep in the detection of

For Research Use Only. Not for use in diagnostic procedures.

Table 2: Illumina DNA PCR-Free Prep variant calling performance

	Illumina DNA PCR-Free (mean, n=20)	PCR-based prep (mean, n=4)	Difference (PCR-Free minus PCR)
Autosome callability	97.54%	97.32%	**0.21%**
Noncallable bases in autosomes (bp)	67,564,679	73,463,941	**-5,899,261**
Autosomal coding exon callability	99.06%	98.914%	**0.16%**
Noncallable bases in autosomal coding exons (bp)	571,201	666,716	**-95,514**
SNV recall	99.37%	99.31%	**0.06%**
False negative SNVs	22,753	24,835	**-2081**
SNV precision	99.82%	99.82%	0.00%
False positive SNVs	6468	6469	-1
Indel recall	98.74%	97.96%	**0.77%**
False negative indels	7242	11,671	**-4429**
Indel precision	97.99%	97.75%	**0.24%**
False positive indels	11,623	12,932	**-1308**
Deletion (50 bp–10 Kbp) Recall	40.40%	39.29%	**1.1%**
Deletion (50 bp–10 Kbp) Precision	85.82%	86.67%	-0.8%
Insertion (50 bp–10 Kbp) Recall	35.13%	31.67%	**3.46%**
Insertion (50 bp–10 Kbp) Precision	94.35%	94.42%	-0.07%

The difference in variant calling is color-coded such that improved performance with Illumina DNA PCR-Free is highlighted in green.

Table 3: Illumina DNA PCR-Free Prep STR expansion accuracy

Cell line	Locus with STR expansion	STR motif	Disease association	Expansion threshold (repeat units)	Replicates with expansion detected
ND12161	*C9ORF72*	CCCCGG	Amyotrophic lateral sclerosis 1 (ALS1)	≥ 31	2/2
ND11917	*C9ORF72*	CCCCGG		≥ 31	2/2
NA23709	*AR*	CAG	Spinal and bulbar muscular atrophy, X-linked (SMAX1)	≥ 36	4/4
NA13716	*ATN1*	CAG	Spinocerebellar ataxia type 1	≥ 36	2/2
NA13717	*ATN1*	CAG		≥ 36	2/2
NA06151	*ATXN3*	CAG	Machado-Joseph disease (MJD)	≥ 45	4/4
NA04567	*DMPK*	CAG	Myotonic dystrophy (DM1)	≥ 35	2/2
NA04648	*DMPK*	CAG		≥ 35	2/2
NA04025	*FMR1*	CGG	Fragile X Syndrome (FXS)	≥ 55	2/2
NA07862	*FMR1*	CGG		≥ 55	2/2
NA03816	*FXN*	GAA	Friedreich ataxia 1 (FRDA)	≥ 66	2/2
NA15850	*FXN*	GAA		≥ 66	2/2
NA13507	*HTT*	CAG	Huntington's disease (HD)	≥ 36	2/2
NA13509	*HTT*	CAG		≥ 36	2/2
NA13536	*ATXN1*	CAG	Spinocerebellar ataxia 1 (SCA1 ATAXIN1, ATX1)	≥ 36	2/2
NA13537	*ATXN1*	CAG		≥ 36	2/2

STR expansions was evaluated in a study of 16 different cell lines containing 9 different pathogenic STR expansions prepared in duplicate or quadruplicate. Libraries were sequenced across eight NovaSeq S4 runs. Illumina DNA PCR-Free Prep resulted in 100% sensitivity for expansion detection (Table 3). This study included two cell lines with expansions in the C9ORF72 locus (ND12161 and ND11917) and two cell lines with expansions in the FMR1 locus (NA04025 and NA07862). These are long GC expansions that were inaccessible to PCR-based methods; however, they were accurately detected using Illumina DNA PCR-Free Prep (Table 3). This study included 72 other Illumina DNA PCR-Free libraries from cell lines and blood samples originating from 19 different individuals with no previous evidence of pathogenic STR expansions in loci targeted by the DRAGEN Germline pipeline. No false positive expansions (100% specificity) were detected in those 72 samples (data not shown).

Summary

Illumina DNA PCR-Free Prep, Tagmentation is an advanced library preparation solution that uses On-Bead Tagmentation as part of an integrated, scalable, and rapid WGS workflow. Library preparation is compatible with options for complete automation of the protocol. For regulated labs or those running large-scale operations, BaseSpace Clarity LIMS can be integrated to help track samples and manage workflows to optimize lab operations for efficiency gains. Illumina DNA PCR-Free Prep libraries result in high-quality sequencing data, enabling uniform coverage, and superior performance in variant calling.

Learn more

To learn more about Illumina DNA PCR-Free Prep, Tagmentation, visit www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html

References

1. Wang K, Yuen ST, Xu J, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet. 2014;46(6):573–582.
2. Zahir FR, Mwenifumbo JC, Chun HE, et al. Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. BMC Genomics. 2017;18(1):403.
3. Fungtammasan A, Ananda G, Hile SE, et al. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res*. 2015;25(5):736-749.
4. Illumina (2018). Nextera DNA Flex Library Preparation Kit. Accessed April 10, 2020.
5. Bruinsma S, Burgess J, Schlingman D, Czyz A, Morrell N, et al. Bead-linked transposomes enable a normalization-free workflow for NGS library preparation. *BMC Genomics*. 2018;19(1):722.

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

© 2020 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html. 770-2020-006-A QB10058

illumina®