Optimized protocol for human whole-exome sequencing

Introduction

Since the first human genome was sequenced in 2003, new sequencing technologies have been evolving rapidly. Even though genome sequencing costs have decreased drastically over the years due to the emergence of next-generation sequencing (NGS), it has yet to become affordable and efficient enough for routine use in diagnostics research. In contrast to whole-genome sequencing (WGS), whole-exome sequencing (WES) and targeted sequencing provide a balance between cost and benefit. WES targets approximately 2% of the whole genome, corresponding to protein-coding genes [1]. It is a proven and cost-effective method to detect disease-causing variants and discover gene targets. This application note describes a workflow solution for NGS library preparation of human whole-exome samples using the Invitrogen™ Collibri™ PS DNA Library Prep Kit for Illumina™ Systems with unique dual indexes (UDIs), from library preparation to sequencing (Figure 1).

Methods

Samples

Well-characterized human genomic DNA (gDNA) samples obtained from the GM12878 human cell line (Ceph/Utah pedigree) were purchased from the Coriell Institute for Medical Research (Cat. No. NA12878) for use in whole-exome library preparation.

Library generation

The gDNA samples were converted into sequencing-ready libraries using physical shearing (PS) for fragmentation using the Collibri PS DNA Library Prep Kit for Illumina Systems with UDIs. Fragmentation was performed on the Covaris™ E220 Focused-ultrasonicator using the following parameters: 10% duty factor, 175 peak incident power, 200 cycles per burst, 280 sec treatment time, and 2–8°C bath temperature. The sample volume for fragmentation was 50 µL; 25 µL of the fragmented DNA sample was used for library generation using the standard protocol for the Collibri PS DNA Library Prep Kit for Illumina Systems.

Figure 1. NGS workflow for human whole-exome sequencing.

Table 1. Composition of hybridization kits (Long and Fast hybridization protocols).

Fast hybridization protocol		
Name	**Cat. No.**	**Kit composition**
Twist Fast Hybridization and Wash Kit (2 boxes), 96 reactions	101175	**1. Twist fast hybridization reagents:**
		Fast Hybridization Mix
		Hybridization Enhancer
		Amplification Primers
		2. Twist fast wash buffers:
		Fast Binding Buffer
		Fast Wash Buffer 1
		Wash Buffer 2
Twist Comprehensive Exome, 96 reactions	102033 (1.1 rev)/ 101953 (1 rev)	Comprehensive Panel
Twist Comprehensive Exome, 12 reactions	102032 (1.1rev)/ 101952 (1 rev)	Comprehensive Panel
Twist Comprehensive Exome, 2 reactions	102031 (1.1 rev)/ 101951 (1 rev)	Comprehensive Panel
Twist Binding and Purification Beads (96 reactions/12 reactions)	100984/100983	Streptavidin binding beads
		DNA purification Beads
Twist Universal Blockers (96 reactions/12 reactions)	100767/100578	Universal Blockers
		Blocker Solution

Long hybridization protocol		
Name	**Cat. No.**	**Kit composition**
Twist Hybridization and Wash Kit (2 boxes), 96 reactions	101026	**1. Twist hybridization reagents:**
		Hybridization mix
		Hybridization enhancer
		Amplification primers
		2. Twist wash buffers:
		Binding buffer
		Wash buffer 1
		Wash buffer 2
Twist Comprehensive Exome, 96 reactions	102033 (1.1 rev)/ 101953 (1 rev)	Comprehensive panel
Twist Comprehensive Exome, 12 reactions	102032 (1.1 rev)/ 101952 (1 rev)	Comprehensive panel
Twist Comprehensive Exome, 2 reactions	102031 (1.1 rev)/ 101951 (1 rev)	Comprehensive panel
Twist Binding and Purification Beads (96 reactions/12 reactions)	100984/100983	Streptavidin binding beads
		DNA purification beads
Twist Universal Blockers (96 reactions/12 reactions)	100767/100578	Universal blockers
		Blocker solution

The adaptor-ligated library was purified according to a customized cleanup protocol (**Pub. No. MAN0025491)** for the Collibri PS DNA Library Prep Kit for Illumina Systems. The library was then amplified for 8 PCR cycles before the enrichment step and subsequent cleanup (**Pub. No. MAN00025492**). The complete manual for the human whole-exome library preparation protocol described in this application note is also available **(Pub. No. MAN0025533).**

Enrichment

The WGS libraries were quantified using the Invitrogen™ Qubit™ Fluorometer and pooled. The exome targets were then enriched using a synthetic probe panel (Twist Human Core Exome Panel from Twist Bioscience). Two separate hybridization reactions were performed:

- **"Fast" hybridization**—using the Twist Fast Hybridization and Wash Kit; 2 hr target enrichment
- **"Long" hybridization**—using the Twist Standard Hybridization and Wash Kit; overnight (~16 hr) target enrichment

The composition of each hybridization kit is noted in Table 1. The enriched libraries were amplified for 8 cycles using the 2X Library Amplification Master Mix from the Collibri PS DNA Library Prep Kit for Illumina Systems instead of the recommended KAPA™ HiFi HotStart ReadyMix from Roche (Table 2). The DNA purification beads from Twist hybridization kits can be replaced with Invitrogen™ Dynabeads™ MyOne™ Streptavidin T1 magnetic beads (Cat. No. 65601), if required.

Library quantification and sequencing

The final concentrations of the sequencing libraries were determined using the Invitrogen™ Collibri™ Library Quantification Kit. Libraries were sequenced on the Illumina™ NovaSeq™ 6000 Sequencing System by paired-end sequencing of 2 x 150 bp.

NGS analysis

The Illumina™ BaseSpace™ Enrichment App v3.1.0 was used to align a subsample of 30M passing filter (PF) for clusters (60M reads) from each sample to the hg19 reference genome. The enrichment parameters were calculated using the Picard CollectHsMetrics tool. Variant calling in the Variant Calling Assessment Tool v4.0.2 was limited by the target regions. Gold-standard data for NA12878 from the Genome in a Bottle (GIAB) consortium hosted by NIST were used to determine the precision and recall parameters.

Table 2. Recommended protocol optimization for library generation using Collibri PS DNA Library Prep Kits.

Step	Standard recommendation	Recommended changes for WES samples	
1. DNA fragmentation	Follow the manufacturer's recommendations to obtain fragmented DNA of desired length and concentration	This protocol has been optimized using a Covaris E220 Focused-ultrasonicator and the 50 µL Covaris microTUBE Strip for a target DNA fragment size of 150–200 bp	
		Setting	**High-quality DNA**
		Duty factor	10%
		Peak incident power (PIP)	175
		Cycles per burst	200
		Treatment time	280 seconds
		Bath temperature	2–8°C
	Input: 1–1,000 ng	Input: 50 ng	
	Target insert size: 350 bp or 550 bp	Target insert size: 150–200 bp	
2. End repair and dA-tailing	Described in the "End conversion" section in the Collibri PS DNA Library Prep Kit for Illumina User Guide	Described in the "End conversion" section in the Collibri PS DNA Library Prep Kit for Illumina User Guide (Pub. No. MAN0018546)	
3. Adapter ligation	Described in the "Dual-Indexed Adaptor ligation" section in the Collibri PS DNA Library Prep Kit for Illumina User Guide	Described in the "Dual-Indexed Adaptor ligation" section in the Collibri PS DNA Library Prep Kit for Illumina User Guide (Pub. No. MAN0018546)	
3. Post-ligation cleanup or double-sided size selection	Cleanup or double-sided size selection to match target insert size	Check product page for the protocol (Pub. No. MAN0025491)	
4. PCR amplification of the DNA library	The number of PCR cycles depends on the starting amount of input DNA	Amplify the library for 10 PCR cycles	
5. Amplified DNA library purification	Post-amplification cleanup of the DNA library	Check product page for the protocol (Pub. No. MAN00025492)	
6. Human exome enrichment	Not applicable	Target enrichment protocol	
7. Post-capture library amplification with Library Amplification Master Mix	Not applicable	**Component**	**Volume**
		Enriched library in the Target Enrichment and Amplification Beads slurry (brown)	20 µL
		2X Collibri Library Amplification Master Mix (blue)	25 µL
		Collibri Primer Mix	5 µL
		Total volume (brown mixture)	50 µL
		Amplify the enriched DNA library for 8 PCR cycles	
8. Purification of the amplified post-capture libraries	Not applicable	Purification of the amplified post-capture libraries	

Results

Some of the main parameters that are usually calculated in WES include enrichment efficiency, coverage of enriched targets, and single-nucleotide variant (SNV) analysis. In this study, we were able to compare the parameters of the Long and Fast enrichment protocols so that customers may identify and choose the performance they desire for their experiments.

Single-nucleotide variants

Accurate SNV determination is one of the key parameters in WES. Using the modified Collibri PS DNA library prep protocol described in Table 2 in combination with the Twist Bioscience enrichment kits (for Long and Fast hybridization protocols, Table 1), we were able to reach an SNV precision greater than 99%. An SNV recall of 96.4% was achieved using the Fast hybridization protocol, whereas the Long enrichment protocol yielded an SNV recall of 98.3% (Figure 2).

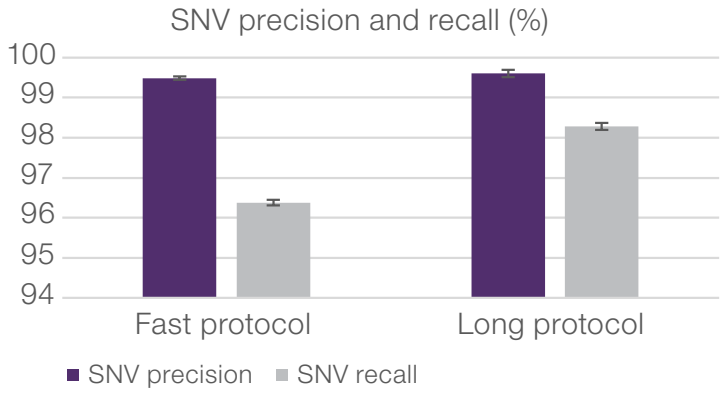


Figure 2. SNV precision and recall comparison for the Fast and Long enrichment protocols.

Coverage and enrichment

Efficient target enrichment and high coverage are key to successful WES. In addition, high coverage uniformity is required for efficient sequencing, so that the targeted regions are covered as uniformly as possible. In this study, the enrichment and coverage parameters were calculated using the 150 bp padded target. High coverage uniformity was achieved using a modified Collibri PS DNA library prep protocol in combination with the Twist Bioscience enrichment module. The results showed 96.2% coverage uniformity when using the Fast hybridization protocol and 98.2% uniformity when target enrichment was performed using the Long hybridization protocol (Figure 3, Table 3).

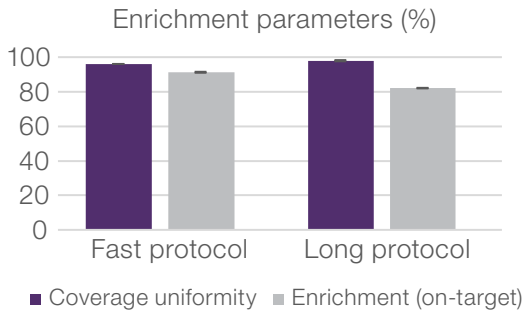
The overall target enrichment of 91.4% was achieved when using the Fast protocol, compared to the 82.2% target enrichment achieved with the Long protocol. We also observed that the percentage of targets with coverage greater than 20x is 98.3% when using the Long enrichment module but only 96.6% when the Fast hybridization module is used (Figure 3, Table 3).

Table 3. Comparison of whole-exome sequencing performance parameters for Fast and Long enrichment protocols. All of the parameters were calculated using 16 repetitions of the same sample preparation.

Parameter	PS Fast	PS Long
Panel	Twist Comprehensive Exome	
Panel size	36.8 Mb	
Probe size	120 bp	
Padded read enrichment (on-target)	91.4 ± 0.3%	82.2 ± 0.2%
Coverage at 20x	96.6 ± 0.0%	98.3 ± 0.3%
Uniformity of coverage	96.2 ± 0.0%	98.2 ± 0.4%
Read depth per sample	30M PF clusters	
SNV precision	99.5 ± 0.0%	99.6 ± 0.0%
SNV recall	96.4 ± 0.1%	98.3 ± 0.1%

Summary

Strong target coverage and sensitive variant detection for the human whole exome were accomplished using two different enrichment methods (Long and Fast) and a modified protocol for the Collibri ES DNA Library Prep Kit for Illumina Systems. Over 96% of exome bases were covered with only 20x coverage. Enrichment quality corresponded to the percentage of on-target reads, and all of the samples were sufficiently enriched to detect single-nucleotide polymorphisms (SNPs). The optimized protocol for the Collibri PS DNA Library Prep Kit for Illumina Systems, when used in combination with the Twist Bioscience hybridization kit, was extremely precise (>99%) and resulted in outstanding recall rates (>96%) for detecting SNPs.



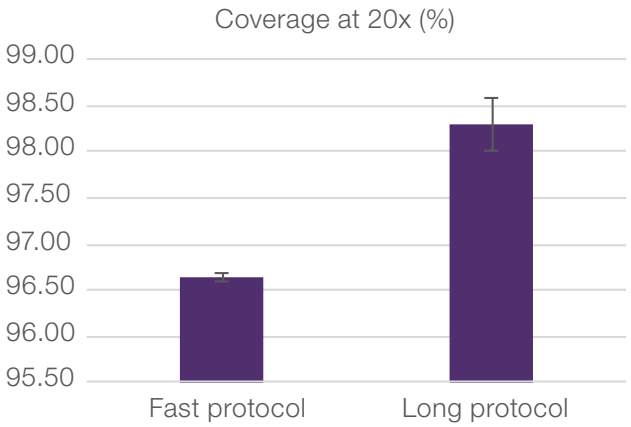


Figure 3. Coverage and enrichment parameters for the Fast and Long enrichment modules.

invitrogen

Conclusions

- The optimized method described in this application note for the Collibri PS DNA Library Prep Kit enables highly sensitive detection of variants for human whole-exome samples.
- Uniformity and high on-target rates show the efficiency of targeted sequencing using the optimized Collibri PS DNA Library Prep Kit.
- The Collibri Library Amplification Master Mix containing Invitrogen™ Platinum™ SuperFi™ DNA Polymerase provides uniform exome amplification and minimal error rates similar to those of PCR-free methods; this is crucial for detecting SNPs accurately.

Reference

1. Suwinski P, Ong C, Ling MHT et al. (2019) Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Front Genet* 10:49. doi: 10.3389/fgene.2019.00049

Ordering information

Description	Quantity	DNA index type	Cat. No.
Collibri PS DNA Library Prep Kit for Illumina Systems, with indexes	24 preps	CDI*	A38612024
	96 preps	CDI	A38614096
	24 preps	UDI* set A (1–24)	A38613024
		UDI set B (25–48)	A43611024
		UDI set C (49–72)	A43612024
		UDI set D (73–96)	A43613024
	96 preps	UDI set A–D (1–96)	A38614196
	96 preps	—	A38614096W
Collibri Library Quantification Kit	100 reactions	—	A38524100
Collibri Library Quantification Kit	500 reactions	—	A38524500
Collibri Library Amplification Master Mix with Primer Mix**	50 reactions	—	A38540050
Collibri Library Amplification Master Mix with Primer Mix**	250 reactions	—	A38540250

* CDI = combinatorial dual indexes, UDI = unique dual indexes.

** Collibri Library Amplification Master Mix with Primer Mix can be purchased separately for use with the PCR-free version of the Collibri DNA Library Prep Kit.

Find out more at **thermofisher.com/collibridna**

ThermoFisher SCIENTIFIC