Protocol

# Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing

Matthias Meyer[1] and Martin Kircher

*Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany*

*[Supplemental Material is available online at www.cshprotocols.org/supplemental/.]*

## INTRODUCTION

The large amount of DNA sequence data generated by high-throughput sequencing technologies often allows multiple samples to be sequenced in parallel on a single sequencing run. This is particularly true if subsets of the genome are studied rather than complete genomes. In recent years, target capture from sequencing libraries has largely replaced polymerase chain reaction (PCR) as the preferred method of target enrichment. Parallelizing target capture and sequencing for multiple samples requires the incorporation of sample-specific barcodes into sequencing libraries, which is necessary to trace back the sample source of each sequence. This protocol describes a fast and reliable method for the preparation of barcoded ("indexed") sequencing libraries for Illumina's Genome Analyzer platform. The protocol avoids expensive commercial library preparation kits and can be performed in a 96-well plate setup using multi-channel pipettes, requiring not more than two or three days of lab work. Libraries can be prepared from any type of double-stranded DNA, even if present in subnanogram quantity.

## RELATED INFORMATION

Illumina's "indexing" system differs from other sample barcoding methods for high-throughput sequencing in that the barcodes ("indexes") are placed within one of the adapters rather than being directly attached to the ends of template molecules (e.g., Craig et al. 2008; Meyer et al. 2008b). The barcode sequence is identified in a separate short sequencing read. This setup allows for a high degree of flexibility in experimental design, because libraries are first prepared with universal adapters and different indexes can repeatedly be added by amplification with tailed primers just before target capture or sequencing. The library preparation protocol described here (see Fig. 1 for an overview) is based on the general principle of library preparation originally developed for 454 sequencing (Margulies et al. 2005). By exchanging adapter sequences, removing and shortening several reaction steps, and introducing an amplification scheme, the protocol has been redesigned for rapid preparation of Illumina multiplex sequencing libraries using a 96-well plate format. In the example shown in Figure 2, the protocol was used to simultaneously capture and sequence target regions from 50 human samples using microarrays (HA Burbano, E Hodges, RE Green, AW Briggs, J Krause, M Meyer, JM Good, T Maricic, PLF Johnson, Z Xuan, et al., in prep.).

## MATERIALS

**CAUTIONS AND RECIPES:** Please see Appendices for appropriate handling of materials marked with <**!**>, and recipes for reagents marked with <**R**>.

[1]Corresponding author (mmeyer@eva.mpg.de).
**Cite as: Cold Spring Harb Protoc; 2010; doi:10.1101/pdb.prot5448**

## Reagents

Agarose gel (2%) and reagents for agarose gel electrophoresis
AMPure XP 60 mL Kit (Agencourt-Beckman Coulter A63881)
ATP (100 mM) (Fermentas R0441)
*Bst* DNA polymerase, large fragment (supplied with 10X ThermoPol reaction buffer) (New England BioLabs M0275S)
DNA ladder (e.g., GeneRuler; Fermentas) (optional; see note before Step 6)

*For unknown reasons, ladders from New England BioLabs do not work for this purpose.*

dNTP mix (25 mM each) (Fermentas R1121)
<R>EBT buffer
Ethanol (70%, freshly prepared)
$H_2O$ (HPLC grade)
Illumina reagents for DNA sequencing (Illumina, Inc.)
Cluster generation kit (e.g., GD-103-4001 [Standard Cluster Generation Kit v4], PE-203-4001 [Paired-End Cluster Generation Kit v4])
Multiplexing sequencing primer kit (PE-400-1002 [Multiplexing Sequencing Primers and PhiX Control Kit v1])

*Alternatively, the following primers may be used for sequencing:*

*Read 1 Sequencing Primer: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′*

*Index Read Sequencing Primer: 5′-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3′*

*Read 2 Sequencing Primer: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′*

Sequencing kit (FC-104-4002 [36 Cycle Sequencing Kit v4])
MinElute PCR Purification Kit (QIAGEN) (optional)
<R>Oligo hybridization buffer (10X)
Oligonucleotides (Sigma-Aldrich) (see Table 1)
Phusion Hot Start High-Fidelity DNA Polymerase (New England BioLabs F-540L) (supplied with 5X Phusion HF buffer)
Positive control DNA (200- to 300-bp fragment, generated via PCR using unmodified primers and a polymerase with terminal transferase activity, e.g., *Taq* DNA polymerase) (200-500 ng)
Sample DNA

*This protocol works reliably with as little as 100 pg and up to 1 µg of double-stranded sample DNA (e.g., genomic DNA, long-range PCR products, or cDNA). The amount of starting material should be chosen so that the representation of target molecules in the final library is sufficient. The final yield of the library preparation process is ~10%-20%. Therefore, a library prepared from 1 ng of human genomic DNA (about 300 copies of the haploid genome), will contain 30 to 60 copies of the human genome.*

Standard for quantitative PCR (qPCR) (see Steps 21.i-21.ii)
SYBR Green qPCR master mix (e.g., DyNAmo Flash SYBR Green qPCR Kit; New England BioLabs)
Tango buffer (10X; Fermentas BY5)
T4 DNA ligase (5 U/µL; Fermentas EL0011) (supplied with 10X T4 DNA ligase buffer and 50% PEG-4000 solution)
T4 DNA polymerase (5 U/µL; Fermentas EP0062)
T4 polynucleotide kinase (10 U/µL; Fermentas EK0032)
<R>TET buffer
Tween 20

## Equipment

Centrifuge for 96-well plates
DNA shearing device (e.g., Bioruptor UCD-200 [Diagenode]; Covaris E210 [Covaris Inc]) (for high-molecular-weight DNA; see Step 3)

*The Bioruptor UCD-200 can process 12 samples in parallel. Among the many alternative systems that are available for this step, the Covaris E210 system may be preferable, because it is compatible with the 96-well plate format.*

Equipment for agarose gel electrophoresis

Equipment and reagents for target capture from sequencing libraries (optional)

*Several systems are available; see e.g., Hodges et al. (2009) for a target capture approach using Agilent microarrays.*

Ice
Multichannel pipettes
Multichannel reagent basins (e.g., Thermo Scientific 9510027)
PCR plates (96-well, 200-µL capacity) and strip caps
Real-time PCR cycler (e.g., Mx3005P QPCR System; Agilent Technologies-Stratagene)
Sequencing machine (Genome Analyzer II/IIx/IIe or HiSeq2000; Illumina)
Spectrophotometer for DNA quantification (e.g., NanoDrop; Thermo Scientific)
SPRIPlate 96R-Ring Super Magnet Plate (Agencourt-Beckman Coulter A32782)
Thermal cycler
Tubes (microcentrifuge, 0.5-mL)
Tubes (PCR)
Vortex mixers for tubes and 96-well plates

## METHOD

*The protocol can be interrupted after Steps 3, 12, 16, 19, 24, and 26 by freezing the DNA at –20°C. Up to 94 samples can be processed in parallel on a 96-well reaction plate; two wells should be reserved for a blank and a positive control. Seal each reaction plate with strip caps and centrifuge to 2000*g *in a plate centrifuge after setting up each reaction in order to collect the liquid in the bottom of the wells. This prevents cross-contamination while removing the caps.*

### Preparation of Adapter Mix

*This step produces sufficient adapter mix for 200 reactions. The adapter mix can be used repeatedly and stored at –20°C before and after usage.*

1. Assemble the following hybridization reactions in separate PCR tubes:

| Reagent | Volume (µL) | Final concentration in 100-µL reaction |
|---|---|---|
| Hybridization mix for adapter P5 (200 µM): | | |
| IS1_adapter_P5.F (500 µM) | 40 | 200 µM |
| IS3_adapter_P5+P7.R (500 µM) | 40 | 200 µM |
| Oligo hybridization buffer (10X) | 10 | 1X |
| $H_2O$ | 10 | |
| Hybridization mix for adapter P7 (200 µM): | | |
| IS2_adapter_P7.F (500 µM) | 40 | 200 µM |
| IS3_adapter_P5+P7.R (500 µM) | 40 | 200 µM |
| Oligo hybridization buffer (10X) | 10 | 1X |
| $H_2O$ | 10 | |

2. Mix and incubate the reactions in a thermal cycler for 10 sec at 95°C, followed by a ramp from 95°C to 12°C at a rate of 0.1°C/sec. Combine both reactions to obtain a ready-to-use adapter mix (100 µM each adapter).

### Fragmentation and Purification of Sample DNA

*This step in the method is not always required. Prior to library preparation, high-molecular-weight sample DNA must be sheared into fragments of suitable size for Illumina sequencing (<600 bp). If samples other than high-molecular-weight DNA are used (e.g., short PCR products, highly degraded DNA, or short double-stranded cDNA), fragmentation may not be necessary. Step 3 describes DNA shearing by sonication using the Bioruptor UCD-200.*

3. Shear the DNA as follows:

   i. Transfer the samples to 0.5-mL tubes, and add $H_2O$ to reach final volumes of 50 µL.

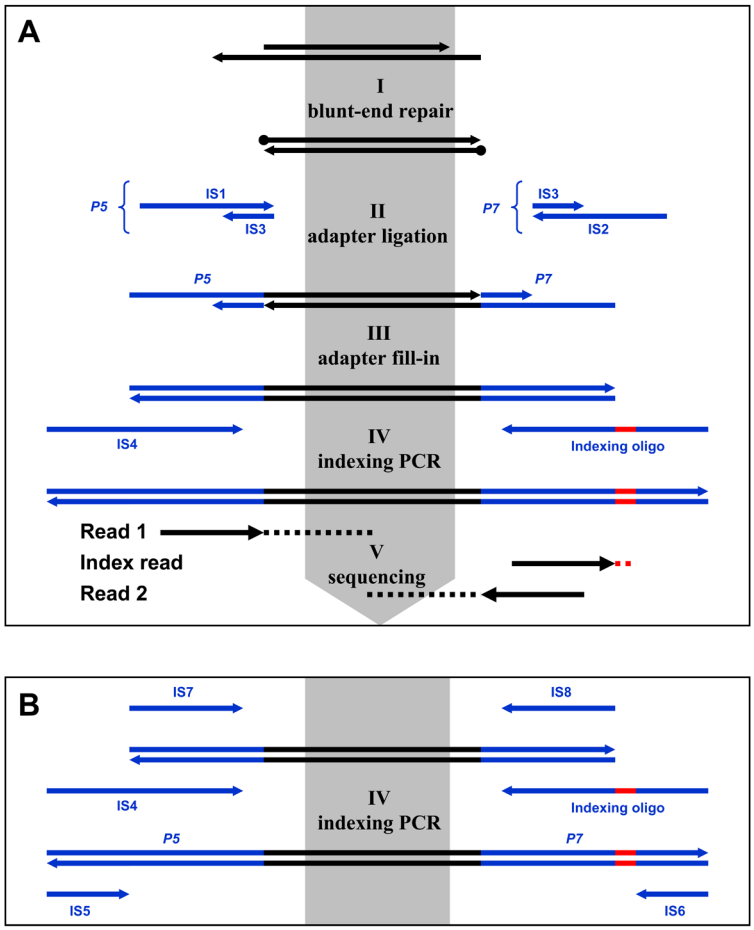


**FIGURE 1.** Schematic overview of the protocol and alternative amplification schemes. (*A*) Sample DNA is sheared into small fragments (not depicted). During blunt-end repair, overhanging 5′- and 3′-ends are filled in or removed by T4 DNA polymerase. 5′-phosphates are attached using T4 polynucleotide kinase (Steps 1-13). Two different adapters, P5 and P7, are ligated to both ends of the molecules using T4 DNA ligase (Steps 14-16). Ligation is nondirectional and also produces molecules which have the same adapters attached to both ends (*not depicted*). Such molecules do not interfere with sequencing and—due to the formation of hairpin structures—amplify very poorly during indexing PCR. Since the adapters do not carry 5′-phosphates, ligation joins only single strands. Nicks are removed in a fill-in reaction with Bst polymerase, which possesses strand-displacement activity (Steps 17-21). Indexes and full length adapter sequences are added by amplification with 5′-tailed primers (Steps 22-26). Indexed libraries are pooled in equimolar ratio. The pool is ready for target capture and/or sequencing on one of Illumina's sequencing platforms (Steps 27-28). Indexes are read in a separate sequencing read. Read 2, the paired end read, is optional. (*B*) Alternative amplification schemes can be used. Using the primers IS7 and IS8, libraries can be amplified prior to indexing. Using IS5 and IS6, single or pooled indexed libraries can be amplified, for example after target enrichment. (For color figure, see doi: 10.1101/pdb.prot5448 online at www.cshprotocols.org.)

ii. Expose the DNA four times to sonication cycles of 7 min, using the energy setting "HIGH" and an "ON/OFF interval" of 30 sec. If liquid spills to the tube walls, shake it down to the bottom of the wells after each sonication cycle.

*This produces a fragment size distribution between 100 bp and 400 bp, with a mean around 200 bp.*

iii. Transfer the sheared DNA samples to a 96-well PCR plate.

*The fragment size distribution obtained from sonication is well-suited for sequencing. However, if a very narrow fragment size distribution is desired, the fragmented DNA may be separated on an agarose gel and isolated from a gel slice to obtain a more narrow distribution. In the example given in* Figure 2*, no gel excision was performed.*

A

| | |
|---|---|
| total sequences (quality filtered) | 56,793,495 |
| sequences with correct index | 53,712,293 |
| sequences mapped to the human genome | 51,187,603 |
| sequences mapped to a target | 12,703,605 (24.8 %) |

B

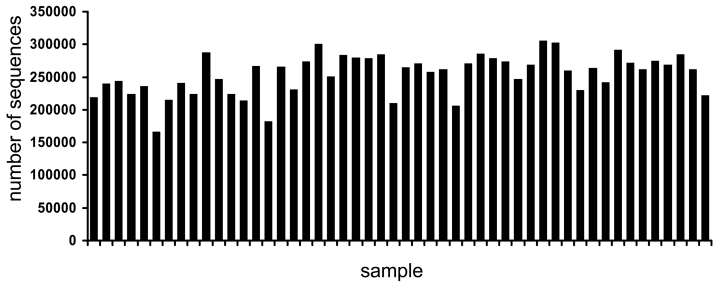


**FIGURE 2.** Example of a result from multiplex target capture and sequencing. Indexed libraries were prepared from 50 human samples from the CEPH human genome diversity panel as described in this protocol. Shearing was performed using the Bioruptor with no subsequent gel excision (see Step 3). The pool of libraries was loaded on a million-feature array from Agilent to capture 12,871 targets from the human genome with an average size of 232 bp (overall 2.9 million bp), following the protocol of Hodges et al. (2009). The array eluate was amplified for 12 cycles using primers IS5 and IS6 and sequenced on 5 lanes of the Illumina flow cell (2 × 100 cycles + 6 cycles index read). Shown are the results from mapping the sequences against the human genome (*A*) and the distribution of sequences among samples (*B*).

### Blunt-End Repair

*If the sample DNA is not dissolved in $H_2O$, Tris-Cl buffer (e.g., QIAGEN's Buffer EB), or TE buffer, purify the DNA as described in Steps 6-13 prior to beginning Step 4. If the sample volume exceeds 50 µL, purification can be used for concentrating the DNA. We strongly recommend carrying a positive and a blank control through Steps 4-18 of the protocol. As a positive control, 200-500 ng of a purified PCR product with a discrete size of 200-300 bp may be used. The product should be generated using unmodified PCR primers and a polymerase with terminal transferase activity (e.g., Taq DNA polymerase).*

4. Add a blank control (50 µL of $H_2O$) and a positive control to two empty wells of the reaction plate. Prepare a master mix as below for the required number of reactions. Mix carefully by flicking the tube with a finger. Avoid vortexing after addition of enzymes.

| Reagent | Volume (µL) per sample | Final concentration in 70-µL reaction |
|---|---|---|
| $H_2O$ | 7.12 | |
| Buffer Tango (10X) | 7 | 1X |
| dNTPs (25 mM each) | 0.28 | 100 µM each |
| ATP (100 mM) | 0.7 | 1 mM |
| T4 polynucleotide kinase (10 U/µL) | 3.5 | 0.5 U/ µL |
| T4 DNA polymerase (5 U/µL) | 1.4 | 0.1 U/ µL |

5. Using a multichannel pipette, add 20 µL of master mix to 50 µL of sample. Mix and incubate in a thermal cycler for 15 min at 25°C followed by 5 min at 12°C. Place plate on ice or immediately proceed to the next step.

### Reaction Clean-Up Using Solid Phase Reversible Immobilization (SPRI)

*Carboxyl-coated magnetic beads (SPRI beads) are ideally suited for reaction purification in a 96-well plate setup. However, under the conditions described here, SPRI purification does not retain molecules shorter than 100-150 bp. The exact size cutoff may vary among different batches of beads. If retention of short molecules is desired, the size cutoff can be adjusted by varying the volume of SPRI bead/buffer suspension added to the sample. The appropriate ratio of SPRI suspension to sample volume can be empirically determined using a DNA ladder (e.g., GeneRuler ladders). If retention of very short molecules is desired (30-80 bp), all SPRI purification steps should be replaced by spin column purification using the MinElute PCR Purification Kit.*

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.