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

© 2010 Cold Spring Harbor Laboratory Press

IDT Exhibit 1060.0001

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

IDT Exhibit 1060.0002

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 2000g 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.

IDT Exhibit 1060.0003

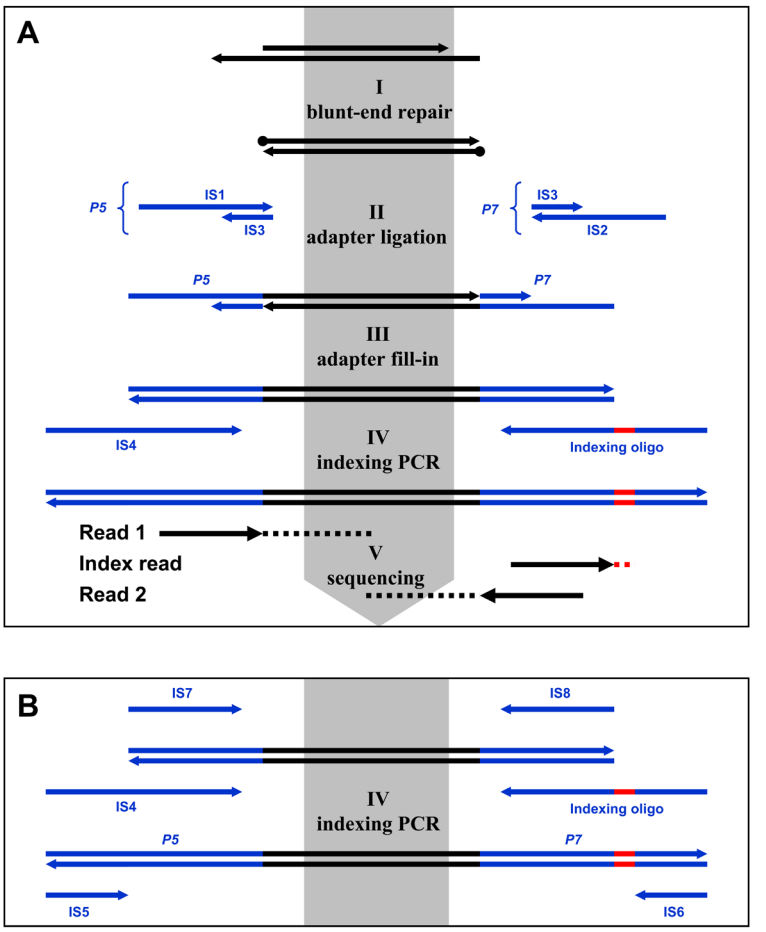

FIGURE 1. Schematic overview of the protocol and alternative amplification schemes. (*A*) Sample DNA is sheared into small fragments (not depicted). During blunt-end repair, overhanging 5′- and 3′-ends are filled in or removed by T4 DNA polymerase. 5′-phosphates are attached using T4 polynucleotide kinase (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.*

IDT Exhibit 1060.0004

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

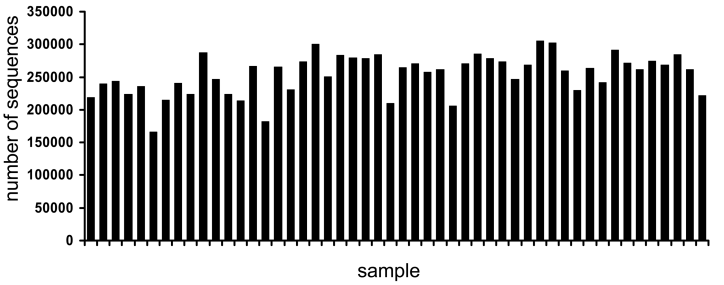

FIGURE 2. Example of a result from multiplex target capture and sequencing. Indexed libraries were prepared from 50 human samples from the CEPH human genome diversity panel as described in this protocol. Shearing was performed using the Bioruptor with no subsequent gel excision (see Step 3). The pool 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.

IDT Exhibit 1060.0005

Table 1. Oligonucleotides and sequences

Oligo ID	Sequence[a]
IS1_adapter.P5	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
IS2_adapter.P7	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*T*C*T
IS3_adapter.P5+P7	A*G*A*T*CGGAA*G*A*G*C
IS4_indPCR.P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT
IS5_reamp.P5	AATGATACGGCGACCACCGA
IS6_reamp.P7	CAAGCAGAAGACGGCATACGA
IS7_short_amp.P5	ACACTCTTTCCCTACACGAC
IS8_short_amp.P7	GTGACTGGAGTTCAGACGTGT
BO1.P5.F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-Pho
BO2.P5.R	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-Pho
BO3.P7.part1.F	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-Pho
BO4.P7.part1.R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Pho
BO5.P7.part2.F	ATCTCGTATGCCGTCTTCTGCTTG-Pho
BO6.P7.part2.R	CAAGCAGAAGACGGCATACGAGAT-Pho

[a]5′-3′; * indicates a PTO bond; Pho indicates a 3′-phosphate.
See Supplemental Material (Indexing_Oligo_Sequences.doc) for indexing oligo sequences.
All oligos (HPLC purified, 0.2 µmol synthesis scale) should be dissolved in TE or H_2O. Oligos 1-3 should be dissolved to 500 µM, oligos BO1-BO6 to 200 µM, and all other oligos to 10 µM. The indexing oligos should be transferred to a 96-well plate to allow for multichannel pipetting. HPLC purification can potentially introduce cross-contamination among indexing oligos. It is therefore important to (1) instruct the company to properly wash the HPLC column before loading a new oligo and (2) synthesize the oligos in a different order than listed here. This makes sure that cross-contamination induced during synthesis can be detected after sequencing by the appearance of index sequences that were not used in the experiment. When designing index sequences, the following criteria were taken into account: (1) Index sequences differ by at least three substitutions. This reduces the chance of converting one index into another by sequencing and amplification errors. (2) Indexes cannot be converted into one another by deleting the first base, which is the only insertion/deletion error common with Illumina sequencing. (3) Index sequences do not contain three or more identical bases in a row to ensure that they can be differentiated from artifact sequences. (4) Stretches of bases illuminated with the same laser (ACA, CAC, GTG, and TGT) are avoided. Software for designing alternative index sequences, for example, with a length of 6 or 8 nt, and for selecting appropriate subsets for pooling is provided at http://bioinf.eva.mpg.de.

6. Resuspend the stock solution of SPRI bead suspension (AMPure kit) by vortexing. To make subsequent pipetting easier, add Tween 20 to the bead suspension to a final concentration of 0.05% (i.e., add 1 µL of Tween 20 to 2 mL of bead suspension).
7. Add SPRI bead suspension to the reactions as follows:
   - **i.** Add a 1.8-fold volume of SPRI bead suspension to each reaction (e.g., add 126 µL of SPRI beads to a 70-µL sample or 72 µL of SPRI beads to a 40-µL sample).
   - **ii.** Seal the wells with caps and vortex for several seconds. Ensure the beads are properly suspended and repeat vortexing if necessary.
   - **iii.** Let the plate stand for 5 min at room temperature.
   - **iv.** Collect the liquid at the bottom of the wells by briefly centrifuging in a plate centrifuge to 2000*g*.
8. Place the plate on a 96-well ring magnetic plate, and let it stand for 5 min to separate the beads from the solution. Pipette off and discard the supernatant without removing the beads.
9. Leave the plate on the magnetic rack, and wash the beads by adding 150 µL of freshly prepared 70% ethanol. Let stand for 1 min and remove the supernatant.
10. Repeat Step 9.
11. Using a multichannel pipette, remove residual traces of ethanol. Let the beads air-dry for 20 min at room temperature without caps.

IDT Exhibit 1060.0006

12. Elute as follows:

    i. Add 20 µL of EBT to the wells and seal the plate with caps.

    ii. Remove the plate from the magnetic rack, and resuspend the beads by repeated vortexing.

    iii. Let stand for 1 min, and then collect the liquid in the bottom of the wells by briefly centrifuging the plate to 2000*g*.

    Occasionally the beads may appear clumpy after vortexing; this does not have a negative effect on DNA recovery.

13. Place the plate back on the magnetic rack, let stand for 1 min, and transfer the supernatant to a new 96-well reaction plate.

    Carryover of small amounts of beads will not inhibit subsequent reactions.

Adapter Ligation

14. Prepare a master mix for the required number of ligation reactions as shown below. If white precipitate is present in the 10X DNA ligase buffer after thawing, warm the buffer to 37°C and vortex until the precipitate has dissolved. Since PEG is highly viscous, vortex the master mix before adding T4 DNA ligase and mix gently thereafter.

Reagent	Volume (µL) per sample	Final concentration in 40-µL reaction
H_2O	10	
T4 DNA ligase buffer (10X)	4	1X
PEG-4000 (50%)	4	5%
adapter mix from Step 2 (100 µM each)	1	2.5 µM each
T4 DNA ligase (5 U /µL)	1	0.125 U /µL

    When starting from low template quantities (50 ng or less), the amount of adapter mix can be reduced to 0.2 µL per reaction.

15. Add 20 µL of master mix to each eluate from Step 13 to obtain reaction volumes of 40 µL. Mix and incubate for 30 min at 22°C in a thermal cycler.

16. Perform reaction purification exactly as described in Steps 6-13. Elute in 20 µL of EBT.

Adapter Fill-In

17. Prepare a master mix for the required number of reactions.

Reagent	Volume (µL) per sample	Final concentration in 40-µL reaction
H_2O	14.1	
ThermoPol reaction buffer (10X)	4	1X
dNTPs (25 mM each)	0.4	250 µM each
Bst polymerase, large fragment (8 U/µL)	1.5	0.3 U/µL

18. Add 20 µL of master mix to each eluate from Step 16 to obtain reaction volumes of 40 µL. Mix well and incubate in a thermal cycler for 20 min at 37°C.

19. Perform reaction purification exactly as described in Steps 6-13. Elute the library in 20 µL of EBT.

Library Characterization

In addition to agarose gel electrophoresis (Step 20), performance of qPCR (Step 21) prior to indexing PCR (Steps 22-24) is strongly recommended, particularly if little sample DNA was used for library preparation. This is the only option to directly measure the number of molecules in the library. If the mean average fragment length and the size of the genome are known, this number can be used to determine whether the average coverage of genomic targets in the library is sufficiently high for subsequent target capture or direct sequencing. Step 21 describes a qPCR assay using SYBR Green (for more details, see Meyer et al. 2008a).

20. To verify the success of the library preparation, load 10 µL of the positive control library side-by-side with 100 ng of the original positive control sample and a size marker on a 2% agarose gel and perform electrophoresis.

    If all enzymatic reactions worked properly, the band produced by the control library should be shifted upward by 67 bp.

IDT Exhibit 1060.0007

See Troubleshooting.

21. Measure the number of molecules by qPCR:

    i. Prepare a standard dilution series by incrementally diluting an indexed sequencing library of known molecular concentration 10-fold in TET buffer.

    ii. If no such library is available, amplify 0.5 µL of the positive control in an indexing PCR (see Step 22). Purify the PCR product as described in Steps 6-13, determine its mass concentration on a spectrophotometer, calculate the molecular concentration, and use it as a standard as described in Step 21.i.

    iii. In a real-time PCR machine, amplify in parallel 1 µL of each standard dilution and each sample using primer IS4 and one of the indexing oligos; we recommend using a commercial PCR master mix containing SYBR Green (e.g., DyNAmo Flash SYBR Green qPCR kit). Set the annealing temperature to 60°C, and otherwise follow the instructions provided by the manufacturers of the kit and the real-time PCR machine.

    The concentration of molecules in the blank library (adapter dimers) should be at least one order of magnitude lower than in the sample libraries.

    It is often necessary to measure dilutions of the samples (e.g., 1000-fold in EBT) to obtain values within the detection range of the qPCR system.

Indexing PCR and Pooling

To avoid a downstream failure of Illumina's image analysis software, subsets of indexes must be chosen in a way that prevents unbalanced usage of the four nucleotides or the two laser channels during any cycle of index sequencing. The indexes provided with this protocol (see Supplemental Material [Indexing_Oligo_Sequences.doc]) are in an appropriate order to fulfill these requirements and should be used accordingly. For example, the first 22 indexes should be used if 22 indexes are needed. Fewer than four indexes should never be used in any experiment. Additional sets of indexes with different length and varying edit distance between indexes are provided on http://bioinf.eva.mpg.de. It will often not be necessary to use the entire library as template for indexing PCR. In this case, it is advisable to keep a backup that can be later used to add a different barcode to the sample.

Note that Phusion polymerase has proofreading activity. If this property is not desired (e.g., if deoxyuracil is present in the template DNA), another polymerase can be chosen for indexing PCR.

22. Prepare a PCR master mix for the required number of reactions. Dispense the master mix into a 96-well reaction plate, and then add template DNA and a different indexing primer to each well using a multichannel pipette.

Reagent	Volume (µL) per sample	Final concentration in 50-µL reaction
Master mix:		
H_2O	37.1 – *x*	
Phusion HF buffer (5X)	10	1X
dNTPs (25 mM each)	0.4	200 µM each
Primer IS4 (10 µM)	1	200 nM
Phusion Hot Start High-Fidelity DNA Polymerase (2 U/µL)	0.5	0.02 U/µL
Add separately to each well:		
Indexing primer (10 µM)	1	200 nM
Template DNA (library)	*x*	

If large amounts of sample DNA were used for library preparation (>100 ng), only a fraction of the library containing the equivalent of ~100 ng of starting material should be used for indexing PCR in order to prevent saturation of the PCR with template DNA.

23. Mix and perform cycling using the following temperature profile:

Initial denaturation	98°C	30 sec
Denaturation/cycle	98°C	10 sec
Annealing/cycle	60°C	20 sec
Elongation/cycle	72°C	20 sec
Final extension	72°C	10 min

IDT Exhibit 1060.0008

The optimal number of PCR cycles, that is, the number of cycles required to reach PCR plateau, will depend on the amount and concentration of template DNA and can be directly inferred from the amplification plots of the qPCR (Step 21). The cycle number can also be adjusted by rule of thumb according to the lowest amount of sample DNA that was used for library preparation: >100 ng → 12 cycles; >10 ng → 16 cycles, >1 ng → 20 cycles, >100 pg → 24 cycles.

24. Perform reaction purification exactly as described in Steps 6-13. Elute the indexed libraries in 25 µL of EBT.

25. Load 3 µL of some of the PCR products on a 2% agarose gel to verify amplification success.

    Indexed libraries prepared from sheared DNA should produce a smear. Due to the formation of heteroduplexes in the plateau phase of PCR (Ruano and Kidd 1992), the fragment size distribution inferred from the agarose gel may deviate slightly from the true distribution. However, no low-molecular-weight artifacts, such as primer dimers or adapter dimers, should be visible in the indexed sample libraries.

    See Troubleshooting.

26. Determine the DNA concentration, and pool the indexed libraries in equimolar ratios.

    The pool of indexed libraries is now ready for target capture or direct sequencing on one of Illumina's sequencing platforms.

    Due to the presence of heteroduplexes, qPCR is the only means of exactly determining the DNA concentrations in indexed libraries. However, concentration estimates derived from measurements with a spectrophotometer are sufficient in this step and more convenient. End product yield of indexing PCR is usually similar for all samples, particularly if there are no major differences in fragment size distribution. If this is the case, as can be confirmed by measuring DNA concentrations in a subset of indexed libraries, pooling equal volumes of all libraries will be sufficient.

Target Capture and/or Sequencing on the Illumina Platform

27. For target capture on microarrays, follow, for example, the exact procedure given in the protocol of Hodges et al. (2009) with the following modifications:

    i. Use a different set of blocking oligos (BO1-BO6).

    ii. Use primers IS5 and IS6 at an annealing temperature of 60°C for amplifying the library pool after capture.

28. For sequencing and data analysis, use the recipes, kits, and analysis tools for multiplex sequencing provided by Illumina.

    A tool for splitting up the qseq sequence files according to indexes is available in CASAVA 1.6 and later versions (demultiplex.pl). However, when using the 7-nt index sequences given in this protocol, the —qseq-mask parameter must be set to seven (the default is six). No modifications to the recipes provided by the Illumina machine control software (SCS) are required, because seven cycles of index sequencing are carried out by default. Additional software for data analysis on FastQ files (SplitFastQIndex.py), a file format created for example by the alternative base caller Ibis (Kircher et al. 2009), is provided on http://bioinf.eva.mpg.de. If single mismatches are allowed during index identification, the fraction of unidentified index sequences typically reduces to ~5%, as compared to ~15% when a perfect match is required. Using alternative base callers like Alta-Cyclic (Erlich et al. 2008), BayesCall (Kao et al. 2009), or IBIS (Kircher et al. 2009) may also increase the fraction of correctly identified indexes.

    Indexed sequencing libraries are compatible with all capture methods requiring sequencing libraries. It is recommended to carry the blank library all the way through target capture and/or sequencing. To avoid cross-contamination of samples through jumping PCR (Meyerhans et al. 1990), pools of indexed libraries should be amplified with a minimum number of PCR cycles or sequenced without amplification if possible.

    See Troubleshooting.

TROUBLESHOOTING

Problem: No size shift of the positive control library is visible on the agarose gel, or the size shift is incomplete.

[Step 20]

Solution: Consider the following:

1. Make sure the positive control PCR was generated using primers with unmodified 5′-ends.

2. One of the enzymes may have degraded. Replace all the enzymes and repeat.

IDT Exhibit 1060.0009

Problem: The positive control library shows no band on the agarose gel.
[Step 20]
Solution: Verify that the SPRI bead suspension is functional and the size cutoff is appropriate, for example, by purifying a DNA ladder as described in Steps 6-13.

Problem: Artifact bands are visible on the agarose gel after indexing PCR.
[Step 25]
Solution: If enough sample DNA was used for library preparation, artifact bands are only expected from the blank control. Repeat library preparation using more sample DNA or reduce the amount of adapters to 0.2 µL per reaction. Make sure a hot start polymerase was used for the indexing PCR.

Problem: Sequencing results in a low percentage of reads with correct index sequences.
[Step 28]
Solution: On the Genome Analyzer II, up to 5% of the raw sequences can generally be expected to be artifacts and up to 25% of low quality. If "N" base calls are more frequent in the index read than the other sequencing read(s), image analysis and downstream base calling partially or completely failed due to unbalanced usage of nucleotides or laser channels. Prepare a new pool of libraries with a more balanced composition of indexes (see Step 22) and repeat sequencing.

ACKNOWLEDGMENTS

We thank Michael Hofreiter, Svante Pääbo, Hernán Burbano, and Adrian Briggs for helpful discussions; Mark Whitten, David López Herráez, and Tomislav Maricic for comments on the manuscript; and the Max-Planck-Society for financial support.

REFERENCES

Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, et al. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* **5:** 887–893.

Erlich Y, Mitra PP, delaBastide M, McCombie WR, Hannon GJ. 2008. Alta-Cyclic: A self-optimizing base caller for next-generation sequencing. *Nat Methods* **5:** 679–682.

Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, Richard McCombie W, Hannon GJ. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat Protoc* **4:** 960–974.

Kao WC, Stevens K, Song YS. 2009. BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing. *Genome Res* **19:** 1884–1895.

Kircher M, Stenzel U, Kelso J. 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* **10:** R83. doi: 10.1186/gb-2009-10-8-r83.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Meyer M, Briggs AW, Maricic T, Hober B, Hoffner B, Krause J, Weihmann A, Paabo S, Hofreiter M. 2008a. From micrograms to picograms: Quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res* **36:** e5. doi: 10.1093/nar/gkm1095.

Meyer M, Stenzel U, Hofreiter M. 2008b. Parallel tagged sequencing on the 454 platform. *Nat Protoc* **3:** 267–278.

Meyerhans A, Vartanian JP, Wain-Hobson S. 1990. DNA recombination during PCR. *Nucleic Acids Res* **18:** 1687–1691.

Ruano G, Kidd KK. 1992. Modeling of heteroduplex formation during PCR from mixtures of DNA templates. *PCR Methods Appl* **2:** 112–116.

IDT Exhibit 1060.0010