Multiplex Illumina Sequencing Using DNA Barcoding

Koon Ho Wong,[1] Yi Jin,[1] and Zarmik Moqtaderi[1]

[1]Harvard Medical School, Boston, Massachusetts

ABSTRACT

The amount of sequence obtained by modern sequencing machines greatly exceeds the sequencing depth requirements of many experiments, especially those involving organisms with small genomes. In the interest of economy and efficiency, various strategies have been developed for multiplexing, in which samples are uniquely tagged with short identifying sequences known as barcodes, pooled, and then sequenced together in a single lane. The resulting combined sequence data are subsequently sorted by barcode before bioinformatic analysis. This unit contains a barcoding protocol for the preparation of up to 96 ChIP samples for multiplex sequencing in a single flow cell lane on the Illumina platform. This strategy may be extended to even larger numbers of samples and may also be generalized to other sequencing applications or sequencing platforms. *Curr. Protoc. Mol. Biol.* 101:7.11.1-7.11.11. © 2013 by John Wiley & Sons, Inc.

Keywords: barcoding • multiplexing • deep sequencing • next-generation sequencing • ChIP-seq

INTRODUCTION

Current sequencing machines routinely deliver well over 150 million reads per flow cell lane, a considerable improvement over even a few years ago, and this sequencing power is expected to increase further over time. This output is vastly more than necessary for many whole-genome experiments. For many species, especially those such as yeasts with small genomes, a large excess of genome coverage may be obtained with just a small fraction of the sequence obtained from a single run/reaction. Up to a point, any additional sequence may be useful in improving detection of low-frequency events or affording extra genome coverage and hence greater data reliability. Beyond this, however, there is little additional benefit. Given the high per-lane cost of deep sequencing, there is significant economic incentive to devise ways for different samples to share the sequencing capacity of a single flow cell lane. Such sample-pooling methods are generally referred to as multiplexing (Church and Kieffer-Higgins, 1988).

In the simplest approach, dissimilar DNA samples can be multiplexed so their sequences can be easily distinguished and deconvoluted. For example, in experiments involving species with sufficiently different genomes, it is possible to combine samples obtained from different species into a single sequencing lane. Provided that the sequence reads are long enough for unique mapping, the reads may be assigned to the correct organism after completion of the sequencing run (Tirosh et al., 2011).

More typically, though, experiments involve multiple samples derived from the same species, and other strategies must be employed. An increasingly popular method for allowing multiple samples to be distinguishable after sequencing in a mixed population is to append a small identifying sequence tag or "barcode" to the nucleic acid molecules of each sample. The barcoded samples are pooled and sequenced together, after which the sequence output is sorted by barcode and assigned to the proper samples. Multiplexing approaches have been successfully used for genome sequencing and

Current Protocols in Molecular Biology 7.11.1-7.11.11, January 2013
Published online January 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471142727.mb0711s101
Copyright © 2013 John Wiley & Sons, Inc.

Tecan Ex. 2014, Page 1 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

chromatin immunoprecipitation-sequencing (ChIP-seq) applications using different deep-sequencing platforms (Meyer et al., 2007; Lefrançois et al., 2009; Wong and Struhl, 2011).

BASIC PROTOCOL

CONSTRUCTION OF BARCODED ChIP-seq LIBRARIES

The following barcoding protocol describes the preparation of barcoded libraries for sequencing on the Illumina platform. The procedure (outlined in Fig. 7.11.1) is modified from the basic ChIP-seq protocol described by Illumina (*http://www.illumina.com*) and is intended for use with DNA samples from ChIP experiments. However, the experimental

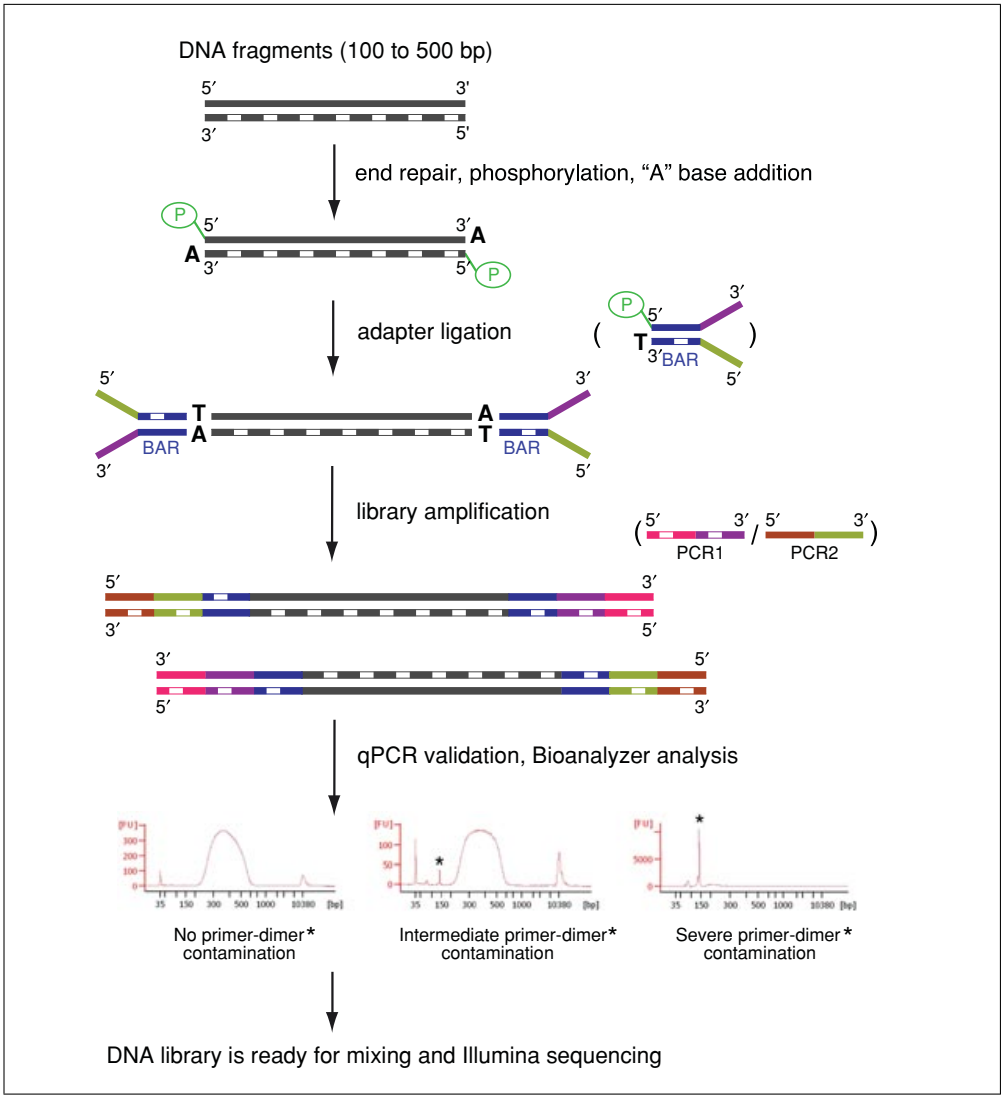

Figure 7.11.1 Outline for preparation of 5′ barcoded DNA libraries for Illumina sequencing. DNA fragments ranging from 100 bp to 500 bp are first subjected to end-repair, 5′ phosphorylation, and 3′ A base addition. Y-adapters with unique barcode sequences (depicted as blue lines) are then ligated onto both ends of DNA fragments. Purified ligation products are subjected to PCR amplification with primers (PCR1/2) containing the full sequences required for hybridization to the Illumina sequencing flow cell. The quality and quantity of purified library DNA can be evaluated by qPCR and Bioanalyzer analyses, respectively. DNA libraries are then ready for mixing and Illumina sequencing. When mixing libraries with severe primer dimer contamination, several rounds of gel or SPRI purifications may be needed to completely eliminate the contaminating primer dimers. A Bioanalyzer assay may be used to check for the presence of primer dimers in the final pooled library. For the color version of this figure, go to *http://www.currentprotocols.com/protocol/mb0711*.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 2 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

setup described can also be used for other applications, such as RNA-seq (*UNIT 4.11*) or genome re-sequencing, and the same general principle may be used in designing multiplexing strategies for other sequencing platforms.

Materials

1 to 10 ng ChIP DNA (see *UNIT 21.3*)
Nuclease-free water
10× T4 DNA Ligase buffer
10 mM dNTP mix
T4 DNA polymerase, 3000 U/ml (NEB)
T4 Polynucleotide kinase, 10,000 U/ml (NEB)
DNA polymerase I, large (Klenow) fragment, 5000 U/ml (NEB)
QIAquick PCR Purification kit (QIAGEN), or equivalent, containing Elution buffer (EB)
10× NEBuffer 2 (NEB)
10 mM dATP
Klenow Fragment (3′→5′ exo-), 5000 U/ml (NEB)
QIAquick MinElute PCR Purification kit (QIAGEN)
10 μM annealed Y-adapters (from Support Protocol 1)
Quick Ligation Kit (NEB) containing:
  2× Quick Ligation Reaction buffer
  Quick T4 DNA ligase
Phusion High-Fidelity DNA Polymerase, 2000 U/ml (NEB) containing 5× Phusion HF Reaction buffer
10 μM PCR primer mix (PCR1 and PCR2; see Table 7.11.1)
25 mM dNTP Mix
QIAquick Gel Extraction kit (QIAGEN)

1.5-ml microcentrifuge tubes
37°C heat block
0.5-ml PCR tubes
Thermal cycler
Gel electrophoresis apparatus (*UNIT 2.5A*)

End repair and 5′ phosphorylation

1. Start with 1 to 10 ng ChIP DNA with a size range of 100 to 500 bp in no more than 40.8 μl H_2O or EB.

   Lower DNA amounts have not been tested. DNA fragments should be less than 500 bp as cluster formation in the HiSeq flowcell is less efficient for longer DNAs. On the other hand, DNA smaller than 100 bp should theoretically work, but the final library size after adapter ligation would be similar to the size of contaminating primer dimers, making the purification challenging.

2. Combine and mix the following in a 1.5-ml microcentrifuge tube:

DNA + H_2O:	40.8 μl
10× T4 DNA Ligase buffer:	5 μl
10 mM dNTP mix:	2 μl
T4 DNA Polymerase:	1 μl
T4 Polynucleotide Kinase:	1 μl
DNA polymerase I, Large (Klenow) Fragment:	0.2 μl
Total:	50 μl.

3. Incubate the mixture for 45 min at room temperature.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 3 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

Table 7.11.1 Oligonucleotide Sequences for Illumina Paired-End Y-Adapter and PCR Primers

Oligo	Sequence
Barcoded adapter oligo 1 (BA1)	5′ ACACTCTTTCCCTACACGACGCTCTTCC GATCT-*barcode*-T 3′
Barcoded adapter oligo 2 (BA2)	5′ p-*barcode*(reverse complement)-AGATCGGAAG AGCGGTTCAGCAGGAATGCCGAG 3′[a]
PCR primer forward (PCR1)	CAAGCAGAAGACGGCATACGAGATCGGTCTCGG CATTCCTGCTGAACCGCTCTTCCGATC*T[b]
PCR primer reverse (PCR2)	AATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCCGATC*T[b]

[a]5′ p indicates 5′ phosphorylation.
[b]*Indicates phosphorothioate bond.

4. Purify DNA using a QIAquick PCR Purification kit. Elute DNA with EB buffer to obtain a final volume of 43 µl.

Add the 3′ A overhang

5. To the DNA sample from step 4, add the following:

10× NEBuffer 2:	5 µl
10 mM dATP:	1 µl
Klenow fragment (3′→5′ exo-):	1 µl
Total:	50 µl.

6. Incubate for 30 min at 37°C.

7. Purify DNA using a MinElute PCR Purification kit. Elute DNA with EB buffer to obtain a final volume of 10 µl.

Perform adapter ligation

8. Assign a unique barcode to each sample to be sequenced in the same sequencing run.

   Use the adapter pair corresponding to the selected barcode. The generic format of the barcode-containing adapters is shown in Table 7.11.1, and a list of 96 unique barcodes is presented in Table 7.11.2.

9. Dilute barcoded adapters with water as follows:

   a. For 1 to 2 ng of starting DNA, dilute the adapter 50×.
   b. For 3 to 10 ng of starting DNA, dilute the adapter 10×.

   The recommended amount of adapter only serves as a general guide and should be reduced accordingly if a high level of primer dimer contamination is detected in the final library. We generally perform this final adapter dilution immediately before use.

10. In a fresh 1.5-ml microcentrifuge tube, combine the following:

DNA from step 7:	10 µl
Diluted Y-adapter:	1 µl
2× Quick Ligation Reaction buffer:	15 µl
Quick T4 DNA ligase:	4 µl
Total:	30 µl.

11. Mix well and incubate for 15 min at room temperature.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 4 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

Remove free and self-ligated adapters

12. Add 70 µl H_2O to the reaction. Perform a one-step SPRI purification to remove DNA smaller than 200 bp (see Support Protocol 2). Elute the DNA with 40 µl EB.

    This step may be substituted with a gel purification step to isolate DNA between 200 and 500 bp. In addition, an optional gel purification step may be performed after the SPRI purification to remove any remaining self-ligated adapters from the sample before PCR amplification. This is recommended for completely eliminating primer dimers from the final library.

Perform PCR Amplification

13. In a 0.5-ml PCR tube, combine and mix the following:

DNA from step 12:	20 µl
H_2O:	16.6 µl
5× Phusion HF Reaction buffer:	10 µl
10 µM PCR primer mix:	2.5 µl
25 mM dNTP mix:	0.4 µl
Phusion DNA Polymerase:	0.5 µl
Total:	50 µl.

    Reserve half of the ligated material for use in case of PCR failure or over-amplification. Alternatively, the entire sample may be used, which is strongly encouraged if the starting sample contains less than 1 ng of DNA.

14. Perform PCR as follows:

Initial step:	30 sec	98°C (initial denaturation)
18 cycles:	10 sec	98°C (denaturation)
	30 sec	65°C (annealing)
	30 sec	72°C (extension)
1 cycle:	5 min	72°C (final extension)
	Indefinite	4°C (hold).

    Cycle number should be empirically determined when a different amount of starting DNA is used. Ideally, the minimum level of amplification that generates sufficient material for sequencing should be performed to minimize PCR artifacts and avoid over-amplification.

15. Add 50 µl H_2O to the reaction. Remove free primers and primer dimers using SPRI purification (see Support Protocol 2). Elute in 50 µl EB buffer or H_2O.

Evaluate the library

16. Dilute 1 µl of the purified library 50× in 200× in EB.

    The exact dilution factor depends on the DNA-associated protein being analyzed and should be empirically determined.

17. Using qPCR or PCR followed by gel electrophoresis, determine the enrichment level of the factor of interest at a few selected regions. The relative enrichment levels across the regions in the final library should be similar to those in the original ChIP DNA.

    This step is strongly recommended. A change in ChIP pattern after library construction is most likely due to PCR over-amplification. If this is observed, the level of amplification should be reduced accordingly.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 5 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

Quantitation and combination of barcoded libraries

18. Quantify DNA molarity by Bioanalyzer High sensitivity DNA assay (or equivalent, e.g. TapeStation).

    Undiluted sample should be used in the Bioanalyzer assay.

19. Combine the uniquely barcoded libraries at the appropriate molar ratio.

    The amount of material required from each library depends on the number of libraries to be included, the relative amounts of sequences each library desired, and the amount of material requested by sequencing facility. For example, if a sequencing facility requires 0.2 pmol of library DNA, for equimolar multiplexing of 24 libraries in a lane, approximately 0.0083 pmol is required from each of the 24 libraries. We routinely pool ~4 times more material than needed (~0.033 pmol) to allow for any DNA loss during the final purification(s).

Select for DNA between 200 and 500 bp in size

20. Run mixed DNA libraries on a 2% agarose gel (*UNIT 2.5A*). Excise the gel fragment containing DNA between 200 to 500 bp. Purify DNA from the gel slice using a QIAGEN Gel Extraction kit. Elute DNA in 20 µl EB.

    Reduce sample volume by SpeedVac, if necessary before this step. This purification step removes any residual free primers and primer dimers, as well as DNA fragments larger than 500 bp, which are less efficient in cluster formation. Alternative methods for size selection may be used. The final library may be checked on a Bioanalyzer for signs of primer dimer contamination. An additional SPRI purification should be performed to completely remove the contamination.

Sequence

21. Submit pooled samples for Illumina sequencing.

Analyze the data

22. After sequencing, separate the data by barcode and assign the reads to the proper samples. Sequencing facilities frequently provide this service. Otherwise, it may be accomplished via Unix commands or by using custom scripts. A simple Python script for the deconvolution of sequence data by barcode is available at *https://zarmik.hms.harvard.edu/Barcoding*.

SUPPORT PROTOCOL 1

PREPARATION OF BARCODED Y-ADAPTERS

Sequences of adapters and barcodes are provided in Tables 7.11.1 and 7.11.2, respectively. All oligonucleotides are synthesized, modified, and purified commercially.

Materials

Oligo BA1 and BA2 (see Table 7.11.1)
Annealing buffer (see recipe)

1.5-ml microcentrifuge tubes
0.5-ml PCR tubes
Thermal cycler

1. In separate 1.5-ml tubes, dissolve freeze-dried oligo BA1 and BA2 in annealing buffer to a final concentration of 200 µM.

2. For each barcode, mix 25 µl of the relevant BA1 and BA2 oligos in a 0.5-ml PCR tube.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 6 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

Table 7.11.2 Unique Barcode Sequences

AACACCTA	ACTAACTG	ATCGCCAG	CATTCCAA	CGCATTAA	CTTCGCGC
ACGTAGCT	ATCCTATT	CAGGAGGC	CGACTGGG	CTTAAGAT	GCAAGTAG
ATATAGGA	CACGTGTT	CGAACTGT	CTCTGTCT	GATGGAAT	GGCAGACG
CACAGTTG	CCTTTACA	CTAGTCAT	GATCCAGC	GGATATGG	GTGACTAC
CCTACAAC	CTAGATTC	GAGTTAAC	GGACGAGA	GTCTACAT	TATACCGT
CGTCGGCT	GAGAACTC	GCTGGCGA	GTCCATTA	TAGTCACA	TGACGCAT
GACGTCAA	GCTCAGTT	GTAGAGCT	TAGCTAGT	TGAATTCG	TTCCTCAC
GCGTTTCG	GTACTTGC	TACTGCGC	TCGGTACC	TTAAACAG	AATGCTGA
GGTCTGAC	TACGAATC	TCGCGTAC	TGTGCTAT	AATCACAC	AGGTCAGT
GTTTCACT	TCCTACTA	TGTAGGTC	AAGATTGC	AGGCAATG	ATTGCATC
TCCAGCCT	TGGTCTTC	AAGCGGTC	AGAACACC	ATGCATCC	CCATACAC
TGCGGTTA	AACCGTGT	ACTCTAAG	ATGAGGAA	CCAGCACG	CGCTTCTG
GCAGCCTC	GGTCCTTG	TGAGAGTG	TTGAGTGT	CAAGACCA	CCGTCTGA
GGCGAGGA	GTTGTCCC	TTCTGATG	ACCGTTAT	CCGGATAG	CGGTTGAT
GTGGGATA	TCATTAGG	ACAGTGCA	AGTTGCTA	CGGACGTG	GACACTCT
TATCTCCG	TGATCCGA	AGTAGTGG	CAATCATC	GAACGCTG	GCGGCGAA

3. Anneal oligos in a thermal cycler as follows:

   Step 1: 95°C for 5 min
   Step 2: Ramp −1°C per minute to a final temperature of 25°C
   Step 3: Hold at 4°C.

4. Dilute annealed adapters 10-fold with H_2O to obtain a working stock of 10 μM.

   To avoid repeated freeze-thaw cycles, store the working stock in small volumes in several tubes indefinitely at −20°C.

DNA PURIFICATION AND SIZE SELECTION USING SPRI-MAGNETIC BEADS

SUPPORT PROTOCOL 2

The protocol described below is for the DNA SizeSelector-I SPRI magnetic beads from Aline Biosciences. There are equivalent products from other companies, and their protocols may vary from what is described here. DNA SizeSelector beads are supplied in a solution containing PEG. When this slurry is mixed with the sample, the final PEG concentration determines the range of DNA sizes that will be retained by the beads.

Materials

DNA SizeSelector-I beads (Aline Biosciences)
DNA sample
70% Ethanol
EB (from Qiaquick kit) *or* H_2O

Vortex mixer
1.5-ml microcentrifuge tubes
Magnetic stand

1. Equilibrate DNA SizeSelector beads to room temperature for 30 min. Vortex the bead slurry thoroughly before use.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 7 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

2. Add 2 vol DNA SizeSelector bead slurry to the DNA sample (e.g., add 200 µl DNA SizeSelector slurry to 100 µl of sample) in a 1.5-ml microcentrifuge tube, and mix well by pipetting.

   The recommended ratio of DNA SizeSelector bead slurry to sample volume is specific for purifying DNA fragments larger than 200 bp. The volume should be empirically determined (e.g., using a 100-bp ladder) if a different size range is desired. Small variations in the final PEG concentration will significantly affect the size range retained by the beads. For this reason, we recommend avoiding small DNA sample volumes (e.g., <50 µl) to minimize sample-to-sample variability caused by potential pipetting error.

3. Incubate for 5 min at room temperature, and place the tube on a magnetic stand for 2 min.

4. Remove the supernatant carefully to avoid losing any beads. Discard the supernatant.

5. Wash the beads twice with 200 µl of 70% ethanol. Mix thoroughly by pipetting.

6. Remove ethanol and air-dry the tube on a magnetic stand for 10 min or until the beads have completely dried.

7. Add the desired volume of buffer EB or H_2O to elute the DNA from the dried beads. Mix by pipetting. Replace the tube back on the magnetic stand for 1 min.

   Dried beads have a tendency to disperse when the tube is removed from the magnetic stand; therefore, keep the tube on the magnetic stand until EB or H_2O is added.

8. Recover the supernatant and transfer into a new microcentrifuge tube. To avoid transferring any beads, leave ~2 µl of elution buffer in the tube.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps. For common stock solutions, see ***APPENDIX 2****; for suppliers, see* ***APPENDIX 4****.*

Annealing buffer

10 mM Tris·Cl, pH 7.5 (*APPENDIX 2*)
50 mM NaCl
1 mM EDTA, pH 8
Store indefinitely at room temperature

COMMENTARY

Background Information

As the sequence yield of next-generation sequencing machines continues to improve, the advantages of pooling samples for parallel sequencing become increasingly obvious. A single flow cell lane provides sufficient sequence output for a large number of whole-genome experiments. Multiplexing 48 samples in a single sequencing lane results in a sequencing cost per sample (not including library preparation) of approximately US $25. Thus, the cost of obtaining whole-genome results by this method compares favorably with the cost of testing even a small number of specific target genes by qPCR. The multiplexing procedure described in this unit, therefore, allows a substantial reduction in the per-sample cost for many sequencing applications, thereby making whole-genome experiments very affordable.

In the simplest sense, the number of samples that can be multiplexed is limited by total sequence output per lane and by the amount of sequence required for each sample. If necessary, more barcodes can be generated, and if needed, the barcode length may be increased, as long as the remainder of the sequence read after the barcode is long enough to allow unique mapping to the genome. With current read lengths of ~50 nt, this is not a problem. In practice, the number of barcoded samples actually pooled also depends on a balance of cost and practical considerations such as the time needed to accumulate (or cooperate with one's neighbors to gather) larger numbers of samples for pooling.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 8 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

The protocol provided here was developed for ChIP samples but may be generalized to other types of sequencing experiments. Care must be taken when combining samples to take into account the different requirements of ChIP-seq and other applications such as genome re-sequencing, for which greater sequence depth (and hence fewer samples per lane) is generally desirable. For a ChIP-seq experiment with a site-specific DNA-binding transcription factor in *S. cerevisiae*, computational simulation indicates that about 2.5 million reads are required to identify over 95% of binding sites with an average of 2.5-fold enrichment (Lefrançois et al., 2009). Therefore, an average HiSeq run with an output of 150 million reads could be used for as many as 60 transcription factor ChIP-seq experiments. For genome re-sequencing, with a typical 50-nt read length (from which we subtract 9 nt for the barcode and final T), 25-fold sequencing coverage of the *Saccharomyces cerevisiae* genome may be obtained with 7.3 million reads. Approximately 20 such experiments could therefore be multiplexed in one Illumina HiSeq lane. Of course, it is not necessary that all the pooled experiments be of the same type: samples requiring different sequencing depths may be multiplexed as long as care is taken to mix them at a molar ratio that will yield the requisite number of reads devoted to each sample.

The protocol and reagents described in this unit are specific for multiplexing using adapters with 5′ barcodes. Illumina has an alternative commercially available multiplexing system known as "Indexing," in which barcode sequences are embedded within the PCR primers instead of the ligated adapters. The indexing system uses a different set of adapters and PCR primers (see *http://www.illumina.com* for details). Additionally, indexing requires a preliminary sequencing run to acquire the index information, which may involve additional sequencing cost. By contrast, barcode sequences are part of the actual output reads in the 5′ barcoding method, and therefore incur no additional cost. However, the usable information for each read is shortened by the length of barcodes, and in an average run of 100 to 150 million reads, the barcodes could actually take up a significant fraction of the output, which makes the 5′ barcoding approach less cost-effective for applications that depend heavily on actual sequence coverage (e.g., whole-genome sequencing). On the other hand, the inclusion of 5′ barcodes as part of the reads means that the level of multiplexing can be significantly increased simply by extending the barcode length. This is not feasible for the indexing setup, as increasing index length may change the intrinsic adapter structure.

When using multiplexing methods, a small amount of extra time is required for deconvolution and sorting of the combined sequence output. This is frequently accomplished by the sequencing facility, but if such a service is not provided, it may be achieved as the first step in bioinformatic analysis by using UNIX commands or custom-made programs in simple scripting languages.

Critical Parameters

Size of input DNA material

The starting DNA material should be around 100 to 500 bp in size. For ChIP, we recommend extensive fragmentation of cross-linked chromatin to an average of 200 bp before immunoprecipitation. For other applications, DNA may be sheared using a sonicator or nebulizer. If a DNA sample includes a broad range of fragment sizes, the DNA fragments of 100 to 500 bp should first be enriched (e.g., by gel purification or other relevant techniques) prior to this procedure.

Barcode design and selection

A list of 96 unique 8-bp barcodes is presented in Table 7.11.2. Other combinations are possible, and additional 8-bp barcodes can be included to increase multiplexing capacity. In theory, the number of possible barcodes increases by a factor of four with every one bp extension in barcode length. However, it is critical that each barcode be unambiguously distinguishable from the others within the entire set even in the event of sequence errors within the barcode region. Additionally, as the Illumina cluster-calling algorithm expects roughly equal base composition at any given position, it is important to have a balanced base composition at every position of the barcode mixture across the entire collection or among the subset of barcodes used in any given multiplexing experiment. When using barcodes of the same length as those shown in Table 7.11.2, all reads would have a "T" nucleotide at position 9, immediately after the barcode, as a result of the "T" included in the adapter for ligation to the 3′ A-overhang ChIP DNA. While this may be tolerated, sequencing of such libraries is more prone to failing at high clustering density as compared to libraries not uniformly

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 9 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504

possessing a "T" at position 9, presumably due to a failure in the cluster-calling process. This issue may be more prevalent for libraries with short barcodes (e.g., 6 bp or less) in which the constant "T" occurs at an early position, before clusters are successfully assigned. In this case, we recommend adding an additional base (e.g., alternating with A, C or G) after the barcode for half the adapters, such that the position of the final "T" is no longer uniform across all samples.

Number of samples multiplexed per sequencing lane

As the initial synthesis of barcoded adapters can be costly, the number of available unique barcoded adapters usually imposes the upper limit for multiplexing capacity. In general, the number of libraries to be included in a run depends on the sequencing application, and is directly related to the amount of sequence required for each sample. An average HiSeq run with an output of 150 million reads could be used for as many as 60 site-specific transcription factor ChIP-seq experiments. On the other hand, a transcription profiling experiment involving ChIP-seq of RNA polymerase II, which binds over the entire body of active genes, would require a greater number of reads for adequate sequencing depth. In this instance, fewer samples would be multiplexed or, instead of equimolar mixing, the Pol II samples would be allocated a higher proportion in the pooled sample.

Number of PCR cycles for amplification

This is one of the most important parameters in the library preparation procedure. In addition to amplifying the DNA, the PCR step serves to introduce relevant adapter sequences for hybridization to the flow cell and for cluster formation (Fig. 7.11.1). As such, the number of PCR cycles used could influence the clustering behavior of the final library, i.e., some library DNA fragments may lack the necessary adapter sequences if too few PCR cycles are used. On the other hand, given that PCR can introduce biases, excessive PCR cycles may compromise the quality of the result for downstream analyses. DNA libraries should always be checked for signs of over-amplification (change in ChIP signal pattern; see Troubleshooting), and the amount of PCR amplification should be reduced if necessary. Ideally, the minimum level of amplification that generates sufficient material for sequencing should be performed. For ChIP-seq libraries, we find that 14 to 18 cycles of amplification are ideal for 1 to 10 ng of starting DNA material. However, the optimal PCR cycle number should be determined empirically if a significantly different amount of starting DNA is used.

Troubleshooting

No or low library content after PCR amplification

A lower-than-recommended starting DNA amount might have been used. When the sample DNA concentration is low, we recommend estimating the concentration using highly sensitive methods (e.g., Quant-iT, qPCR or pico-green assays) over conventional A_{260} measurement.

Sample contamination by primer dimers

Moderate primer dimer contamination in the purified DNA library after PCR amplification (i.e., step 18) is acceptable (see sample Bioanalyzer profiles in Fig. 7.11.1), and can be readily removed by gel purification after library pooling. However, when libraries are severely contaminated with primer dimers, special care should be taken to ensure their absence from the final pooled sequencing library. A gel purification step for DNA of 200 to 500 bp may be included after the SPRI purification of ligated material (i.e., step 12) to reduce any remaining trace amounts of self-ligated adapters in the DNA sample before PCR amplification.

High percentage of final output fails to map to the genome of interest

One major cause for this is contamination by primer dimers and/or self-ligated adapters in the final combined sequencing library. The problem can be resolved by performing an additional purification (e.g., gel or DNA SizeSelector-I purification) after the final size selection step (i.e., after step 20). This is strongly recommended when some of the pooled libraries are severely contaminated with primer dimers (see sample Bioanalyzer profiles in Fig. 7.11.1). Alternatively, follow the guideline discussed above to reduce primer dimer contamination in the libraries. A Bioanalyzer run may be used to evaluate the level of contamination by primer dimers. An optimal library to be sequenced should be free from any primer-dimer contamination (see Fig. 7.11.1 for an ideal Bioanalyzer profile).

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ChIP signal pattern change after library construction

The relative signals among different target regions should be maintained after library construction to ensure good data quality. Skewed ratios of ChIP signals are most likely caused by PCR over-amplification. Repeat the library construction procedure with fewer amplification cycles and reevaluate the new library.

Anticipated Results

With 18 rounds of PCR amplification on half of the ligated material, 10 ng and 1 ng of 100 to 300 bp DNA would yield ~1 to 6 pmol and 0.05 to 0.75 pmol, respectively, of library DNA. The sequencing output (e.g., read length and number of reads) depends mainly on the Illumina system and the type of sequencing run. The quality of libraries can significantly influence cluster formation, and as a result, final read output. A competent sequencing facility would empirically determine the optimal concentration of DNA libraries to load on a flow cell for maximum possible output based on the results and clustering behaviors of previous samples. If each barcoded library is mixed at equal molarity, the same proportion of the total output is expected across all libraries. Based on our experience, when primer dimers are effectively removed, over 85% of raw output reads can be mapped to the relevant genome.

Time Considerations

Preparation (i.e., annealing and dilution) of barcoded Y-adapters takes ~2 hr. These annealed adapter stocks may be stored for use in multiple experiments. The amount of time required for generating starting DNA materials depends on the application. The end-repair, A-addition, ligation, size selection, and various purification steps require about 4 hr in total. Although all of these reactions can be performed on the same day, samples can be purified and stored up to 2 days at 4°C or indefinitely at −20°C after each enzymatic step. Library amplification by PCR and the subsequent clean-up take around 2 hr. While determination of library concentrations by Bioanalyzer takes only 30 to 40 min, this step is often carried out as a service by a core facility, and the time required is subject to that facility's schedule. The combining of different barcoded libraries and final purification steps take ~1 to 3 hr.

Acknowledgments

We thank Dr Joe Mellor for generously sharing the list of barcodes. This work was carried out in the laboratory of Prof. Kevin Struhl. We thank Prof. Struhl for his comments on the protocol. KHW is supported by a postdoctoral fellowship from the Croucher Foundation.

Literature Cited

Church, G.M. and Kieffer-Higgins, S. 1988. Multiplex DNA sequencing. *Science* 240:185-188.

Lefrançois, P., Euskirchen, G.M., Auerbach, R.K., Rozowsky, J., Gibson, T., Yellman, C.M., Gerstein, M., and Snyder, M. 2009. Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. *BMC Genomics* 10:37.

Meyer, M., Stenzel, U., Myles, S., Prüfer, K., and Hofreiter, M. 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res.* 35:e97.

Tirosh, I., Wong, K.H., Barkai, N., and Struhl, K. 2011. Extensive divergence of yeast stress responses through transitions between induced and constitutive activation. *Proc. Natl. Acad. Sci. U.S.A.* 108:16693–16698.

Wong, K.H. and Struhl, K. 2011. The Cyc8-Tup1 complex inhibits transcription primarily by masking the activation domain of the recruiting protein. *Genes Dev.* 25:2525–2539.

Internet Resource

http://zarmik.hms.harvard.edu/Barcoding
Python scripts for data pre-processing and analysis.

19343647, 2013, 1, Downloaded from https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471142727.mb0711s101 by Stanford University, Wiley Online Library on [22/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tecan Ex. 2014, Page 11 of 11
IDT, Inc. v. Tecan Genomics - IPR2024-01504