

Preparing samples for 3' Tag-based RNA-Seq

Last updated 4 April 2012 by Eli Meyer for Illumina HiSeq.

Introduction

This protocol describes in detail the procedures used to prepare cDNA fragment libraries for quantitative analysis of gene expression (RNA-Seq) by deep sequencing on Illumina HiSeq. About 0.6-1 µg of DNase-treated total RNA is required per sample, and this starting material should be carefully quantified and analyzed by gel electrophoresis prior to beginning these procedures to verify that the RNA is intact, and free of genomic DNA contamination.

The procedure can be reasonably completed within three days. Day 1: RNA is fragmented and used to synthesis cDNA (steps 1-2). Day 2: cDNA is amplified, sample-specific barcodes are incorporated, and size-selection is accomplished by means of gel extraction (steps 3-4). Day 3: the preparations are tested by amplification with the Illumina primers to verify that the material can be amplified in emPCR for Illumina sequencing. Finally, prior to sequencing the finished product on Illumina, we recommend confirming the constructs by cloning a small aliquot of the Illumina-ready product (step 4i) for Sanger sequencing. These additional quality control steps obviously require additional time, but allow the user to detect any major artifacts prior to investing resources in Illumina sequencing. The sequences of all oligonucleotide primers used in this protocol are provided at the end of this document, to allow synthesis of custom oligos as needed.

PLEASE NOTE: At several stages in the protocol, samples are tested to optimize cycle numbers or times, and to verify the absence of contamination. If you are working with a large number of samples, there is no need to conduct these tests on ALL samples. I suggest selecting a representative subset (n~4-6) at each stage. This comment applies to steps 1b, 3a-e, 4a-g, and the optional step 5.

1. RNA fragmentation

NOTE: the buffer in which the original RNA is incubated is critically important for the success of fragmentation, as are the volume and concentration of the RNA. Prior to working with the precious experimental samples, we recommend testing a range of different incubation times to identify the duration that produces the appropriate size range in these samples. RNA should be treated with DNase prior to cleanup (e.g. using an on-column DNase treatment) so that DNA does not interfere with quantification.

- a. Aliquot 1 µg of total RNA in 10 µl of 10 mM Tris (pH 8.0) in a 96-well PCR plate. If concentrations are not sufficiently high to allow this loading, RNA

samples can be concentrated in the Speedvac prior to loading. Set aside an additional sample (~100 ng) of the original intact RNA for comparison with the fragmented samples.

- b. Carefully seal all wells and incubate RNA at 95°C for 5-30 minutes to fragment to the desired size range (~100 – 500 bp). This can be most easily accomplished in a thermocycler. In our previous work with grasses the optimum time has been ~10 minutes.
- c. Analyze 100 ng from each sample of fragmented RNA, alongside the intact RNA from the same sample, on a standard 1% agarose gel to evaluate the molecular weight of the fragmented RNA.

2. First-strand cDNA synthesis

NOTE: if RNA quantity is not limiting, first-strand cDNA should be synthesized using 1 µg of fragmented RNA. The reaction shown below is intended for ~ 1 µg; this can be doubled or halved as needed if the amount of RNA is very different from 1 µg.

- a. Measure the volume remaining after fragmentation using a pipette. The following recipe assumes a starting volume of 9 µl (10 µl minus evaporation), so if the volume is lower than this, add additional water to achieve 9 µl.
- b. Add 1 µl of the 10 µM oligonucleotide 3ILL-20TV to each well. Incubate at 65°C for 3 minutes in a thermocycler, then transfer immediately onto ice.
- c. Prepare a cDNA synthesis master mix. The following volumes are intended for a single reaction, so multiply these values by the number of reactions plus a small amount (~10%) to account for pipetting error.

(all volumes given in µl)

H ₂ O	1
dNTP (10 mM ea)	1
DTT (0.1 M)	2
5X first-strand buffer	4
10 µM 5ILL-SW	
(RNA oligonucleotide; stored at -80°C)	1
SuperScript II Reverse Transcriptase	
(Invitrogen #18064-022).	1

- d. Add 10 μl of this master mix to the RNA from (2b), mix thoroughly, and incubate in a thermocycler for one hour at 42°C.
- e. Incubate at 65°C for 5 minutes to inactivate the RT, dilute 1:5 in H₂O, and store on ice or at -20°C until ready to proceed to the next step.

3. cDNA amplification

NOTE: it is important not to over-amplify the cDNA at this stage, to avoid artifacts and distortion of expression ratios. The fewer cycles used, the better. If a visible smear is not produced within 17-20 cycles, try repeating the PCR with additional template (diluted FS-cDNA from step 2e), and correspondingly less water in the master mix. If no smear is detected this indicates a problem with the first-strand cDNA synthesis.

- a. Prepare a set of master mixes for small-scale PCR tests. The following volumes are intended for a single reaction each, so multiply these values by the total number of reactions plus a small additional amount to account for pipetting error. This recipe assumes 3 μl of template, so if you use a different amount of template, adjust the water accordingly.

	(Volumes given in μl)			
	A	B	C	D
H ₂ O	12.6	12.2	12.2	11.8
dNTP (2.5 mM ea)	2	2	2	2
10X PCR buffer	2	2	2	2
10 μM 5ILL oligo	0	0.4	0	0.4
10 μM 3ILL-20TV oligo	0	0	0.4	0.4
Titanium Taq polymerase (Clontech #639208)	0.4	0.4	0.4	0.4

- b. For each of the original RNA samples, prepare four PCR tubes labeled A-D. Add 18 μl of the appropriate master mix to each tube.
- c. Add 2 μl diluted FS-cDNA from (2e).
- d. Amplify in a thermocycler using the following profile:
95°C 5 min, (95°C 40 sec, 63°C 1 min, 72°C 1 min) X 17 cycles

- e. After 17 cycles check 5 μ l of the PCR products for all reactions on a standard 1% agarose gel. A smear of cDNA (~100-500 bp) should be faintly visible in reaction D, and nothing should be detected in the reactions A-C. If nothing is detected in reaction D, you can continue the reaction for additional cycles (up to a maximum of 20), or repeat the reaction with additional template. If product ever appears in reactions A or C, this indicates too much template, too many cycles, or contamination in one or more reagents. A small amount of amplification can be tolerated in reaction B, since the 3ILL-20TV oligo omitted in that reaction is still present in the original FS-cDNA. The goal at this stage is to identify the minimum cycle number that produces a visible smear in reaction D while remaining clean in reactions A-C.
- f. Once the optimum amount of template and number of cycles have been determined (3e), prepare a single large-scale reaction for each cDNA sample as follows. This recipe assumes 15 μ l of template, so if you use more template adjust the water accordingly.

(all volumes given in μ l)

H ₂ O	59
dNTP (2.5 mM ea)	10
10X PCR buffer	10
10 μ M 5ILL oligo	2
10 μ M 3ILL-20TV oligo	2
Titanium Taq polymerase (Clontech #639208)	2

- g. After PCR, check 5 μ l of the product on a gel to verify that the reaction worked as expected before freezing or purifying the product.
- h. Purify PCR products using the NucleoMag 96 PCR Cleanup Kit, according to the manufacturer's instructions.
- i. Working in a 96-well plate or PCR strip, split each PCR into two 50- μ l portions.
- j. Add 6 μ l P-beads and 138 μ l buffer MP1 to each well and mix thoroughly by back pipetting.
- k. Place plate on magnetic separator plate and hold undisturbed for 1 minute to collect beads. Remove supernatant without disturbing beads.
- l. Wash beads with 200 μ l MP2; repeat step k.
- m. Wash beads with 200 μ l MP3; repeat step k.

- n. Repeat step m. On this final rinse combine the two aliquots of each sample by suspending each in 100 μl of MP3, then combining into a single pool of 200 μl . Be careful to remove all visible supernatant after magnetic separation.
- o. Dry the beads 10 minutes at room temperature.
- p. Add 25 μl elution buffer (MP4) to the dried beads and mix well by back pipetting. Hold (not on the magnetic plate) for 5 minutes at room temperature.
- q. Place plate on magnetic separator plate and hold undisturbed for 1 minute to collect beads. Collect eluted DNA without disturbing beads, and transfer into a new 96-well plate or PCR strip.
- r. Quantify the purified products by OD260 (e.g. Nanodrop).

4. Adaptor extension and size selection

NOTE: Because the size distribution of templates is a critical factor for successful cluster amplification, any templates intended for sequencing on Illumina should be carefully size-selected prior to sequencing. The directions below outline a simple procedure for selecting fragments ranging from 250-300 bp in size that does not require any special equipment. Other methods of size selection could be substituted provided they achieve this same size range.

Note that on HiSeq, it is possible to “barcode” both ends of the construct. This opens the possibility of combinatorial barcodes (e.g. 10 unique 5’ oligos and 10 unique 3’ oligos allow for 100 unique combinations). Example barcodes are listed at the end of this document. These barcodes are incorporated at this stage.

- a. First, four test-scale PCRs are prepared for each sample to verify yield and specificity of the reaction, each using 10 ng PCR product (step 3, above) as template. It may be useful to dilute an aliquot from each sample to be tested so that all templates are at the same concentration. (e.g. you could prepare a 5 ng μl^{-1} solution for each, and use 2 μl of that dilution as a template).
- b. Prepare a small amount of diluted barcode oligo (1 μM), with a different barcode used for each sample. Be sure to write down which barcode is assigned to each sample at this stage, since this cannot be easily determined later in the process.
- c. Prepare four separate master mixes for small-scale test PCR. The following volumes are for a single reaction, so multiply these values by the total number of samples plus a small additional amount to account for pipetting error. The

values shown here assume the use of 2 μ diluted PCR product as template, so if you change this be sure to change the volume of water accordingly.

	(volumes given in μ l)			
	A	B	C	D
H ₂ O	5.8	3.8	3.8	1.8
dNTP (2.5 mM ea)	1	1	1	1
10X PCR buffer	1	1	1	1
Titanium Taq polymerase	0.2	0.2	0.2	0.2
Master mix per reaction	8	6	6	4
TruSeq-Mpx oligo (1 μ M)	0	2	0	2
TruSeq-BC oligo (1 μ M)	0	0	2	2

- d. Aliquot the specified amounts of master mix into each well, then add diluted (1 μ M) multiplex and barcode oligos to reactions B-D as specified in the above table.
- e. Add 2 μ l of the diluted cDNA (steps 3l, 4b) to each of these four primer combination reactions (A-D). The total volume of each reaction is 10 μ l.
- f. Amplify in a thermocycler using the following profile:
95°C 5 min, (95°C 40 sec, 63°C 1 min, 72°C 1 min) X 4 cycles
- g. After 4 cycles check 5 μ l of the PCR products for all reactions on a gel. The ideal result is a faint smear in reaction D, with no visible product in reactions A-C. If products appear in these control reactions, consider using less template. If nothing is detected in reaction D, add 1-2 more cycles and check the results on a gel. A small amount of product in the controls can be tolerated, but if reactions A-C are comparable in intensity to reaction D something is wrong.
- h. When the optimum number of cycles and volume of template have been determined, prepare a large-scale reaction based on those parameters with 50 ng template in 50 μ l total volume. The following master mix assumes the use of 10 μ l of template per 50 μ l reaction; if you adjust this template volume be sure to adjust the volume of water accordingly. This recipe is for a single

reaction, so multiple these values by the number of samples to be prepared plus a small additional amount for pipetting error.

(volumes given in μl)

H ₂ O	27
dNTP (2.5 mM ea)	5
10X PCR buffer	5
Titanium Taq polymerase	1

- i. Aliquot 38 μl of master mix to each well, then add 1 μl of the appropriate barcode oligo (10 μM), 1 μl of the appropriate multiplex oligo (10 μM), and 10 μl of PCR product (step 3l).
- j. Be sure to write down the barcode assignments! (e.g. sample X got multiplex Y and barcode Z).
- k. Amplify these reactions using the same profile and cycle number as determined above.
- l. Prepare a gel for size selection. This preparative gel should consist of 2% agarose in 1X TBE buffer, with SYBR Safe DNA staining dye (Invitrogen # S33102) added according to the manufacturers' instructions (1:10,000 dilution). A low-molecular weight ladder is required for accurate selection of the appropriate sizes; we recommend pBR322 DNA-MspI Digest (New England Biolabs # N3032S) or Low MW Ladder (NEB, #N3233S). Be sure to use large volume combs to allow loading of the entire 50 μl reaction into a single well.
- m. Load samples and run the gel until marker bands in the 50-200 bp size range are well separated. Illuminate the gel very briefly (< 30 seconds total exposure time) on a UV-transilluminator set at low intensity, for just long enough to mark the appropriate region (250-300 bp) with a clean razor blade. Turn off the UV light and carefully cut out the marked region, transferring it into a microcentrifuge tube.
- n. Extract the cDNA from this gel slice by adding 40 μl of nuclease-free water to the tube, spinning down briefly to bring water and gel slice into contact, and incubating overnight at 4°C. This product is now ready for sequencing on

Illumina. However, before proceeding with sequencing, we strongly recommend the following quality control steps (step 5, below).

5. Amplification of size-selected fragments

NOTE: this step is a useful test of the final constructs, but is not required for Illumina sequencing, which only requires about 0.5-1 pg of input cDNA. The goal of this step is to verify that the constructs produced in this procedure include the intended primer binding sites, are free of artifacts e.g. poly-A inserts, and fall within the appropriate range of molecular weights.

- a. Quantify the cDNA constructs from step 4n, using 0.5X TBE as the blank, and calculate the volume required to use 5 ng as template for the following reaction. Prepare a small volume of diluted template for each sample such that the same volume can be used in each reaction.
- b. Prepare a PCR master mix. The following volumes are for a single reaction, so multiply these values by the total number of reactions plus a small additional amount to account for pipetting error. The values shown here assume a cDNA concentration of 5 ng μl^{-1} ; the volumes of cDNA and water should be adjusted according to the concentration of cDNA in your samples:

(volumes given in μl)

H ₂ O	5.6
dNTP (2.5 mM ea)	1
10X PCR buffer	1
ILL-Lib1-20 oligo (10 μM)	0.2
ILL-Lib2 oligo (10 μM)	0.2
Titanium Taq polymerase	1

- c. Add 5 ng of cDNA template to each reaction (from 4n), for a total reaction volume of 10 μl .
- d. Amplify in a thermocycler using the following profile:
95°C 5 min, (95°C 40 sec, 63°C 1 min, 72°C 1 min) X 15 cycles
- e. After 15 cycles check 3 μl of the PCR products for all reactions on a gel. If a clean, narrow smear ranging from 250-300 bp is detected the constructs are correct. If no product is visible, amplify for an additional 1-2 cycles and check

- the product on a gel. If no product is visible before 18 cycles, this suggests a problem; consider repeating the PCR with additional template.
- f. Once the constructs pass the PCR tests in (5e) they are ready for Illumina sequencing. The purified template from step (4n) is the material to be sequenced. We strongly recommend cloning and sequencing a small number of clones by Sanger sequencing (n=6-24), to verify that no major artifacts are present (e.g., adaptor concatenation or excessive poly-A tracks), prior to sequencing this material on Illumina.

Sequences of oligonucleotide primers used in this protocol

Minimal oligo set

5ILL-SW

ACCCCAUGGGGCUACACGACGCUCUCCGAUCUGGG

(Note: this is an RNA oligo. Store small aliquots at -80 and minimize freeze-thaw cycles).

3ILL-20TV

ACGTGTGCTCTTCCGATCTAATTTTTTTTTTTTTTTTTTTT

5ILL

CTACACGACGCTCTTCCGATCT

TruSeq-Mpx-01

ATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCT

TruSeq-BC01

CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

(Note: this long oligo calls for either special synthesis, e.g. Ultramer synthesis, or special purification, e.g. HPLC).

ILL-Lib1-20

AATGATACGGCGACCACCGA

ILL-Lib2

CAAGCAGAAGACGGCATAACGA

Example barcode oligos

As described above, the HiSeq platform allows a unique barcode to be assigned to each end of the construct. This makes combinatorial barcode assignments possible, so that a set of 10 unique 5' oligos ("MPX") and 10 unique 3' oligos ("BC") can be used to produce 100 unique combinations. This approach reduces the total cost of barcode oligos, and increases the number of samples that can be combined within a lane.

The following list of oligos allows for 96 unique combinations. Additional barcodes can be found at:

<https://wikis.utexas.edu/display/GSAF/Illumina+-+all+flavors>

TruSeq-BC51	CAAGCAGAAGACGGCATAACGAGATAGTCCCCTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC52	CAAGCAGAAGACGGCATAACGAGATCTTACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC53	CAAGCAGAAGACGGCATAACGAGATTTCTTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC54	CAAGCAGAAGACGGCATAACGAGATAGCATTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC55	CAAGCAGAAGACGGCATAACGAGATCCTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC56	CAAGCAGAAGACGGCATAACGAGATACGCGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC57	CAAGCAGAAGACGGCATAACGAGATGGTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC58	CAAGCAGAAGACGGCATAACGAGATTGTGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC59	CAAGCAGAAGACGGCATAACGAGATATAATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC60	CAAGCAGAAGACGGCATAACGAGATTGGGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC61	CAAGCAGAAGACGGCATAACGAGATAGAGACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-BC62	CAAGCAGAAGACGGCATAACGAGATTCCTATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
TruSeq-Mpx-2	AATGATACGGCGACCACCGAAAAATACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-3	AATGATACGGCGACCACCGACTCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-4	AATGATACGGCGACCACCGACAGTAACACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-5	AATGATACGGCGACCACCGAGCACCACACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-6	AATGATACGGCGACCACCGACGGCGACACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-7	AATGATACGGCGACCACCGAGCGTTACACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-8	AATGATACGGCGACCACCGATTGTTACACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq-Mpx-9	AATGATACGGCGACCACCGACTGTGACACTCTTTCCCTACACGACGCTCTTCCGATCT
