Tag-based RNA-Seq sample preparation, for sequencing on the Illumina HiSeq

Galina Aglyamova, Eli Meyer, and Mikhail Matz

aglyamova@austin.utexas.edu, matz@utexas.edu

Updated 26 Jan 2011 to include changes for high-throughput sample preparation. This version of the protocol was optimized for working in 96-well plates.

Updated January 1, 2013 and April 2, 2013 to reflect the switch to Illumina HiSeq sequencing platform and to add qPCR-based quantification of the resulting samples.

Updated August 21, 2013 to reflect minor changes in procedure (maximizing cDNA representation).

At least 100 ng, and ideally 0.5-1 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 synthesize 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 quantified by qPCR.

The sequences of all oligonucleotides used in this protocol are provided at the end of this document.

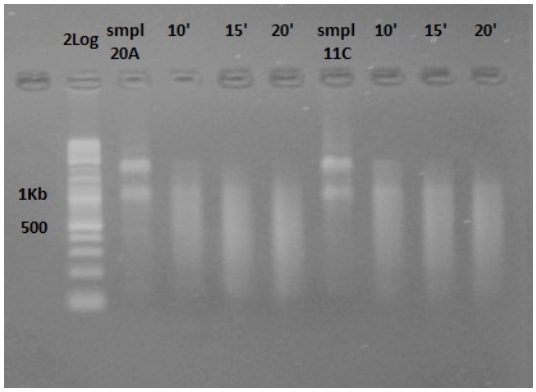
1. RNA fragmentation

   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.

   a. Aliquot 1 µg of total RNA in 10 µl of 10 mM Tris (pH 8.0). To achieve this concentration, RNA samples can be concentrated by drying in Speedvac (without heating) or by standard ethanol or LiCl precipitation. 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 to fragment the RNA. This can be most easily accomplished in a thermocycler. In our previous work the optimum time has been ~10-15 minutes.

   c. Analyze 100 ng from fragmented RNA, alongside the intact RNA from the same sample, on a standard (as for DNA) 1% agarose gel to evaluate the extent of RNA fragmentation. The smear must extend all the way up into the region where ribosomal RNA bands were, while the bands themselves should be mostly gone. In the figure on the next page, 15' result is close to the ideal, but in fact all three incubation times are acceptable.



2. First-strand cDNA synthesis
   NOTE: Although we had occasional success with amounts as low as 100 ng of fragmented RNA per reaction, we strongly recommend using 0.5-1 µg to ensure adequate representation of all transcripts.

   a. The following recipe assumes a starting volume of 10 µl (11 µl minus evaporation), so if the volume is lower than this, add additional water to achieve 10 µl.

   b. Add 1 µl of the 10 µM oligonucleotide 3ILL-30TV to each well. Incubate at 65°C for 3 minutes in a thermocycler, then transfer immediately onto ice for 2 minutes.

   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)
dNTP (10 mM ea)	1
DTT (0.1 M)	2
5X first-strand buffer	4
10 µM S-ILL-swMW (RNA oligonucleotide; stored at -80°C)	1
SMARTScribe Reverse Transcriptase (Clontech 639537)	1

   d. Add 9 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 15 minutes to inactivate the RT. Store First Strand cDNA (FS-cDNA) on ice or at -20°C until ready to proceed to the next step.

3. cDNA amplification

a. Prepare a set of master mixes for small-scale PCR tests: the control amplification (A), lacking the 5'-specific primer, should be empty. 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 2 µl of template (step 2e), so if you use a different amount of template, adjust the water accordingly.

(Volumes given in µl)

	A	B
H_2O	12.2	11.8
dNTP (2.5 mM ea)	2	2
10X PCR buffer	2	2
10 µM 5ILL oligo	0	0.4
10 µM 3ILL-30TV oligo	0.4	0.4
Titanium Taq polymerase (Clontech #639208)	0.4	0.4

b. For each of the original RNA samples, prepare two PCR tubes labeled A and B. Add 18 µl of the appropriate master mix to each tube.

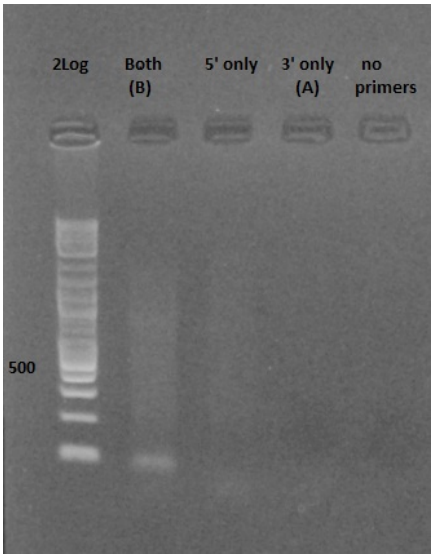
c. Add 2µl FS-cDNA (step 2e).

d. Amplify using the following profile:
95°C 5 min, (95°C 40 sec, 63°C 2 min, 72°C 1 min) X 13 cycles;
after which save 3 µl aliquots from all tubes.

e. Put the tubes back into the thermocycler; perform additional 2 cycles:
95°C 40 sec, 63°C 2 min, 72°C 1 min,
and once again save 3 µl aliquots from all tubes.

f. Repeat step (3e) two more times (until the total number of PCR cycles is 19).

g. Run the saved aliquots on a standard 1% agarose gel. The desired pattern (see figure) is a smear of cDNA (~100-1200 bp) which should be <u>just becoming visible</u> in reaction B, while nothing should be detectable in the reaction A.



NOTES:
- If you started with large amount (1 µg) of total RNA you might see a carry-over degraded RNA smear on the gel, in both A and B reactions. Do not confuse it with the PCR product! Make sure the product actually accumulates as you are adding more cycles.
- Different samples might require slightly different number of

cycles, this is OK since all the potential biases due to PCR amplification will be removed at the data analysis stage by discarding PCR duplicates.
- Very important: if a smear is not produced after 19 cycles, the representation of the cDNA is not adequate for RNA-seq; you must optimize previous stages. Ideal RNA-seq results can be obtained for samples that are amplified in 15 cycles or less, 17 cycles is OK.

h. Once the optimum number of cycles have been determined, prepare a single large-scale reaction for each cDNA sample as follows.

	(all volumes given in µl)
H_2O	32
dNTP (2.5 mM ea)	5
10X PCR buffer	5
10 µM 5ILL oligo	1
10 µM 3ILL-30TV oligo	1
Titanium Taq polymerase (Clontech #639208)	1
First-strand cDNA	5

i. Perform as in step (3d) but for the determined optimal number of cycles for each sample. Run 5 µl of the product on a gel to verify that the reaction worked.
j. Purify PCR products using PCR-clean up kit (Fermentas K0702), according to the manufacturer's instructions.
k. Quantify the purified products by OD260 (Nanodrop). The DNA concentration should be about 15-20 ng μl^{-1}.
l. Prepare 30 µl of the purified PCR products diluted to 5 ng μl^{-1} (in 10 mM tris HCl pH 8, or the elution buffer from the PCR-cleanup kit).

4. Barcoding and size selection

a. First, four test-scale PCRs are prepared for a representative subset (8-12) of samples to verify yield and specificity of the reaction. Each PCR will be using 10 ng of PCR product (2 µl from step 3l) as template.

b. Prepare a small amount of diluted barcoding oligos[1] (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.

[1] The barcoding oligos given in the table are a selection of standard Illumina TrueSeq barcodes that we used thus far; more barcode sequences can be found elsewhere. We recommend multiplexing 20-30 RNAseq samples per lane, aiming at 5-10 million raw reads per sample; so the selection of 38 barcodes given in the table should be sufficient. If ordering from IDT, order them as "ultrameres" with no purification; this seems to be the best quality-cost balance. Remember that the barcode will be read in a reverse-complement orientation compared to the sequence in the table.

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 being tested plus a small additional amount to account for pipetting error.

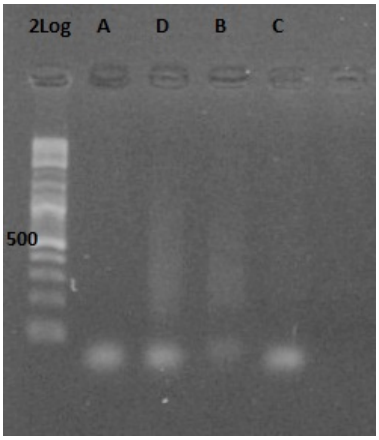
	(volumes given in µl)			
	A	B	C	D
H2O	5.8	5.6	3.8	3.6
dNTP (2.5 mM ea)	1	1	1	1
10X PCR buffer	1	1	1	1
TrueSeqMpx2n, 10 µM	0	0.2	0	0.2
Titanium Taq polymerase	0.2	0.2	0.2	0.2
Total Master mix per reaction	8	8	6	6

d. Aliquot the specified amount of master mix into each well, then add diluted (1 µM) barcode oligos to reactions C and D as specified in the above table.

e. Add 2 µl of the diluted purified cDNA (step 3l) to each of these four primer combination reactions (A-D). The total reaction volume is 10 µl.

f. Amplify using the following profile:
95°C 5 min, (95°C 40 sec, 63°C 2 min, 72°C 1 min) X 4 cycles

g. After 4 cycles check 5 µl 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 nothing is detected in reaction D, add 1-2 more cycles and check the results on a gel, if no product is visible after that, something is wrong – check your PCR reagents and/or oligos. A small amount of product in the controls can be tolerated, as long as reactions A-C show less intensity than reaction D.



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 (10 µl of the diluted purified cDNA, step 3l) in 50 µl total volume. The recipe below 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.

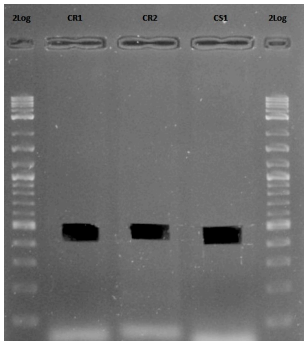
(volumes given in µl)

H2O	18
dNTP (2.5 mM ea)	5
10X PCR buffer	5
Illumina_Universal oligo (10 µM)	1
Titanium Taq polymerase	1

i. Aliquot 30 µl of master mix to each well, then add 10 µl of the appropriate barcode oligo (1µM), and 10 µl of PCR product (step 3l).

j. Amplify these reactions using the same profile and cycle number as determined above.

k. Prepare a gel for size selection. This preparative gel should consist of 2% agarose in 1X TBE buffer, with SYBR Green I nucleic acid gel staining dye (Invitrogen # S7563) added according to the manufacturers' instructions (1:10,000 dilution). Be sure to use large volume combs to allow loading of the entire 50 µl reaction +10 µl loading dye into a single well.



l. Load samples and run the gel slowly, at 100 volts cm^{-1}, for 70 minutes (or until marker bands in the 100 - 500bp size range are well separated on your gel system). Use blue-light gel illuminator to safely cut out the required size range (400-500bp). Cut only the middle of the lane, leave the edges (see picture). Slice each cut-out piece into 4-5 fragments (but not too finely) and put them into a new 0.5 ml tube.

m. Add 20 µl of nuclease-free water to the tubes containing gel slices, make sure the water and gel pieces are in contact, and incubate overnight at 4°C to let the DNA diffuse out of the gel. No further purification procedures are necessary; simply use the water eluate in the subsequent steps.

5. qPCR quantification for mixing on the same HiSeq lane

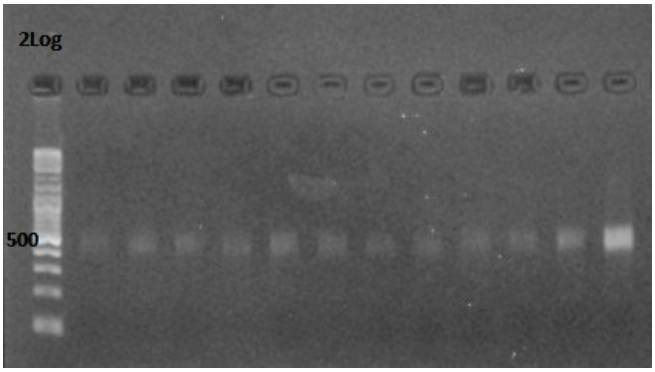
NOTE: For checking quality and quantity of eluted DNA we do two PCRs. The first one is to verify the product size on gel; it should be the same as the band we cut out and no additional products. Second, for mixing the barcoded samples together in equal proportions we perform qPCR with P7 and our custom anti-T30 short primers.

a. For quality check prepare a PCR master mix according to the following recipe. The volumes are given for a single reaction, so multiply these values by the total number of reactions plus a small additional amount to account for pipetting error.

	(volumes given in µl)
H_2O	6.4
dNTP (2.5 mM ea)	1
10X PCR buffer	1
IC2-P7 primer (10 µM)	0.2
IC1-P5 primer (10 µM)	0.2
Titanium Taq polymerase	0.2

b. Add 1 µl of gel-extracted final product DNA template (step 4m) to each reaction, for a total reaction volume of 10 µl.

c. Amplify using the following profile:
   95°C 5 min, (95°C 40 sec, 63°C 1 min, 72°C 1 min) X 12 cycles
   Run 3 µl on gel. The size of the product should match the size you aiming when cut a band for gel-extraction.



d. For qPCR quantification, prepare two dilutions of each sample (eluate from step 4m), 1/10 and 1/50, in 10 mM Tris-HCl. *(Galina prefers making three dilutions – the third one 1/250 – but theoretically, two dilutions should suffice)* Arrange the dilutions in a 96-well plate for easy pipetting. For each sample, make two identical dilution series – these will be your technical replicates. Do not simply split the same dilution series into two – we aim to replicate the process of making those dilutions.

e. Mix SYBR-based qPCR Master mix appropriate for your qPCR instrument with water and two primers, according to the recipe below, and aliquot 14 µl per reaction. Plan for two no-template-control (NTC) reactions.

	(volumes given in µl)
H_2O	6.1
SYBR Green mix	7.5
IC2-P7 primer (10 µM)	0.2
anti-T30 primer (10 µM)	0.2

f. Add 1 µl of template's dilutions from step 5d.

g. Amplify in a qPCR instrument using the following profile:
95°C 5 min, (95°C 40 sec, 63°C 1 min, 72°C 1 min) X 20 cycles
(the product should amplify around 8-12 cycles, the NTC controls should be empty).

h. Arrange the data in Excel in the form of a table with four columns: sam (sample name), lane (intended HiSeq lane), conc (DNA dilution; use 0.1 for 1/10, 0.02 for 1/50 and 0.004 for 1/250), and ct (qPCR result for this sam-conc combination). There must be at least two technical replicates for each combination of sam-conc (i.e. two rows with the same sam and conc and different ct values). If all samples are to be mixed on the same lane, enter '1' throughout the lane column. The order of columns and rows does not matter, but the names of the columns do matter (they are case sensitive).

i. Export the data from Excel as comma-separated values (.csv). Open script mix_illumina_qpcr.R in R, follow the instructions given in the comments within the script.

j. Mix samples (eluates from the gel slices) according to final mixing table produced by the script. This material is in principle ready for Illumina sequencing, except you might need to concentrate the resulting sample 2-3 fold to meet the sequencing facility requirements. In that case, we recommend mixing a larger volume of the all-barcodes mixture and concentrating it using SpeedVac.

oligo	use	Sequence, 5'-3'	notes
3ILL-30TV	cDNA synthesis and amplification	ACGTGTGCTCTTCCGATCTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTV	V=[ACG]
S-Ill-swMW	cDNA synthesis	ACCCCAUGGGGCUACACGACGCUCUUCCGAUCUNNMWGGG	RNA oligo; M=[AC], W=[AU]
5ILL	cDNA amplification	CTACACGACGCTCTTCCGATCT	
ILL-BC23	Barcoding	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT	the barcode is underlined
ILL-BC24	Barcoding	GCTACC	only the barcode
ILL-BC25	Barcoding	ATCAGT	only the barcode
ILL-BC26	Barcoding	GCTCAT	only the barcode
ILL-BC27	Barcoding	AGGAAT	only the barcode
ILL-BC28	Barcoding	CTTTTG	only the barcode
ILL-BC29	Barcoding	TAGTTG	only the barcode
ILL-BC30	Barcoding	CCGGTG	only the barcode
ILL-BC31	Barcoding	ATCGTG	only the barcode
ILL-BC32	Barcoding	TGAGTG	only the barcode
ILL-BC33	Barcoding	CGCCTG	only the barcode
ILL-BC34	Barcoding	GCCATG	only the barcode
ILL-BC35	Barcoding	AAAATG	only the barcode
ILL-BC36	Barcoding	TGTTGG	only the barcode
ILL-BC37	Barcoding	ATTCCG	only the barcode
ILL-BC79	Barcoding	ACGCGG	only the barcode
ILL-BC80	Barcoding	AGGGCG	only the barcode
ILL-BC81	Barcoding	CTGCAG	only the barcode
ILL-BC82	Barcoding	AACTTC	only the barcode
ILL-BC83	Barcoding	GGGTGC	only the barcode
ILL-BC84	Barcoding	TCCTGC	only the barcode
ILL-BC85	Barcoding	CGCGGC	only the barcode
ILL-BC86	Barcoding	ACCGCC	only the barcode
ILL-BC87	Barcoding	TAATAC	only the barcode
ILL-BC88	Barcoding	CACGTA	only the barcode
ILL-BC89	Barcoding	ATGTGA	only the barcode
ILL-BC90	Barcoding	TATAGA	only the barcode
ILL-BC91	Barcoding	TTTGCA	only the barcode
ILL-BC92	Barcoding	GTGCCA	only the barcode
ILL-BC93	Barcoding	CTAACA	only the barcode
ILL-BC94	Barcoding	ATAGAA	only the barcode
TruSeq-Mpx-2n	Barcoding	AATGATACGGCGACCACCGAAAAATACACTCTTTCCCTACACGACGCTCTTCCGAT	extends the linker at the 5' of the cDNA
IC-P7	qPCR, final check	CAAGCAGAAGACGGCATACGA	
IC-P5	Final check	AATGATACGGCGACCACCGA	
antiT30	qPCR	AAATTAGATCGGAAGAGCACAC	