

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

Galina Aglyamova, Eli Meyer, and Mikhail Matz

[aglyamova@austin.utexas.edu](mailto:aglyamova@austin.utexas.edu), [matz@utexas.edu](mailto: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 March 12, 2015 to reflect substitution of qPCR for Picogreen DNA assay for DNA quantification.*

*Updated May 11, 2015 to simplify cDNA amplification procedure.*

At least 100 ng, and ideally 0.5-1 µg of DNase-treated total RNA is required per sample. The RNA does not have to be perfectly intact but should show reasonably high-MW smear on gel. Minor DNA contamination is tolerable but we do recommend DNase treatment.

The procedure can be reasonably completed within three days:

Day 1: RNA is fragmented and used to synthesize cDNA (steps 1-2). cDNA is amplified.

Day 2: PCR product are cleaned and DNA concentrations are quantified with Picogreen DS DNA assay and sample concentrations are equalized. Short PCR (4 cycles) is performed to incorporate sample-specific barcodes. Samples are pooled, cleaned and size-selection by gel extraction is performed.

Day 3: After overnight elution from gel, the final DNA concentrations are quantified by Picogreen assay. Optional control PCR is run to confirm size range of the samples.

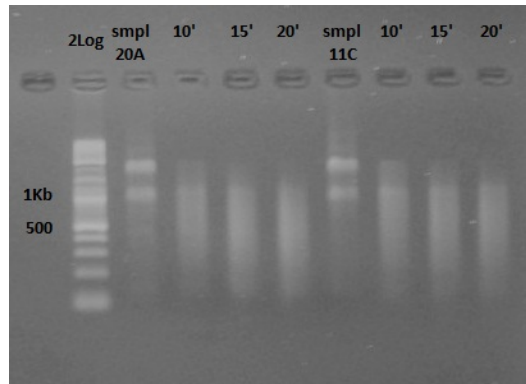
The sequences of all oligonucleotides used in this protocol are provided at the end of this document.

## 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 of 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 have 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.*

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

- 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.
- 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

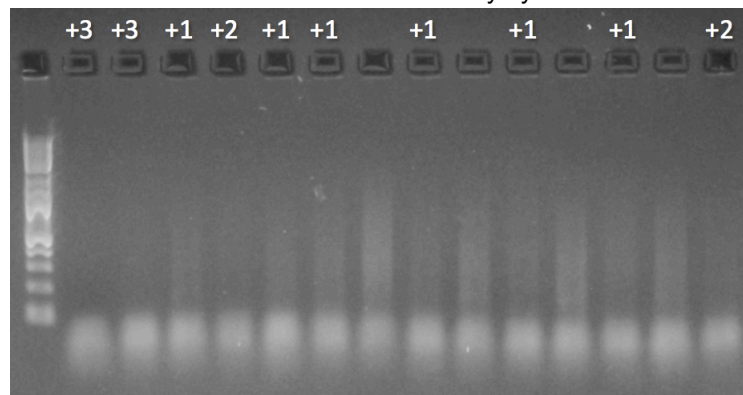
- a. Prepare PCR reactions for each cDNA sample as follows. The recipe below is for a single reaction, so multiple these values by the number of samples to be prepared plus a small additional amount for pipetting error.

	volume in $\mu$ l
H <sub>2</sub> O	32
dNTP (2.5 mM ea)	5
10X PCR buffer	5
10 $\mu$ M 5ILL oligo	1
10 $\mu$ M 3ILL-30TV oligo	1
Titanium Taq polymerase (Clontech #639208)	1
First-strand cDNA	5

94°C 5 min, (94°C 1 min, 63°C 2 min, 72°C 2 min) X 16 cycles

- b. Run 5  $\mu$ l of the product on a 2% agarose gel to verify that the reaction worked. To the reactions where the smear is visible but very faint, add 1-2 more PCR cycles (94°C 1 min, 63°C 2 min, 72°C 2 min). To the ones where the smear is not seen at all, add 3 more cycles. Run 5  $\mu$ l of these “lagging-behind” reactions on the gel again.

This is how to decide how many cycles to add:



#### NOTES:

- If you started with large amount (1  $\mu$ g) of total RNA you might see a carry-over degraded RNA smear on the gel, which can be confused with the PCR product. One way to make sure is to set up a couple of negative control reactions, lacking the 5ILL primer. If doubts remain, add one more PCR cycle to all reactions to confirm that the product actually accumulates.
- 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 visible after 19 cycles, the representation of the cDNA is not adequate for RNA-seq; you must optimize previous stages. Replacing a batch of reverse

*transcriptase or additional RNA purification, such as precipitation by adding equal volume of 5M LiCl, chilling at -20°C for 30 min and spinning at max speed for 15 min, might help. Ideal RNA-seq results are obtained for samples that are amplified quite brightly in 16 cycles.*

- c. Purify PCR products using PCR-clean up kit (Fermentas K0702), according to the manufacturer's instructions.
- d. Quantify the purified products by Quant\_IT Picogreen DS DNA Kit (Life Technologies P7589). See page 6 of this protocol for details.
- e. Prepare 20  $\mu\text{l}$  of the purified PCR products diluted to 5  $\text{ng } \mu\text{l}^{-1}$  (in 10 mM tris HCl pH 8, or the elution buffer from the PCR-cleanup kit). It's extremely important to put the same amount of template into the barcoding PCR.

#### 4. Barcoding and size selection

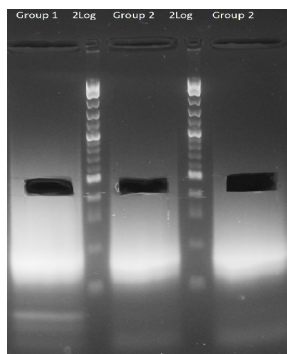
- a. Prepare the following PCR reactions. The recipe below is for a single reaction, so multiple these values by the number of samples to be prepared plus a small additional amount for pipetting error.

	volume in $\mu\text{l}$
H2O	11
dNTP (2.5 mM ea)	3
10X PCR buffer	3
* TruSeq_Un1 (10 $\mu\text{M}$ )	0.6
Titanium Taq polymerase	0.6

(\*) We use four different variations of Illumina Universal Oligo: TruSeq\_Un1, TruSeq\_Un2, TruSeq\_Un3, TruSeq\_Un4, so each sample is barcoded from both ends. It's convenient to prepare four master mixes, one for each TruSeq\_Uni oligo.

- b. Aliquot 18  $\mu\text{l}$  of master mix to each well, then add 6  $\mu\text{l}$  of the appropriate barcode oligo (1  $\mu\text{M}$ ), and 6  $\mu\text{l}$  of 5  $\text{ng } \mu\text{l}^{-1}$  cleaned PCR product (step 3I).
- c. Amplify using the following profile:  
95°C 5 min, (95°C 40 sec, 63°C 2 min, 72°C 1 min) X 4 cycles
- d. Run 5  $\mu\text{l}$  of each product on 2% agarose gel to confirm that amplification across all samples was successful and uniform (as it should be if quantification and dilutions at the previous stage were precise). If just a few of the samples are lagging behind, it is OK to add 1-2 more cycles just to those, but then make sure to run them on gel again alongside a couple of evenly-amplified samples.
- e. Pool 20  $\mu\text{l}$  from each sample in groups of 5-8 (depending on the total number of samples in the experiment). Makes sure the pools all comprise the same (or nearly same) number of samples. Concentrate the pools into 50  $\mu\text{l}$  using PCR-clean up kit (Fermentas K0702), according to the manufacturer's instructions.

- f. Prepare a gel for size selection. This preparative gel should be 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 very wide and large volume combs to allow loading of the 50  $\mu$ l mix +10  $\mu$ l loading dye into a single well.



- g. Load samples and run the gel slowly, at 5 volts  $\text{cm}^{-1}$  (i.e., at 100V if the distance between electrodes is 20cm), for 70 -90 minutes until marker bands in the 100 - 500bp size range are well separated. 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 above). Slice each cut-out piece into 4-5 fragments and put them into a new 0.5 ml tube.
- h. Add 20  $\mu$ l 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. Alternatively use QIAquick Gel Extraction Kit (QIAGEN 28704).

*NOTE: size selection can be most conveniently done on PippinPrep.*

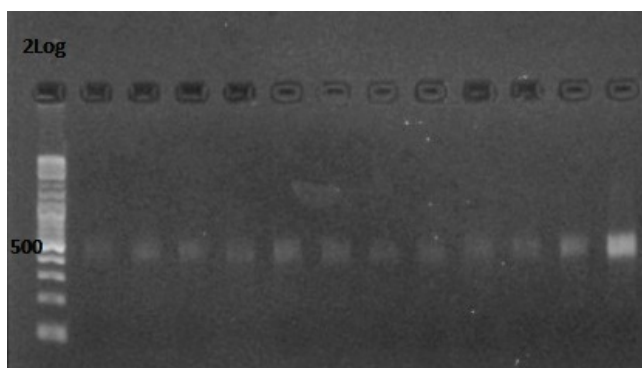
##### 5. Quantification for mixing on the same HiSeq lane

*NOTE: Checking quality of eluted DNA (steps 5a-5c) is optional. We do PCRs to verify the product size on gel; it should be the same as the band we cut out and no additional products, plus you expect to see the same intensity across all samples.*  
*For mixing the barcoded samples together in equal proportions we perform Picogreen DS DNA assay.*

- 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 $\mu\text{l}$ )	
H <sub>2</sub> O	6.4
dNTP (2.5 mM ea)	1
10X PCR buffer	1
IC2-P7 primer (10 $\mu\text{M}$ )	0.2
IC1-P5 primer (10 $\mu\text{M}$ )	0.2
Titanium Taq polymerase	0.2

- b. Add 1  $\mu\text{l}$  of gel-extracted final product DNA template (step 4h) to each reaction, for a total reaction volume of 10  $\mu\text{l}$ .
- c. Amplify using the following profile:  
95°C 5 min, (95°C 40 sec, 63°C 1 min, 72°C 1 min) X 10-12 cycles  
Run 3  $\mu\text{l}$  on gel. The size of the product should match the size you aiming when cut a band for gel-extraction.



- d. Run Quant-IT picogreen DS DNA assay (Life Technologies P7589) to determine the final concentrations of the eluted product in order to mix libraries in equal proportions.

### **PicoGreen assay Protocol:**

- 1) Place 100ul 1X TE into all first column\_wells except B1.
- 2) Add 150ul of DNA standard (@ 2ug/ml, which is the same as 2ng/ul) into B1.
- 3) Serially dilute standards by taking 50ul of B1, mixing into C1, taking 50ul of C1, mixing into D1, and so on until taking 50ul from H1 and throwing it out.
- 4) To all sample wells, add 98ul of 1X TE.
- 5) Add 2ul sample DNA to sample wells.
- 6) Mix Pico Green Master mix: 99.5ul 1XTE + 0.5ul PicoGreen for one sample. Multiply accordingly (plus 8 wells for DNA standard).
- 7) Add 100ul of master mix to all standard and sample wells, bringing up final volumes in each well to 200.
- 8) Read the fluorescence (excitation 480nm, emission 520nm). We use SpectraMax M2 plate reader and Costar assay plates 96 well, no lid, flat bottom, non-treated black with black bottom (Corning 3650) or clear bottom (Corning 3631).
- 9) Save the data into txt file, assemble the results in Excel in two-column form – well, reading - save it as comma-delimited (.csv) file. The file must contain all A1-H1 wells (blank and calibrators) plus an arbitrary number of sample wells, in any order. See file picogreen.csv as an example.
- 10) Use picogreen.R script to calculate sample concentrations (ng/ul in the original sample).

Only the first three oligos are specific for TagSeq, the rest can be substituted for standard Illumina TruSeq™ oligos.

oligo	Protocol stage	Sequence, 5'-3'	notes
<b>TagSeq-specific oligos:</b>			
3ILL-30TV	cDNA synthesis and amplification	ACGTGTGCTCTCCGATCTAATTTTTTTTTTTTTTTTTTTTTTTTTT	V=[ACG]
S-III-swMW	cDNA synthesis	ACCCCAUGGGGCUACACGACGCUUCGCAUCUNNMWGGG	RNA oligo; M=[AC], W=[AU]
5ILL	cDNA amplification	CTACACGACGCTCTCCGATCT	
<b>TruSeq oligos</b>			
ILL-BC23	Barcoding	CAAGCAGAAGACGCATACAGAGAT <u>CCACTC</u> GTGACTGGAGTTCAGACGTGTGCTCTCCGAT	the barcode is underlined
ILL-BC24	Barcoding	GCTACC	only the barcode
ILL-BC25	Barcoding	ATCAGT	only the barcode
ILL-BC26	Barcoding	GTCAT	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	AATGATACGGCGACCAACGAAAAATACACTCTTTCCCTACACGACGCTCTCCGAT	extends the linker at the 5' of the cDNA
TruSeq_Un1	Barcoding	AATGATACGGCGACCAACGAGATCTACAC <b>ATCACG</b> ACACTCTTCCCTACACGACGCTCTCCGATCT	
TruSeq_Un2	Barcoding	AATGATACGGCGACCAACGAGATCTACAC <b>ACTTGA</b> ACACTCTTCCCTACACGACGCTCTCCGATCT	
TruSeq_Un3	Barcoding	AATGATACGGCGACCAACGAGATCTACAC <b>TAGCTT</b> ACACTCTTCCCTACACGACGCTCTCCGATCT	
TruSeq_Un4	Barcoding	AATGATACGGCGACCAACGAGATCTACAC <b>GGGTAC</b> ACACTCTTCCCTACACGACGCTCTCCGATCT	
IC-P7	qPCR, final check	CAAGCAGAAGACGCATACGA	
IC-P5	Final check	AATGATACGGCGACCAACGA	