TagSeq library preparation (version August 2018)

Galina Aglyamova, Eli Meyer and Mikhail Matz, University of Texas at Austin

matz@utexas.edu

RECENT IMPORTANT UPDATES:

September 20, 2016 to simplify protocol and enable micro-amounts (down to 10 ng of total RNA).

December 21, 2017: change to mixture of three RNA oligos for cDNA synthesis; this protocol now works on HiSeq 4000.

May 29, 2018: RNA oligos for cDNA synthesis are replaced by DNA versions with only three 3'-terminal ribo-G bases. This makes them much cheaper (no purification needed) and more stable.

July 13, 2018: Edits by C. Kenkel to clarify protocol steps, correct volume errors and add additional trouble-shooting information.

The procedure now works with as little as 10 ng total RNA per sample at the start (1 ng also works but results are substantially less consistent across replicates). However, this is not a substitute for poor extractions. If you are not tissue limited, we recommend aiming for 1 microgram starting RNA to maximize representation of transcripts of sufficiently high quality (crisp rRNA bands). If samples are limited or precious, the RNA does not have to be perfectly intact but should show reasonably high-MW smear on gel or Bioanalyzer. DNA contamination is OK although we recommend DNAse treatment when the amount of material is not limiting.

The procedure can be reasonably completed within two days:

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

Day 2: Samples are pooled, cleaned and size-selection by gel extraction (or Pippin-prep) is performed. The final DNA concentrations are quantified by Picogreen assay or qPCR. Optional control PCR is run to confirm size range of the samples.

The sequences of all oligonucleotides used in this protocol are provided in Appendix I and in tagseq_oligo_order.xls file (thanks to Sarah Davies for putting it together!).

Extract RNA with RNAqueous micro or BioRad's Aurum Total RNA Mini Kit (note: if using Aurum, extend on-column DNAse treatment time to 1 hour); elute by passing the same 20 ul through the column twice.

1. First-strand cDNA synthesis
    a. Mix:

(all volumes given in µl)	
Total RNA	10-12
dNTP (10 mM ea)	1
DTT (20 mM)	2
5X first-strand buffer	4
3ILL-30TV, 10uM	1

Incubate at 70°C for 10 minutes in a thermocycler, then transfer onto ice for 2 minutes.

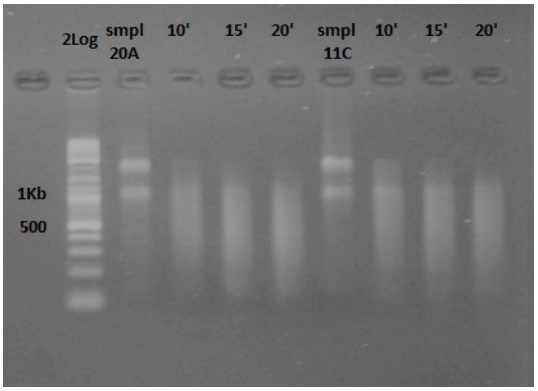
b. Add 1 ul of equal-parts mixture of three RNA "sw" oligos (10 uM) and 1 ul of SMARTScribe Reverse Transcriptase (Clontech 639537).

c. Incubate in a thermocycler for 1 hour at 42°C.

d. Incubate at 65°C for 15 minutes to inactivate the RT.

e. Purify the product using AMPpure beads (Agilent Technologies) according to manufacturer's instructions; elute in 12 ul of water (Appendix II).

NOTE 1: For new species, it is recommended that the exact degradation time be trialed with a few samples to verify that 10 min does not over-degrade samples. For example, we have found that sponge RNA degrades on the order of 3-5 min. For samples in which yield is sufficiently high, prepare 1-2 cDNA synthesis reactions using only the DTT and 5X first-strand buffer and replacing primers with ultrapure H_2O. Sample 2ul of your degradation at different times (e.g. 5, 10 and 15 min) and then run 100 ng intact sample alongside these trials on a 1% gel with a 100bp ladder to determine the time that achieves an RNA smear centered around 500 bp.



NOTE 2: it makes sense to also perform a (–)RT reaction to verify that DNA contamination does not result in any amplification at the next stage. Usually there is none, but this might not hold for all species (for example if there is a particularly strong chance match between our amplification primers and the genome) so we also recommend doing (-)RT control for each new species, at least once.

2. <u>cDNA amplification</u>

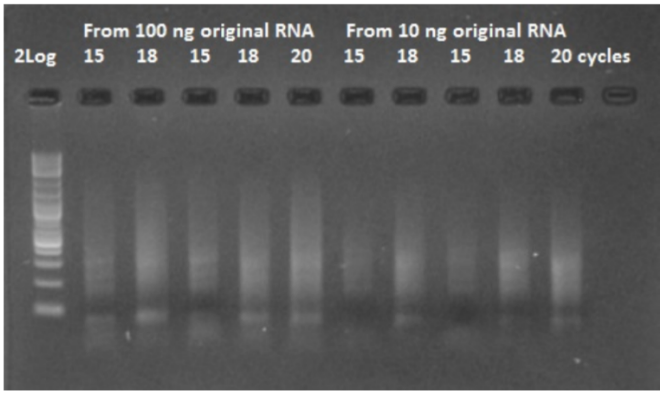
a. Prepare PCR reactions for each cDNA sample as follows. The recipe below is for a single reaction, so multiple these values by the number of samples to be prepared plus a small additional amount for pipetting error.

(all volumes given in µl)

H2O	4.8
dNTP (2.5 mM ea)	2
10X PCR Buffer	2
10 µM 5ILL oligo	0.5
10 µM 3ILL-30TV	0.5
<u>KlenTaq (DNA Polymerase Technology, #100)</u>	<u>0.2</u>
First-strand cDNA	10

94°C 5 min, (94°C 1 min, 63°C 2 min, 72°C 2 min) X 18 cycles
(this reaction can be left overnight at room temperature in the thermocycler

b. Run 2 µl of the product on a 1% agarose gel to verify that the reaction worked.



NOTES:

- If you started with large amount (~1 µg) 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 – example given below:

(Volumes given in µl)

	(-)	(+)
H_2O	5.6	5.4
dNTP (2.5 mM ea)	1	1
10X PCR buffer	1	1
10 µM 5ILL oligo	0	0.2
10 µM 3ILL-30TV oligo	0.2	0.2
KlenTaq	0.2	0.2

a. For each of the original RNA samples, prepare two test PCR tubes labeled A and B. Add 8 µl of the appropriate master mix to each tube.

b. Add 2µl FS-cDNA and run for 18 cycles as in (2a) above. Load 5 ul on gel. Smear should be present in (+) only. Add cycles to confirm product accumulation.

- 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. Do not worry about adding more PCR cycles – the fraction of duplicates does not change after the first 3-4 cycles, see Appendix V.

c. Purify PCR products using AMPure beads (Agilent Technologies, Appendix II), according to the manufacturer's instructions; elute in 20 ul of water.

d. Quantify the purified products by Quant_IT Picogreen DS DNA Kit (Life Technologies P7589). See Appendix III for details.

- *Note: yields may be lower than the recommended* 5 ng µl $^{-1}$. *We have seen successful libraries prepared from samples with as low as* 0.5 ng µl $^{-1}$. *In the event of lower concentrations, dilute all samples to the SAME concentration and increase the amount of template added to the barcode PCR (by reducing the volume of water) accordingly.*

e. Prepare 20 µl of the purified PCR products diluted to exactly 5 ng µl $^{-1}$ in water. It's extremely important to put the same amount of template into the barcoding PCR.

3. Barcoding and size selection

a. Prepare the following PCR reactions. NOTE: this reaction is for those using full dual-indexing. The recipe below is for a single reaction PER "UN" barcode, so multiple these values by the number of samples to be prepared for each "UN" plus a small additional amount for pipetting error (e.g. for a full 96-well plate, 8 UN x 12 BC works nicely; it's also possible to reverse these mixes and use "BC" overlaid with "UN"; see Appendix I).

(all volumes given in µl)

H2O	5.8
dNTP (2.5 mM ea)	3
10X PCR Buffer	3
1 µM "UN" Barcode	6
KlenTaq	0.2

b. Aliquot 18 µl of master mix to each well, then add 6 µl of the appropriate "BC" oligo (1µM), and 6 µl of 5 ng µl^{-1} cleaned PCR product (step 2e).

c. Amplify using the following profile:

95°C 5 min, (95°C 30 sec, 53°C 30 sec, 68°C 30 sec) X 4 cycles

d. Run 5 µl of each product on 1% 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).

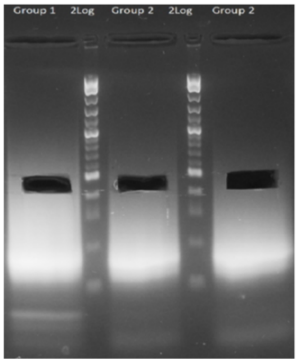
e. Pool 20 µl from each sample in groups of 5-8 (depending on the total number of samples in the experiment). Makes sure the pools all comprise the same (or nearly same) number of samples. Concentrate the pools into 30 µl using AMPure beads, according to the

manufacturer's instructions.

- expected DNA concentration in the cleaned pools is 15-30 ng/ul

- NOTE: if your samples are more concentrated, you may encounter difficulties during gel extraction due to overloading of wells. Reduce the number of samples per pool in this case.

f. Prepare a gel for size selection. This preparative gel should be 1% agarose in 1X TBE or 1X SB buffer, with SYBR Green I nucleic acid gel staining dye (Invitrogen # S7563) added according to the manufacturers' instructions (1:10,000 dilution, or 5ul per 50ml). Be sure to use large volume combs to allow loading of the 30 µl mix +5 µl loading dye into a single well. Use 100bp ladder to visualize.



g. Load samples and run the gel slowly, at 5 volts 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 1.5 ml tube.

h. Use QIAquick Gel Extraction Kit (QIAGEN 28704) to extract DNA. Or "Freeze and Squeeze" as follows:

- – Add 20 ul ultrapure H2O to each tube.
- – Centrifuge tubes 1 min at high speed to bring gel into contact with water.
- Optional: incubate at 4C O/N to maximize yield. Otherwise, proceed directly to
- freezer.
- – Freeze at -80C for at least 30 minutes
- – Centrifuge at maximum speed for 10 minutes
- – Press gel slice against side of tube using pipette tip, and withdraw supernatant (at least 30 ul should be recovered). If less than 30 ul is accessible, repeat centrifugation. Transfer the supernatant to a new PCR tube or plate and proceed with quantification.

NOTE: size selection can be most conveniently done on PippinPrep or BluePippin, with higher yields. Set the size selection window to 350-550 bases.

4. Final Quality Check and Mixing Libraries for Sequencing

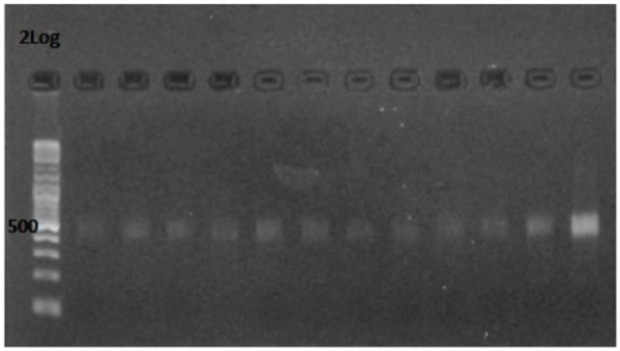
   a. For quality check prepare a PCR master mix according to the following recipe. The volumes are given for a single reaction, so multiply these values by the total number of reactions plus a small additional amount to account for pipetting error.

(all volumes given in µl)	
H2O	6.4
dNTP (2.5 mM ea)	1
10X PCR Buffer	1
IC2-P7 primer (10 µM)	0.2
IC2-P5 primer (10 µM)	0.2
KlenTaq	0.2

   b. Add 1 µl of gel-extracted final product DNA template (step 3h) 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 10-12 cycles
      Run 3 µl on gel. The size of the product should match the size selected in the gel-extraction step.



   d. Run Quant-IT picogreen DS DNA assay (Life Technologies P7589) or qPCR to determine the final concentrations of the eluted product in order to mix libraries in equal proportions. See Appendix III for picogreen protocol. A qPCR protocol is below.

e. Protocol for qPCR based quantitation (recommended if Tagseq preps are to be mixed with other library types, e.g. 2bRAD)

- Prepare a 1:100 dilution by combining 2 ul of the eluted library (Step 3h) with 198 ul NFW
- Prepare sufficient master mix for the number of samples to be quantified (In experienced hands 1-2 reactions per sample are sufficient, but users new to qPCR may benefit from additional replication).

  (all volumes given in µl)

H2O	4.3
2x SYBR master mix (e.g. Bioline #BIO-92005)	7.5
IC2-P7 primer (10 µM)	0.6
IC2-P5 primer (10 µM)	0.6

  Add 13 ul master mix to each well, followed by 2ul of the diluted template.

- Conduct qPCR and calculate Ct for each sample.
- To determine volumes of each library for the combined pool:
  1. Rank samples from lowest to highest Ct and identify reference sample (sample with the highest Ct)
  2. Calculate the proportion of each library to sequence as:
     a. PL = 2 ^ [Ct(sample) – Ct(reference)]
  3. Calculate the volume of each library to use as:
     a. V=PL * 60 ul

  Note 1: If you've chosen a reference with a VERY high Ct (suggesting a failed library prep) relative to the others, very low volumes (<2ul) may be calculated at this step. If so, choose the next sample (i.e. the next lowest Ct) as a reference instead, and continue adjusting until reasonably high volumes are calculated for the majority of samples.

  Note 2: As a rough rule of thumb, the pool of combined samples used for a single lane of sequencing should be AT LEAST 200-500 ul at this stage, and may be substantially higher

- Combine libraries using the volumes calculated from qPCR to produce a pool for sequencing.
- Illumina sequencing typically requires templates in <= 20 ul volume at >= 2nM concentration. The pooled libraries produced above are typically too dilute for sequencing. Use the Ampure beads to purify and concentrate these libraries. Your final concentration should be at least 0.65 ng µl-1 to have enough material for sequencing.

NOTE: If mixing library types (tag-seq and 2bRAD), put two-fold higher amount of TagSeq libraries into your pools to account for template bias.

Appendix I: Oligonucleotides

(also see tagseq_oligo_order.xls file, or KenkelLab_StartupPrimers.xls)

All long oligos should be ordered from IDT as “ultrameres”. No purification is necessary.

Name	Sequence	Notes
3ILL-30TV	ACGTGTGCTCTTCCGATCTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTV	V=[ACG]
5-III-swMW	ACCCCATGGGGCTACACGACGCTCTTCCGATCTNNMWrGrGrG	DNA oligo with three 3’-terminal ribo-G bases; M=[AC], W=[AU]
5-III-swUG	ACCCCATGGGGCTACACGACGCTCTTCCGATCTNNTGCMWrGrGrG	DNA oligo with three 3’-terminal ribo-G bases
5-III-swGC	ACCCCATGGGGCTACACGACGCTCTTCCGATCTNNGCWTCHMWrGrGrG	DNA oligo with three 3’-terminal ribo-G bases
5ILL	CTACACGACGCTCTTCCGATCT	
ILL-BC23	CAAGCAGAAGACGGCATACGAGAT**CCACTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT	Example. Barcode is colored red.
ILL-UN1	AATGATACGGCGACCACCGAGATCTACAC**ATCACG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT	Example. Barcode is colored red. AKA ‘HT’ in 2bRAD
ILL-P5	AATGATACGGCGACCACCGA	AKA Lib1 in 2bRAD
ILL-P7	CAAGCAGAAGACGGCATACGA	AKA Lib2 in 2bRAD

Appendix II: AMPpure Bead Protocol Overview

Note: The initial investment in beads and the SPRI magnetic plate is expensive, but if you are planning to prepare many libraries, it is worth it as the cost per sample for ALL the clean-up steps listed in this protocol equates to ~$2.25, which is the typical cost per sample per clean up for other column-based kits. Though it is also possible to use these kits if it better fits your current budget.

1) Add 1.8x volume AMPpure beads to each sample and mix by vortexing for 30s
2) Incubate at room temperature for 5 min
3) Place tubes into magnetic plate and wait 10 minutes for beads to separate from solution
4) Aspirate cleared supernatant from the reaction tubes and discard
5) Dispense 200ul of 70% EtOH into each well and incubate at room temperature for at least 30s. Aspirate out the EtOH and discard
6) Repeat step 5 for a total of two EtOH washes
7) Let the tubes air dry for 10-20 minutes at room temperature
8) Elute in desired volume, vortex for 30s to mix
9) Place tubes back onto magnetic plate and pipette out supernatant – this now contains your cleaned DNA.

Appendix III: PicoGreen assay Protocol

1) Calculate out amount of 20X TE needed to make sufficient 1X TE for all rxns
    a. 197.5ul * N samples
    b. 1643ul needed for standard curve (Lambda dilution plus Pico dilution)
    c. Example: For full 96-well, 19,023ul 1X TE is needed. Round up to 19,100 ul for error. Use C1V1 = C2V2
    20X TE (X ul) = 1X TE (19,100 ul)
    => Add 955 ul 20X TE to 18,145 ul clean H_2O

2) Place 100ul 1X TE into all first column wells except B1.

3) Make Lambda DNA standard @2ug/ml by adding 3ul of concentrate to 147ul 1X TE

4) Add 150ul of DNA standard (@ 2ug/ml, which is the same as 2ng/ul) into B1.

5) 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.

6) To all sample wells, add 98ul of 1X TE.

7) Add 2ul sample DNA to sample wells.

8) Mix Pico Green Master mix: 99.5ul 1XTE + 0.5ul PicoGreen for one sample. Multiply accordingly (plus 8 wells for DNA standard).

9) Add 100ul of master mix to all standard and sample wells, bringing up final volumes in each well to 200.

10) Read the fluorescence (excitation 485nm, emission 538nm). We use Costar assay plates 96 well, no lid, flat bottom, non-treated black with black bottom (Corning 3650) or clear bottom (Corning 3631).

    a. Note: it's also possible to quantify in a 384-well using the picogreen assay, just reduce all volumes to 25% of those listed above (max 50ul per well).

11) 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.

12) Use picogreen.R script to calculate sample concentrations (ng/ul in the original sample).

Appendix IV: Gel electrophoresis

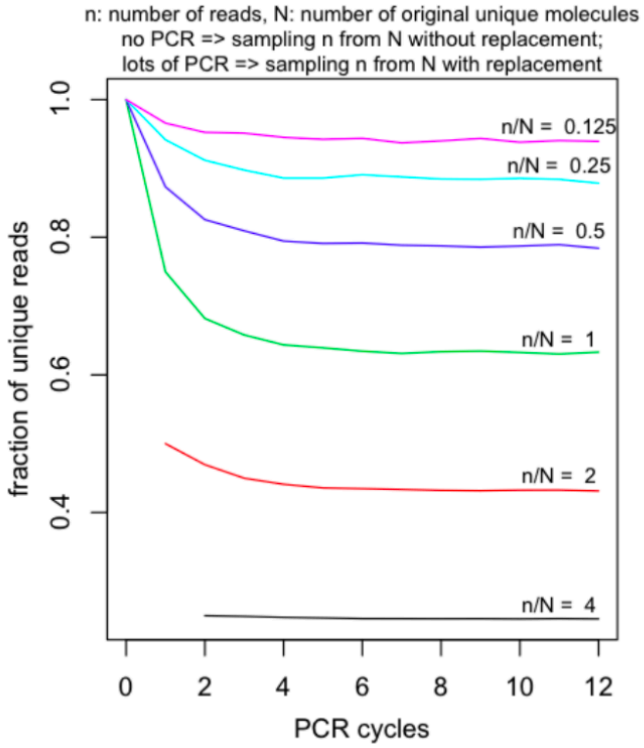
We use a sodium borate buffer with added EDTA (SBE). This buffer is cheap and can be run at higher voltages with better resolution than the usual Tris based buffers. You can make the 20X buffer as shown at https://openwetware.org/wiki/SB

We add EDTA to 20mM (in the 20X solution)

We also use ethidium bromide at 0.1 ug ml -1 for routine gel checks, but SYBR is used for the final gel excision as it can be viewed under blue light. Note that ethidium will illuminate under blue light as well, but it's not the optimal wavelength so band intensity will be lighter than when viewed under UV light.

Appendix V: PCR Duplicates

Perhaps counter-intuitively, the fraction of PCR duplicates among reads depends predominantly on the ratio between the number of reads (n) and the number of original unique molecules (N) and does not increase with additional PCR cycles beyond the first 3-4:



This is because with lots of PCR cycles the fraction of duplicates simply converges on the value expected in a sample of n from N with replacement.