

## TagSeq library preparation (version December 2017)

Galina Aglyamova and Mikhail Matz, University of Texas at Austin

[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 : substitution of qPCR for Picogreen DNA assay for DNA quantification.*

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

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

*Updated February 8, 2017: small change to barcoding amplification protocol*

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

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). 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. 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 II and in [tagseq\\_oligo\\_order.xls](#) file (thanks to Sarah Davies for putting it together!).

---

Extract RNA with RNAqueous micro; elute by passing the same 12 ul through the column twice.

### 1. First-strand cDNA synthesis

#### a. Mix:

(all volumes given in  $\mu$ l)

| Total RNA              | 10-12 |
|------------------------|-------|
| dNTP (10 mM ea)        | 1     |
| DTT (0.1 M)            | 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.

*NOTE: 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 recommend doing (-)RT control for each new species, at least once.*

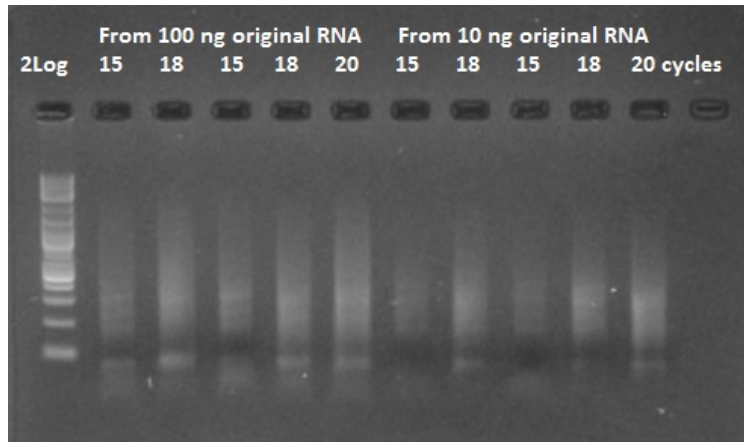
### 3. cDNA amplification

- 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 µl |
|---|--------------|
| H <sub>2</sub> O                          | 4.5          |
| dNTP (2.5 mM ea)                          | 2            |
| 10X PCR buffer                            | 2            |
| 10 µM 5ILL oligo                          | 0.5          |
| 10 µM 3ILL-30TV oligo                     | 0.5          |
| KlenTaq (DNA Polymerase Technology, #100) | 0.2          |
| 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 2% 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.
- 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 III.

- c. Purify PCR products using AMPure beads (Agilent Technologies), according to the manufacturer's instructions; elute in 20 µl of water.
- d. Quantify the purified products by Quant\_IT Picogreen DS DNA Kit (Life Technologies P7589). See Appendix I for details.
  - expected yield is about 10 ng/ul
- e. Prepare 20 µl of the purified PCR products diluted to exactly 5 ng µl<sup>-1</sup> in water. It's extremely important to put the same amount of template into the barcoding PCR.

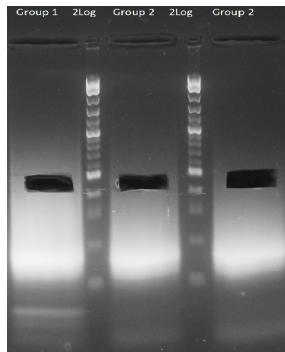
**4. Barcoding and size selection**

- 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 µl |
|----------------------|--------------|
| H2O                  | 11           |
| dNTP (2.5 mM ea)     | 3            |
| 10X PCR buffer       | 3            |
| * TruSeq_Un1 (10 µM) | 0.6          |
| KlenTaq              | 0.2          |

(\*) 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 ng  $\mu\text{l}^{-1}$  cleaned PCR product (step 3I).
- 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  $\mu\text{l}$  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).
- 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 60  $\mu\text{l}$  using AMPure beads, according to the manufacturer's instructions.  
*- expected DNA concentration in the cleaned pools is 15-30 ng/ $\mu\text{l}$*
- 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 30  $\mu\text{l}$  mix +5  $\mu\text{l}$  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. Use QIAquick Gel Extraction Kit (QIAGEN 28704) to extract DNA.

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

5. Quantification for mixing on the same HiSeq lane

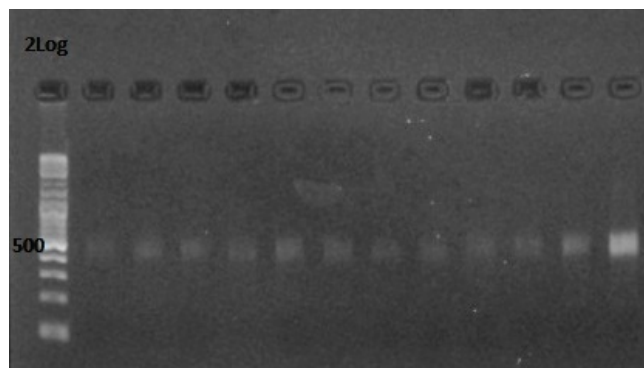
- a. Run Quant-IT picogreen DS DNA assay (Life Technologies P7589) to determine the final concentrations of the eluted product in order to mix libraries in equal proportions. See Appendix I for detailed protocol.

*The following steps are optional, if you want to be extra sure that the correct size range was efficiently extracted:*

- b. 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 |
| KlenTaq                           | 0.2 |

- c. Add 1  $\mu\text{l}$  of gel-extracted final product DNA template (step 4h) to each reaction, for a total reaction volume of 10  $\mu\text{l}$ .
- d. 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.



**NOTE: Put two-fold higher amount of TagSeq libraries on Illumina runs then suggested by P5-P7 quantitative PCR.**

## Appendix I: 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).

## Appendix II: Oligonucleotides

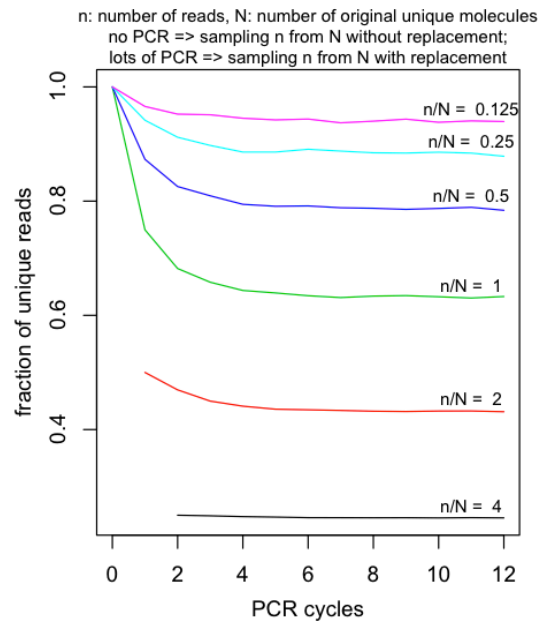
(also see [tagseq\\_oligo\\_order.xls](#) file that can be submitted to IDT)

All the long oligos should be ordered from IDT as “ultrameres”. No purification is necessary except RNA oligos for cDNA synthesis (RNAse-free HPLC for these).

| oligo      | Protocol stage                   | Sequence, 5'-3'   | notes                                |
|------------|----------------------------------|---|--------------------------------------|
| 3ILL-30TV  | cDNA synthesis and amplification | ACGTGTGCTCTCCGATCTAATTTTTTTTTTTTTTTTTTTTTTTTTT                              | V=[ACG]                              |
| 5-III-swMW | cDNA synthesis                   | ACCCCAUGGGGCUACACGACGCUUCCGAUCUNNMWGGG                                      | RNA oligo; M=[AC], W=[AU]            |
| 5-III-swUG | cDNA synthesis                   | ACCCCAUGGGGCUACACGACGCUUCCGAUCUNNUGCMWGGG                                   | RNA oligo; M=[AC], W=[AU]            |
| 5-III-swGC | cDNA synthesis                   | ACCCCAUGGGGCUACACGACGCUUCCGAUCUNNGWUCHMWGGG                                 | RNA oligo; H = [ACT], M=[AC], W=[AU] |
| 5ILL       | cDNA amplification               | CTACACGACGCTCTCCGATCT   |                                      |
| ILL-BC23   | Barcoding                        | CAAGCAGAAGACGGCATAACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCTCTCCGAT              | 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                        | ATTCGG  | only the barcode                     |
| ILL-BC38   | Barcoding                        | AGCTAG  | 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                     |
| TruSeq_Un1 | Barcoding                        | AATGATACGGCGACCACCGAGATCTACAC <b>ATCACG</b> ACACTCTTCCCTACACGACGCTCTCCGATCT |                                      |
| TruSeq_Un2 | Barcoding                        | AATGATACGGCGACCACCGAGATCTACAC <b>ACTTGA</b> ACACTCTTCCCTACACGACGCTCTCCGATCT |                                      |
| TruSeq_Un3 | Barcoding                        | AATGATACGGCGACCACCGAGATCTACAC <b>TAGCTT</b> ACACTCTTCCCTACACGACGCTCTCCGATCT |                                      |
| TruSeq_Un4 | Barcoding                        | AATGATACGGCGACCACCGAGATCTACAC <b>GGCTAC</b> ACACTCTTCCCTACACGACGCTCTCCGATCT |                                      |
| IC-P7      | qPCR, final check                | CAAGCAGAAGACGGCATAACGA  |                                      |
| IC-P5      | qPCR, Final check                | AATGATACGGCGACCACCGA  |                                      |

### Appendix III: PCR duplicates

Perhaps counter-intuitively, the fraction of PCR duplicates among reads depends predominantly on the ratio between 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.