SMART protocol for RNA-seq – January 2015

1. Poly(A) enrichment

a. Aliquot 15 µL of oligo(dT) beads (NEB).

> [NOTE – this amount of the NEB beads suffices for quantities of total RNA below 10 µg. We have not explored larger quantities of RNA, as this seems exorbitant. Also, we have used as little as 100 ng of total RNA with good results. Lower quantities should work as well, but one should keep in mind that non-specific binding of RNA to beads may reduce the recovery of poly(A) RNA. Also, it is probable that the starting amounts of poly(A) RNA may fall below the affinities of RT for template DNA if one starts with less than 100 ng of total RNA. This needs to be accounted for in the following steps, possibly by extending RT times and perhaps peridiocally adding more enzyme.]

b. Wash the beads twice with 100µL of Binding Buffer (20mM Tris-HCl pH 7.5, 1.0M LiCl and 2mM EDTA), and remove the supernatant.
c. Resuspend the beads in 50 µl of Binding Buffer.
d. Bring RNA to 50 µL with RNAse-free water. Heat to 65°C, cool on ice, then add to the washed beads. Incubate at room temperature for 5 min.
e. Collect beads using the magnetic stand, discard supernatant.
f. Wash the beads twice with 100µL of Washing Buffer B (10mM Tris-HCl PH 7.5, 0.15M LiCl, 1mM EDTA).
g. Remove the supernatant from the beads, add 15 µL of 10mM Tris-HCl (this depends on how much RNA is used, and how many samples the polyA RNA is going to be used for) and heat the beads at 80^0C for 3 minutes to elute mRNA.
h. Remove beads with the magnetic stand, save the supernatant.

> [Note that it may be advisable to repeat this, to remove as much rRNA as possible. However, this is still an open issue, since organellar and stable RNAs are known to be polyadenylated, which means repeated rounds of poly(A) enrichment yield reducing returns.
>
> Also, since we use NEB oligo-dT beads, it should be OK to follow NEB's protocol for poly(A) enrichment instead of the one I described in the preceding. NEB sells kits as well as just the beads, and some may prefer to use the ready-made solutions and protocol.]

2. RNA fragmentation, cDNA synthesis, and clean-up

a. mix 14.5 µL RNA, 1 µL (=100 pmol) RT primer + 5 µL 5X 1^{st} strand buffer.
b. heat to 95 °C for 2 min, chill.

> [NOTE – the high temperature in the RT buffer is a suitable and efficient proxy for stand-alone RNA fragmentation kits or systems, such as is sold by

NEB. By folding fragmentation directly into the RT reactions, the process is streamlined.]

c. **Immediately** add (while cold!):
   2.5 µL 10X dNTPs for reverse transcription reactions
   1 µL 100 mM DTT
   0.5 µL RNase Inhibitor
   1 µL SMARTSCRIBE
d. Mix and incubate for 120 min at 42°C.
e. Add 1 µL of the strand-switching primer (SMART7.5) and an additional 1 µL of SMARTSCRIBE, mix and incubate for an additional 120 min at 42°C.
f. Heat to 70 °C for 5 min.
g. Add 20 µL AMPure beads that have been completely mixed and brought to room temperature. Incubate for 8 min at room temperature.
h. Separate beads using the magnet stand (this may take a minute or two, because of the viscosity of the bead solution), remove and discard the supernatant.
i. Add 200 µL **fresh** 80% ethanol, mix, and collect the beads. Remove and discard the supernatant. Repeat the wash.
j. Air dry the washed beads for 5 min at room temperature.
k. Add 25 µL 10 mM Tris HCl, pH 7.5. Mix, and then collect the beads with the magnetic stand. Remove the supernatant to a new 500 µL thin-walled microcentrifuge tube.
l. Repeat steps g-k.

[The two-step process for RT and strand-switching was arrived at empirically in my lab, with guidance from some prior research publications. For us, the goal has been to develop procedures that can be used with sub-microgram quantities of RNA. I reasoned that a limiting factor would be the performance of RT, which is likely working below the Km's for template. The easiest way to address this is to extend incubation times, hence the long reaction times. While we have not tried it for RNA-Seq, by adding an additional step of RT addition and 2 hr incubation (in other words, adding just more RT at step (e) and incubating for 2 hr, then adding even more RT and the strand-switching primer), we have been able to make poly(A) site tags from as little as 5-10 ng of total RNA. It is possible that other permutations may yield acceptable results.]

3. PCR amplifications

Set up Phire reactions using the PE-PCR1 and PE-PCR2 primers. The annealing temp is 60°C (**for 15 sec**), and the extension time is 60 sec. *(In principle, the extension time can be decreased, perhaps to as little as 30 sec; I use 60 sec to make sure every amplicon is completed, so that no odd PCR strand-switching artifacts arise later due to the buildup of truncated products.)* As far as cycle # is concern, a bit of range finding is usually needed. With 1 µg of total Arabidopsis RNA, 15 cycles usually works fine for this. Some members of my lab have done as few as 9 cycles here. At the other

end of the spectrum, we have made libraries after as many as 24 cycles of amplification. As a rule of thumb, we try to avoid libraries involving more than 18 or 21 cycles, but sometimes this is unavoidable. Cycle numbers greater than 18 should not be needed if starting with 100 ng or more of total RNA.

Regardless of these considerations, the goal is a smear between 200 and 500 bp; if successful, amplifications can be scaled up and tags purified using the same AMPure protocol. This could be the final sequencing sample. Alternatively (see notes 2 and 3 below), a narrower range of sizes can be gel-purified; depending on the gels and product yields, the gel-purified sample may be sequenced directly or re-amplified to yield enough product for sequencing.

Notes: 1. We use Phire routinely for this step, because of a combination of the excellent and reliable performance of the enzyme and its relatively low cost (compared with some other hot-start enzymes). On occasion, we have used other Taq polymerases successfully, including "home-made" enzyme. The choice to include Phire here is intended to provide some measure of consistency; different enzymes can probably be used, but amplification conditions need to be optimized.

2. Perhaps the most unsettling aspect of this is the fact that, often, one cannot know about library quality until the sequencing results are returned – this means one runs the risk of spending money on poor libraries. In my lab's experience, gel-purification and re-amplification is a good quality-control check for libraries. The most persistent problem with RNA-seq libraries is the presence of short amplification products that arise due to the direct copying of the strand-switching primer by RT, using the RT primer as a primer for reverse transcription. These can easily over-amplify and give smears that can be mistaken for good libraries. When these artifacts are gel-purified and re-amplified, the results are not a narrow band corresponding to the region excised from the gel, but a long smear extending from 100 bp or so through the 1000+ bp size range.

3. My experience, after working with many labs, is that a combination of lot-to-lot variability with the AMPure beads, variability can sometimes lead to libraries that are enriched for short inserts. The easiest way to avoid this is to gel-purify libraries (or pools) before sequencing.

Primers

>RT-PE6 series (underlined bases "xxx" are where bar codes may be added)
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNxxxNNNNNN

[Note – this is the primer design that we have used to obtain libraries with good coverage and reasonably random distributions of tags. It should be pointed out that, with at least two (of more than 20) sequencing runs we have done, the first base of the run was called "N" on almost all sequences. For this reason, when we assign bar codes and design pooling strategies, we work with the assumption that the second and third "x"'s will be the informative positions. This is why we include an "N" at the very beginning of the sequencing region.

The optimal outcome is to have perfectly random priming via the NNNNNN sequence – this will be reflected in a random distribution of reads through a transcriptome. Of course, with this as well as other methods, perfect randomization is not likely; for this procedure, the nature of the bar code will probably bias priming position. We have not explored this to any extent, but I am confident that the distribution and coverage we get is comparable to other library methods. The best hedge against possible bar-code biases is replication with different bar codes.]

>SMART7.5 ("+G" denotes a locked nucleic acid)
CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGG+G

[This primer is designed to anneal to the expected CCC overhang added by the Clontech RT enzyme to the 3' end of the first strand cDNA, and thereafter serve as a template for extension of the cDNA. The Zhu paper describes this, as do various Clontech manuals.]

>PE-PCR1
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

>PE-PCR2
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

[These are the adapters needed for Illumina sequencing. Note that some of these sequences are present in the RT and SMART primers.]

References (some of the papers that guide development of the method)

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques 30 (4):892-897.

Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature methods 10 (11):1096-1098. doi:10.1038/nmeth.2639.

Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. Nature protocols 9 (1):171-181. doi:10.1038/nprot.2014.006.

Pinto FL, Lindblad P (2010) A guide for in-house design of template-switch-based 5' rapid amplification of cDNA ends systems. Analytical biochemistry 397 (2):227-232. doi:10.1016/j.ab.2009.10.022.