Preparing sequencing libraries for gene expression profiling using 3’ cDNA tag sequencing (Tag-Seq).

Last updated 26 Aug 2016

Overview

- RNA is first fragmented to expose 3’ ends of transcripts and standardize template lengths across genes.
- cDNA is synthesized from the 3’ ends of transcripts using oligo-dT primers and template switching activity to attach known sequences at each end of the fragment.
- Constructs are amplified and purified, then barcodes and sequencing primer binding sites incorporated through PCR.
- Finally, libraries are size-selected using gel purification to select a uniform size distribution appropriate for sequencing on the Illumina platform.

Before you begin

1. The protocol begins with purified, intact RNA. 100-200 ng of RNA from each sample should be evaluated on a gel prior to analysis (if sufficient material is available) to verify the integrity of the starting RNA.

   Note: if your RNA is degraded, take heart: the protocol works fine with degraded RNA. After all, the first step is fragmentation! But if samples are degraded they will be degraded to varying extents, and fragmentation times will have to be optimized separately for each sample (see below).

2. Ideally, 1 µg of total RNA should be used. RNA should be treated with DNAse prior to cleanup (e.g. using an on-column DNAse treatment) so that DNA does not interfere with quantification.

3. The reaction begins with a volume of 10 µl, so if your RNA is <100 ng/µl you may need to concentrate samples first by drying under vacuum or precipitating (with LiCl or EtOH).

4. RNA should be eluted or dissolved into either nuclease-free water (NFW) or 10 mM Tris pH 8.0. Avoid TE buffers or other EDTA-containing storage buffers.

RNA fragmentation

In this step, RNA is fragmented by incubating at high temperatures to expose 3’ ends of transcripts and standardize template lengths across genes.

Before fragmenting your samples, it’s important to optimize the incubation time for each batch of samples, by incubating RNA at 95°C for a range of incubation times and identifying the duration that produces the desired range of molecular weights (100-500 bp) and clearly eliminates rRNA bands.
This is typically 10-15 minutes. Note that fragmentation success is affected by the buffer, volume, and concentration of the RNA.

1. For each sample, combine 1 µg of RNA with 1 µl 100 mM Tris (pH 8.0) and fill to a total volume of 10 µl with NFW (final concentrations, RNA: 100 ng µl, Tris: 10 mM). Reactions are best conducted in PCR strip tubes or a 96-well PCR plate.

2. Carefully seal all wells and incubate RNA at 95°C for the duration identified in preliminary trials to produce the desired size range (~100 – 500 bp). This can be most easily accomplished in a thermocycler.

3. Analyze 100 ng from each sample of fragmented RNA, alongside an intact sample of ~100 ng RNA from the same sample, on a 2% agarose gel* to confirm that fragmentation worked as intended.

Note: If you’re preparing a large number of libraries, it’s sufficient to analyze a represent subset of ~6-8 samples at this stage as long as the initial RNA integrities were similar.

*See note at end of document about gel electrophoresis.

First-strand cDNA synthesis

In this step, cDNA is synthesized from the 3’ ends of transcripts using oligo-dT primers and template switching activity to attach known sequences at each end of the fragment.

If your initial template contains much less than 1 µg of RNA you can save substantially on costs of library preparation by scaling the volume of this reaction, and amount of enzyme, accordingly.

1. Confirm that the volumes of fragmented RNA remain at least 9 µl after losses to evaporation during fragmentation, and replace with additional (NFW) if necessary to bring to a final volume of 9 µl.

2. Add 1 µl of the primer 3ILL-20TV (at 10 µM) to each well. Incubate at 65°C for 3 minutes in a thermocycler, then transfer immediately onto ice.

3. 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)

<table>
<thead>
<tr>
<th>Component</th>
<th>µl</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFW</td>
<td>3</td>
</tr>
<tr>
<td>dNTP (10 mM ea)</td>
<td>1</td>
</tr>
<tr>
<td>5X RT buffer (comes with enzyme)</td>
<td>4</td>
</tr>
<tr>
<td>10 µM 3ILL-4N-TS</td>
<td></td>
</tr>
<tr>
<td>(RNA oligonucleotide; stored at -80°C)</td>
<td>1</td>
</tr>
<tr>
<td>Tetro Reverse Transcriptase</td>
<td></td>
</tr>
<tr>
<td>(Bioline #BIO-65050)</td>
<td>1</td>
</tr>
</tbody>
</table>
4. Add 10 µl of this master mix to each sample of fragmented RNA on ice, mix thoroughly, and incubate in a thermocycler for one hour at 45°C.

5. Incubate at 85°C for 5 minutes to inactivate the RT, and store on ice or at -20°C until ready to proceed to the next step.

cDNA amplification

In this step, cDNA tags are converted to double-stranded cDNA through PCR. It's important to minimize the number of cycles used; the goal is only to produce enough material for accurate quantification and quality control. Mapping efficiency and accuracy of gene expression profiles are reduced by increasing cycle numbers, so the fewer cycles the better (ideally 10-17).

A test-scale PCR is first conducted (using a representative subset of samples) to identify the minimum number of cycles required to produce a visible product on the gel, then large-scale reactions conducted for all samples at the optimum cycle number. Test-scale reactions include controls for different primer combinations, to avoid artifacts that may arise from contamination of reagents or samples.

1. Prepare a set of master mixes for test-scale PCR tests. 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, so if you use a different amount of template, adjust the water accordingly.

   (Volumes given in µl)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFW</td>
<td>13.3</td>
<td>12.7</td>
<td>12.7</td>
<td>12.1</td>
</tr>
<tr>
<td>dNTP (10 mM each)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>5X Q5 reaction buffer</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>10 µM 5ILL primer</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>10 µM 3ILL-20TV primer</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Q5 DNA Polymerase (NEB #M0491S)</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

2. For each sample, prepare four PCR tubes labeled A-D. Add 18 µl of the appropriate master mix to each tube.

3. Add 2 µl of the FS-cDNA prepared above.
4. Amplify in a thermocycler using the following profile:
   
   98°C 30 sec, (98°C 10 sec, 60°C 30 sec, 72°C 30 sec) X 10 cycles

5. After 10 cycles sample 5 µl from each reaction and return to the thermocycler for an additional 2-5 cycles.

6. After sampling each reaction at each of 3 cycle numbers (e.g. 10, 13, & 16 cycles), load each sample on a 2% agarose gel.
   
   - Ideally, reaction D should show a smear ranging from ~100-500 bp, and nothing should be visible in reactions A-C.
   
   - If nothing is detected in reaction D, you can repeat the reactions at higher cycle numbers (up to a maximum of 20), or with additional template.
   
   - If product ever appears in reactions A or C, this indicates too much template, too many cycles, or contamination in one or more reagents. Troubleshoot the FS-cDNA.
   
   - A small amount of product in reaction B is OK, since the 3ILL-20TV oligo omitted in that reaction is still present at low concentrations in the FS-cDNA. The goal at this stage is to identify the minimum cycle number that produces a visible smear in reaction D while remaining reasonably clean in reactions A-C.

7. After determining the optimum amount of template and number of cycles, prepare a larger prep-scale reaction for each cDNA sample as follows. This recipe assumes 10 µl of template (i.e. scaled up five-fold from the default amount in test-scale PCR), so if you use more than 2 µl template in test-scale PCR adjust the template and water accordingly.

   (all volumes given in µl)

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFW</td>
<td>60.5</td>
</tr>
<tr>
<td>dNTP (10 mM ea)</td>
<td>2.5</td>
</tr>
<tr>
<td>5X Q5 reaction buffer</td>
<td>20</td>
</tr>
<tr>
<td>10 µM 5ILL primer</td>
<td>3</td>
</tr>
<tr>
<td>10 µM 3ILL-20TV primer</td>
<td>3</td>
</tr>
<tr>
<td>Q5 DNA polymerase</td>
<td>1</td>
</tr>
</tbody>
</table>

8. After pre-scale PCR, check 5 µl of each product on a gel to verify that the reaction worked as expected before freezing or purifying the product.

9. Purify PCR products with the method of your choice (we have used the NucleoMag 96 PCR Cleanup Kit, and Omega BioTek Cycle-Pure kit). If using any column-based kit, read the instructions carefully and be sure to include any optional steps intended to include low-molecular-weight fragments (<200 bp). Alcohol precipitation is the most cost effective option; use low temperatures and excess alcohol concentrations to maximize yield of LMW fragments if using precipitation.

10. Elute or resuspend purified samples in 25 µl NFW, and quantify by OD260 (e.g. Nanodrop) or fluorescence (e.g. Accublue or QuBit) [Since the material is purified cDNA, either method will give similar results]. You need at least 50-70 ng cDNA (total) for each sample at this stage.
Adaptor extension and barcoding

In this step, sample-specific barcodes and platform-specific primer binding sites are incorporated into the sequencing constructs using PCR with a minimal number of cycles (4-6).

Because the size distribution of templates is a critical factor for successful cluster amplification, constructs are then size-selected prior to sequencing. The directions below outline a simple procedure for selecting fragments ranging from 200-300 bp in size that does not require any special equipment. Other methods of size selection could be substituted provided they achieve this same size range.

1. First, test-scale PCRs are conducted to verify yield and specificity of the reaction, using a variety of primer combinations to control for artifacts that may arise from contamination.

   If preparing libraries in high throughput, choose a representative subset of 6-8 samples for this test.

   For each sample to be tested, dilute an aliquot to 10 ng µl⁻¹ so that all templates are at the same concentration.

2. Prepare four separate master mixes for test-scale PCR. The following volumes are for a single reaction, so multiply these values by the total number of samples plus a small additional amount to account for pipetting error. The values shown here assume the use of 2 µ diluted PCR product as template, so if you change this be sure to change the volume of water accordingly.

   \[
   \begin{array}{cccc}
   \text{NFW} & 5.7 & 3.7 & 3.7 & 1.7 \\
   \text{dNTP (10 mM ea)} & 0.2 & 0.2 & 0.2 & 0.2 \\
   \text{5X Q5 reaction buffer} & 2 & 2 & 2 & 2 \\
   \text{Q5 DNA polymerase} & 0.1 & 0.1 & 0.1 & 0.1 \\
   5' barcode oligo ("HTxx") (1 µM) & 0 & 2 & 0 & 2 \\
   3' barcode oligo ("BCxx") (1 µM) & 0 & 0 & 2 & 2 \\
   \end{array}
   \]

3. Aliquot 8 µl of each master mix into each well.
4. Add 2 µl of the cDNA previously quantified and diluted to 10 ng µl⁻¹ (above) to each of these four reactions (A-D) for a total volume in each reaction of 10 µl.

5. Amplify in a thermocycler using the following profile:
   
   98°C 30 sec, (98°C 10 sec, 60°C 30 sec, 72°C 30 sec) X 4 cycles

6. After 4 cycles check 5 µl of the PCR products for all reactions on a gel.
   
   - Ideally, reaction D should show a smear ranging from ~200-600 bp, and nothing should be visible in reactions A-C.
   
   - If nothing is detected in reaction D, you can continue the reaction for additional cycles (up to a total of 6), or repeat the reaction with additional template.
   
   - Reactions A-C should remain relatively clean regardless of cycle numbers. If you observe products in these reactions it probably indicates contamination of the reagents or libraries.

7. When the optimum number of cycles and volume of template have been determined, conduct a prep-scale reaction based on those parameters, using with 50 ng template in 50 µl total volume. The following master mix assumes the use of 10 µl of template per 50 µl reaction; if you adjust this template volume be sure to adjust the volume of water accordingly. This recipe 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)

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFW</td>
<td>26.5</td>
</tr>
<tr>
<td>dNTP (10 mM ea)</td>
<td>1</td>
</tr>
<tr>
<td>5X Q5 reaction buffer</td>
<td>10</td>
</tr>
<tr>
<td>Q5 DNA polymerase</td>
<td>0.5</td>
</tr>
</tbody>
</table>

8. Aliquot 38 µl of master mix to each well, then add 1 µl of each sample-specific barcode oligo (at 10 µM), and 10 µl of purified cDNA (diluted to 5 ng µl⁻¹).

9. Be sure to record your barcode assignments at this step! (e.g. sample X got BC oligo Y and HT oligo Z).

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

**Gel purification**

In this step the sequencing constructs are gel-purified to eliminate residual genomic DNA and primer dimers.

1. Prepare a diluted loading buffer to minimize dye carryover during gel purification. Dilute any standard DNA loading buffer (e.g. NEB B7021S) 1:10 using a 50% glycerol solution prepared with NFW.

2. Prepare a 2% agarose gel*. Use a thick comb that can accommodate 120 µl volumes, or tape together two wells if required.
3. Combine each 50 μl PCR product with 10 μl of diluted loading buffer, load entire mixture into the gel, and run the gel long enough for good resolution in the 50-500 bp range.

4. View the gel briefly (<30 seconds) on a UV transilluminator set at low intensity to verify the presence of target bands and adequate separation of molecular weight standards to resolve bands in the 50 to 500 bp range. Typically ~5 cm run distance (well to dye front) is sufficient.

5. Cut out the target molecular weight range (200-300 bp), being careful to cut the same range for each sample.

*See note at the end of document about gel electrophoresis.*

Note: at this stage a commercial gel-extraction kit can be used instead of the following steps.

6. Transfer each gel slice into a 2 ml microcentrifuge tube and add 40 μl NFW.

7. Centrifuge tubes 1 min at high speed to bring gel into contact with the water. Optional: incubate at 4°C overnight to maximize yield. Otherwise, proceed directly to freezer.

8. Freeze at -80°C for at least 30 minutes.

9. Centrifuge at maximum speed for 10 minutes.

10. Press gel slice against side of tube using pipette tip, and withdraw supernatant (at least 60 μl should be recovered). If less than 60 μl is accessible, repeat centrifugation. Transfer the supernatant to a new PCR tube or plate.

Combining libraries for sequencing

Finally, the libraries are pooled in equal ratios in an effort to sequence all samples at equal coverage.

1. Prepare a 1:100 dilution of each library by combining 2 μl of the eluted library (above) with 198 μl NFW. Typically this is done using a multichannel pipette and micro wellplates.

2. To quantify each library using qPCR, prepare a master mix. The following recipe is for one sample, so adjust accordingly. (In experienced hands 1-2 reactions per sample are sufficient, but users new to qPCR may benefit from additional replication).

   NFW 4.3 μl
   SYBR qPCR master mix, 2X 7.5 μl (e.g. Bioline #BIO-92005)
   ILL-Lib1, 10 μM 0.6 μl
   ILL-Lib2, 10 μM 0.6 μl

Note: Other PCR master mixes can obviously be substituted here. The important thing is to use a master mix compatible with your PCR instrument; this reaction is shown only as an example.

3. Pipette 13 μl qPCR master mix into each well of a PCR plate, then add 2 μl of each diluted library to the appropriate well.

4. Conduct qPCR and calculate C_T for each sample.
5. To determine volumes of each library for the combined pool:
   a. Rank samples from lowest to highest C_T and identify reference sample (sample with highest C_T)
   b. Calculate proportion of each library to sequence as:
      \[ P_L = 2^{^\text{CT (sample)} - \text{CT (reference)}} \]
   c. Calculate the volume of each library to use as:
      \[ V = P_L \times 60 \mu l \]

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

Note: As a rough rule of thumb, the pool of combined samples used for a single lane of sequencing should be at least 200-500 µl at this stage (it may be substantially higher).

6. Combine libraries using the volumes calculated from qPCR to produce a pool for sequencing.

7. Illumina sequencing typically requires templates in ≤20 µl volume at ≥2 nM concentration. The pooled libraries produced above are typically too dilute for sequencing. To purify and concentrate these libraries, they may be precipitated with isopropanol (below) or using a commercial PCR cleanup kit.

8. To precipitate:
   a. add 0.1 volumes of 3M sodium acetate and 3 volumes 100% isopropanol. Incubate 30 minutes at -20°C.
   b. Centrifuge 20 minutes at maximum speed, 4°C.
   c. Dry pellet 15’ at room temperature and resuspend in 25 µl NFW.

9. (Optional) Quantify your pooled library using a method sufficiently precise and sensitive to quantify concentrations in the 1-10 ng µl⁻¹ range. You need a concentration of at least 1.0 ng µl⁻¹ at this stage to have enough for sequencing.
Sequences of oligonucleotide primers used in this protocol

Minimal oligo set
To prepare for RNA-Seq using this protocol, you'll need to have the following oligonucleotides synthesized.

<table>
<thead>
<tr>
<th>Oligo Set</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILL-4N-TS*</td>
<td>ACCGCAUGCGGCUACACGACGCUCUUCCGAUCUNNNNGG</td>
</tr>
<tr>
<td>3ILL-20TV</td>
<td>ACGTGTGCTCTTCCGATCTAAATTTTTTTTTTTTTTTTV</td>
</tr>
<tr>
<td>5ILL</td>
<td>CTACACGACGCTCTTCCGATCT</td>
</tr>
<tr>
<td>ILL-Lib1</td>
<td>AATGATACGGCGACCACCGA</td>
</tr>
<tr>
<td>ILL-Lib2</td>
<td>CAAGCAGAACGGCGCATACGA</td>
</tr>
</tbody>
</table>

*This is an RNA oligo. Store small aliquots at -80 and minimize freeze-thaw cycles.

Barcoding primers
Buy a variety of barcode 1 and barcode 2 oligos to provide a sufficient number of unique combinations for your project (e.g. 8 BC and 8 HT oligos = 64 unique combinations). Because these are long oligos, they should be specially synthesized or purified to ensure purity (e.g. Ultramer synthesis or HPLC purification).

(Example barcode 1 oligo, x in sequence indicates barcode positions)
BCxx:
CAAGCAGAACGGCGCATACGAGATxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

(Example barcode 2 oligo, x in sequence indicates barcode positions)
HTxx:
AATGATAACGGCGACCACCGAGATCTACACxxxxACACTCTTTTCCCTACACGACGCTCTTCCGATCT
Gel electrophoresis for preparation of RNA-Seq libraries

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 http://openwetware.org/wiki/SB. We add EDTA to 20 mM (in the 20X solution). We use 1% agarose gels to analyze RNA fragmentation and 2% to analyze and size-select libraries.

We also use ethidium bromide at 0.1 µg ml⁻¹, but other dyes with comparable sensitivity can be used instead. EtBr is very cheap and sensitive, but somewhat toxic and must be illuminated with UV light which damages DNA, so some users prefer to use more expensive dyes that can be viewed with blue light.

Generally, there is a lot of flexibility in the details of gel electrophoresis but users should strive to maximize resolution and minimize damage to the DNA.