

## **Revised - Improved protocol for preparation of cDNA samples for *de novo* transcriptome sequencing with 454-Titanium technology**

Galina Aglyamova, Eli Meyer, and Mikhail V. Matz  
Integrative Biology Section, University of Texas at Austin

Jade Buchanan-Carter, Keithanne Mockaitis, and John K. Colbourne  
Center for Genomics and Bioinformatics, Indiana University

Revised by:

Joshua A. Udall  
Dept. of Plant and Wildlife Sciences, BYU, Provo, UT-84602

Edward R. Wilcox  
BYU DNA Sequencing Center, BYU, Provo, UT-84602

[Modified steps are highlighted in yellow.]

### **Background**

The preparation of appropriately modified cDNA is a critical step ensuring the overall success of transcriptome diversity characterization using next-generation sequencing methods. Researchers at UT Austin have developed an integrated protocol specifically adapted for the use with 454 technology, with the primary focus on protein-coding transcriptome data assembly and annotation *de novo* (i.e., in the absence of the reference genome data). This method generates pools of fragmented cDNAs flanked by two standard 454 amplification/sequencing primers, ready for amplification of individual sequences on microbeads and sequencing. The method requires as little as 50 ng total RNA at the start, and solves three most important problems inherent in comparable protocols: artifacts due to long 46 homopolymer regions, large proportion of unusable (adaptor) sequences in the 454 output, and coverage bias towards 3'-termini of transcripts.

### **Invention Description**

The developed method uses PCR-suppression effect to eliminate problems associated with improper primer annealing, adapter ligation, and adaptor concatenation. Modification of the cDNA synthesis procedure avoids incorporation of long A/T-stretches originating from the polyA tails of the mRNA, which would create problems during pyrosequencing stage. cDNA fragments in samples produced by this method bear the sequencing primer only on the ends corresponding to the fragmentation sites of the original mRNAs rather than 5' or 3' termini, facilitating even coverage and further lowering the proportion of unusable adaptor sequences in the output. To further reduce the 3'-end bias, the method uses two approaches. First, the desired distribution of lengths within the originally produced cDNA can be achieved by varying the conditions of the amplification reaction (there is no physical separation procedure involved). Second, the final products are generated in three separate samples, specific to 3'-terminal, 5'-terminal, and middle cDNA fragments, which can be then mixed in a desired proportion or sequenced independently. To enable simultaneous sequencing of several samples, the method uses its own cDNA barcodes incorporated into adaptor sequences.

## Benefits

- Requires small amount of total RNA as a starting material
- High output of useful sequence due to elimination of adaptor-related artifacts (2-5 fold more new sequences per run than in analogous data published applications)
- Provides even coverage of the length of individual transcripts due to possibility to enrich the initial cDNA sample with longer copies, strategic placement of the sequencing primer, and generation of separate final samples for 5', 3', and middle cDNA fragments
- Eliminates the need for strand-selection prior to emulsion step PCR due to the inherent control over adaptor configurations
- Allows simultaneous sequencing of several samples through adaptor barcoding.

## Protocol Overview:

The initial cDNA is produced using SMART cDNA amplification kit (Clontech) (Zhu et al, 2001) but with different cDNA synthesis primer:

Cap Trsa-CV (first strand cDNA synthesis primer):

5'- AAGCAGTGGTATCAACGCAGAGT CGCAGTCGGTACTTTTTTCTTTTTT -3'  
("cap" primer sequence) ("broken chain" polyT)

The purpose of the "broken chain" T-primer is to reduce read artifacts during 454 pyrosequencing, which may get thrown out of calibration by a too strong signal produced from a long mononucleotidestretch (such as polyT or PolyA).

The cDNA is then:

- [optionally] normalized using Trimmer kit (Evrogen) and re-amplified using cap primer;
- nebulized or sonicated to the average fragment size of 500-1000;
- end-polished (by incubation with a DNA polymerase and dNTPs in appropriate buffer) and ligated to the mixture of 'Atitn+' and 'Btitn+' adapters.

Each of these adapters is an equimolar mixture of two oligos (typically, 1  $\mu$ M each in the working concentration), a long one that actually gets ligated by its 3' end and a short one that complements to the 3' end of the longer one to mimic the double-stranded blunt end for the ligase. The short oligo is not getting ligated since it does not have a 5'-phosphate.

### Atitn+ adapter

Long oligo: 5'- TCCCTGCGTGTCTCCGACTCAG CCGCGCAGGT -3'  
Atitn primer sequence suppression tag + barcode (underlined)

Short oligo: 5'- ACCTGCGCGG -3'

This one has a CAG barcode. Here are some other possible variants of barcoded Atitn+ adapters (pairs of long and short oligos)

GAC: TCCCTGCGTGTCTCCGACTCAG CCGCGGACGT  
ACGTCCGCGG  
AGC: TCCCTGCGTGTCTCCGACTCAG CCGCGAGCGT  
ACGCTCGCGG  
CGA: TCCCTGCGTGTCTCCGACTCAG CCGCGCGAGT  
ACTCGCGCGG  
ACG: TCCCTGCGTGTCTCCGACTCAG CCGCGACGGT  
ACCGTCGCGG



## Detailed protocol

**1. RNA template preparation.** These steps are recommended but may not be necessary, depending on your protocol of choice for isolating total RNA.

- a. Begin with about 0.5-1 µg RNA from the organism of your choice (note: the latest version of the Clontech's SMART kit claims the amount can be as low as 50 ng).
- b. Precipitate RNA by adding 1 volume 13.3M LiCl, incubating 30 minutes at -20°C, and centrifuging 20 minutes at 16g at room temperature.
- c. Rinse RNA pellets briefly with 80% ethanol (don't centrifuge) air dry at room temperature and dissolve pellets in EB (10 mM).
- d. Analyze RNA on a gel to evaluate integrity.

[While we do analyze RNA quality in Bioanalyzer (Agilent), we skip step #1 and start with cDNA synthesis, step #2.]

**2. First strand cDNA synthesis** (at this and the next stage, follow Clontech's SMART cDNA amplification protocol, but replace the cDNA synthesis primer by Cap-TRSA-CV).

- a. Combine 4 µl RNA (for a total of 1000 ng RNA) with 1 µl 10 µM Cap-TRSA-CV primer. Incubate 3 minutes at 65°C, then chill on ice.
- b. To the above tube, add a premixed solution containing the following:
  - 2 µl 5X first-strand synthesis buffer
  - 0.5 µl 10mM dNTP
  - 0.1 µl 0.1 M DTT
  - 1 µl 10 µM SMART IIA Oligonucleotide
  - 1 µl SMART Scribe RT (Clontech)
- c. Incubate at 42°C for 1 hour.
- d. Terminate the reaction by incubating at 65°C for 15 minutes, then return the tube to ice.
- e. Dilute 5-fold in water to minimize carryover of primers into subsequent reactions.

### 3. cDNA Amplification

- a. For each first-strand-cDNA sample, set up 13 PCR reactions (30 µl each).
  - 3 µl diluted FS-cDNA (from step 2e).
  - 19 µl H<sub>2</sub>O
  - 3 µl 10X HF buffer (Clontech)
  - 3 µl 10 mM HF dNTP (Clontech)
  - 1.45 µl 12 µM cDNA amplification primer (5' IIA, Clontech)
  - 0.65 µl Advantage HF polymerase (Clontech)
- b. Amplify using the following profile:  
94°C for 5 minutes; (94°C for 40 seconds 65°C for 1 minute, 72°C for 6 minutes) x X number of cycles, depending on the sample.  
After PCR, hold product at room temperature.
- c. Starting at the 15<sup>th</sup> cycle (earlier, or later for other samples, depending on the sample), aliquot 5 µl PCR reaction from a PCR tube at every 2 cycles. Aliquot the reaction sample towards the end of 72°C amplification stage. Use reaction sample from the same tube for subsequent withdrawals.
- d. Evaluate PCR product by loading 5 µl on a gel and visualizing with Ethidium bromide. Alternatively, fast Lonza gels can be used. There should be a faintly visible smear with some or no bands, with the majority of product falling between 500 and 3000 bp in length. Add more cycles if there is nothing visible on the gel, then evaluate again. If the product is not amplified in 20-24 cycles, something is wrong - start over from the cDNA synthesis step.

*NOTE 1: To avoid the failure as mentioned in step 3d, it is safer to start the amplification with one sample. After evaluating this on a gel, a larger (12X) amplification can be started.*

*NOTE 2: The total amount of cDNA product per tube should not exceed 200 ng, which means that the smear on the agarose gel (20 ng per lane) should be really faint. Make sure you don't over-amplify cDNA beyond that. However, if Lonza gel is used for evaluation, the smear might appear brighter than in the normal agarose gels. In this case, evaluate the amplification by comparing with samples from earlier steps that show smears from barely-started amplification.*

- e. Combine together 12 separate reactions prepared from each first-strand-cDNA sample, and purify this PCR product on a column (we use Qiagen Qiaquick PCR Purification kit). Elute the final sample in 50  $\mu\text{l}$  of EB (10 mM Tris-HCl pH 8.0). Do not use water in this step! Measure the concentration of DNA using Nanodrop spectrophotometer or any other appropriate method; there should be at least 2  $\mu\text{g}$  of DNA in total.

Then go directly to Sonication (step 6) or do optional Normalization step. (We go through the normalization step.)

#### 4. Normalization (optional)

- a. EtOH precipitate the product and dissolve it appropriate volume of milliQ water (we use 10  $\mu\text{l}$ ).  
5  $\mu\text{l}$  (1/10 volume) of 3 M NaAcetate pH 4.8-5.2 [We use 7  $\mu\text{l}$  of 2.14 M NaAcetate, pH 5.2.]  
125  $\mu\text{l}$  (2.5 volume) 96-100% EtOH [We use 100% EtOH.]  
Incubate at least 30 minutes at  $-20^{\circ}\text{C}$   
Spin 20 minutes at maximum speed (20000 rpm) at  $4^{\circ}\text{C}$ , rinse with 500  $\mu\text{l}$  70% EtOH, air dry, dissolve in appropriate volume of milliQ water to achieve a concentration of 2  $\mu\text{g}$  in 12  $\mu\text{l}$ . [We use 10-15  $\mu\text{l}$  of Chromosolve water (Sigma-Aldrich).] Measure the quantity of cDNA using a spectrometer.
- b. The Trimmer kit from Evrogen is used essentially according to the manufacturer's instructions; here we are just replicating their protocol.
- c. Prepare a hybridization master mix by combining:  
2  $\mu\text{g}$  cDNA from step 3e or 4a in  $\leq 12$   $\mu\text{l}$  volume  
4  $\mu\text{l}$  4X hybridization buffer  
H<sub>2</sub>O to a total volume of 16  $\mu\text{l}$   
(Note that final cDNA concentration= 125 ng/ $\mu\text{l}$ )
- d. Aliquot his out into 4 individual PCR tubes (4  $\mu\text{l}$  each) and overlay each with a drop (~ 50  $\mu\text{l}$ ) of sterile mineral oil; centrifuge briefly to collect liquid and separate phases.
- e. Using a thermal cycler, incubate at  $98^{\circ}\text{C}$  for 2 minutes, then at  $68^{\circ}\text{C}$  for a minimum of 5 hours, then proceed immediately to the next step.
- f. Near the end of the hybridization period (step 4e), warm the DSN master buffer (Trimmer kit) to  $68^{\circ}\text{C}$ .
- g. Prepare a  $\frac{1}{2}$  and  $\frac{1}{4}$  strength dilutions of the double-strand specific nuclease (DSN) using DSN storage buffer as the diluent; store on ice until ready to use.
- h. At the end of the hybridization period, add 5  $\mu\text{l}$  preheated master buffer to each tube. Spin briefly in a bench-top centrifuge and return immediately to the thermocycler. It is important to maintain the temperature at  $68^{\circ}\text{C}$  during this period. So minimize time spent out of the thermal cycler (no more than a few seconds).
- i. To the four tubes from step 4c, add the following, while maintaining temperature:

Tube	Add
A	1 $\mu\text{l}$ undiluted DSN enzyme
B	1 $\mu\text{l}$ $\frac{1}{2}$ dilution DSN enzyme
C	1 $\mu\text{l}$ $\frac{1}{4}$ dilution DSN enzyme
D	1 $\mu\text{l}$ DSN storage buffer (diluent)
- j. Incubate at  $68^{\circ}\text{C}$  for 25 minutes.

- k. Add 10  $\mu\text{l}$  of DSN stop solution (Trimmer kit) to each tube, mix well, and spin briefly to collect contents.
- l. Incubate at 68°C an additional 5 minutes.
- m. Add 20  $\mu\text{l}$  H<sub>2</sub>O to each tube then store at -20°C or proceed with next steps.

## 5. Amplification of normalized cDNA

- a. Set up 4 separate PCR reactions, each containing:
  - 1  $\mu\text{l}$  diluted normalized cDNA (from step 4m), one PCR reaction per DSN treatment
  - 21  $\mu\text{l}$  H<sub>2</sub>O
  - 3  $\mu\text{l}$  10X HF buffer (Clonetech)
  - 3  $\mu\text{l}$  10 mM HF dNTP (Clonetech)
  - 1.45  $\mu\text{l}$  12  $\mu\text{M}$  cDNA amplification primer (5' IIA, Clonetech)
  - 0.65  $\mu\text{l}$  Advantage HF polymerase (Clonetech)
- b. Amplify using the following profile:
  - 94°C for 5 minutes; (94°C for 40 seconds 65°C for 1 minute, 72°C for 6 minutes) x X number of cycles, depending on the sample
  - After PCR, hold product at room temperature.
- c. Starting at the 9<sup>th</sup> cycle (earlier, or later for other samples, depending on the sample), aliquot 5  $\mu\text{l}$  PCR reaction from a PCR tube at every 2 cycles. Aliquot the reaction sample towards the end of 72°C amplification stage. Use reaction sample from the same tube for subsequent withdrawals.
- d. Evaluate PCR product by loading 5  $\mu\text{l}$  on a gel and visualizing with Ethidium bromide. Alternatively, fast Lonza gels can be used. There should be a faintly visible smear with some or no bands. Add more cycles if there is nothing visible on the gel, then evaluate again.
- e. After evaluation, pick the DSN-treated cDNA which gives the best results and start a 12X reaction using the same reaction settings mentioned in steps 5a and 5b.
  - NOTE: It is safer to have 13X reaction samples so that 1 tube can be used during the 'large 12X' amplification for another evaluation on a gel.*
- f. Combine together the 12 separate reactions prepared from the normalized-cDNA sample, and purify this PCR product on a column (we use Qiagen Qiaquick PCR Purification kit). Elute the final sample in 50  $\mu\text{l}$  of EB (10 mM Tris-HCl pH 8.0). Measure the concentration of DNA using Nanodrop spectrophotometer or any other appropriate method.

The eluted DNA can be stored in -20°C or used in the next step: Nebulization.

## 6. Fragmentation (Nebulization)

For this step, the required Nebulizers and other reagents are used from the GS FLX Titanium Library Preparation Kit.

- a. We start with 2  $\mu\text{g}$  of Qiagen-cleaned normalized cDNA (step 5f) for nebulization. Pipette this amount of cDNA to the bottom (cup) of the Nebulizer.
- b. Add TE Buffer to a final volume of 100  $\mu\text{l}$ .
- c. Add 500  $\mu\text{l}$  of Nebulization Buffer, and mix thoroughly by swirling or pipetting up and down.
- d. Direct 30 psi (2.1 bar) of nitrogen through the Nebulizer for 1 minute to nebulize the DNA.
- e. After Nebulization, turn off the gas flow and disconnect the tubing from both the Nebulizer and the nitrogen tank.
- f. Tap the Nebulizer on a table top to collect as much of the material as possible. Total recovery should be greater than 300  $\mu\text{l}$ .

- g. Add 2.5 ml of Qiagen's Buffer PBI directly into the Nebulizer cup and swirl to collect all the material droplets.
- h. Purify the nebulized DNA using two columns from a MinElute PCR Purification Kit (Qiagen). Elute each column with 10  $\mu$ l of Buffer EB (room temperature) and pool the eluates for a total volume of 20  $\mu$ l.

## 7. Polishing and ligation with adaptors

- a. In a microcentrifuge tube, add the following reagents (GS FLX Titanium Library Preparation Kit), in the order indicated:
  - 20  $\mu$ l (2  $\mu$ g) of nebulized DNA fragments
  - 5  $\mu$ l 10X Polishing Buffer
  - 5  $\mu$ l BSA
  - 5  $\mu$ l ATP
  - 2  $\mu$ l dNTP mix
  - 5  $\mu$ l T4 PNK
  - 5  $\mu$ l T4 DNA polymerase (NEB, not from the GS FLX Kit)
  - H<sub>2</sub>O to final volume of 50  $\mu$ l
- b. Mix well and incubate the polishing reaction for 15 minutes at 12°C.
- c. Immediately continue incubation at 25°C for an additional 15 minutes.
- d. Purify the polished fragments using one Qiagen MinElute column using 250  $\mu$ l of PBI and elute the DNA using 10  $\mu$ l of Buffer EB (room temperature).
- e. In a microcentrifuge tube, add the following reagents (GS FLX Titanium Library Preparation Kit), in the order indicated:
  - 10  $\mu$ l cleaned, Polished DNA (step 7d)
  - 20  $\mu$ l 2X Ligase Buffer (NEB, not from the GS FLX Kit)
  - 5  $\mu$ l Adaptors
  - 2.5  $\mu$ l Ligase (NEB, not from the GS FLX Kit)
  - H<sub>2</sub>O to final volume of 40  $\mu$ l
- f. Mix well, spin briefly and incubate the ligation reaction at 25°C for 15 minutes.
- g. Purify the DNA using one Qiagen MinElute column using 200  $\mu$ l of PBI and elute the DNA using 100  $\mu$ l of Buffer EB (room temperature).

## 8. Small fragment removal

- a. Add AMPure beads to the 100  $\mu$ l eluate (step 7g) and vortex to mix. The quantity of AMPure beads is used as suggested for "Gel Cut Libraries", found on page 40 of GS FLX Titanium General Library Preparation Method Manual, October 2008 by Roche.
- b. Incubate for 5 minutes at room temperature.
- c. Using a Magnetic Particle Concentrator (MPC), pellet the beads against the wall of the tube.
- d. Remove the supernatant and wash the beads twice with 500  $\mu$ l of 70% Ethanol, incubating for 30 seconds each time.
- e. Remove all the supernatant and allow the AMPure beads to air dry completely.
- f. Remove the tube from the MPC, add 30  $\mu$ l of 10 mM Tris-HCl, pH 8.0 (or Qiagen's Buffer EB), and vortex to resuspend the beads.
- g. Using the MPC, pellet the beads against the wall of the tube once more, and transfer the supernatant containing the purified, Adaptor-ligated DNA to a fresh microcentrifuge tube.

## 9. PCR testing the ligation

- a. For each 454 cDNA library produced, prepare 5 different PCR reactions, each with a different combination of primers. Including water controls (no-template controls) is recommended. The primer combinations are as follows (final primer concentrations are shown):

<u>Tube</u>	<u>Primers</u>
1	0.2 $\mu$ M Atitn
2	0.2 $\mu$ M Btitn
3	0.1 $\mu$ M Atitn, 0.1 pM Btitn
4	0.1 $\mu$ M Atitn, 0.1 pM Btitn+halfswitch
5	0.1 $\mu$ M Atitn, 0.1 pM Btitn+TrsaC

The reactions 3, 4, and 5 specifically amplify internal fragments, 5' –ends and 3' –ends of the original cDNAs, respectively.

- b. Each PCR reaction is assembled s follows:  
3  $\mu$ l 10X PCR buffer  
0.75  $\mu$ l 10 mM dNTP  
1  $\mu$ l ligation product (from step 8g)  
0.6  $\mu$ l Advantage2 polymerase  
Primers as shown above (9a)  
H<sub>2</sub>O to final volume of 30  $\mu$ l
- c. Amplify these reactions using the following profile:  
94°C for 5 minutes;  
(94°C for 40 seconds, 65°C for 1 minute, 72°C for 1 minute) x20-23 cycles
- d. Load 3  $\mu$ l of these products on a gel, hold remainder at room temperature while the gel runs in case additional cycles are required.
- e. A visible smear ranging from 300-2000 bp should be visible in reaction #3, #4 and #5. None of the other two reactions should produce any product.
- f. If nothing is visible in any lanes, amplify for an additional 2 cycles and repeat the gel analysis.
- g. Repeat step 9f until visible smears are produced to allow determination of optimum cycle number. If more than 17 cycles were required to produce visible smears, try adding more template and using fewer cycles. In our experiments, 1  $\mu$ l of purified ligation product as PCR template and 20 cycles produced visible smears based on loading 3  $\mu$ l on a gel.

## 10. Amplification of samples for gel extraction

- a. Set up 'bulk' amplification based on the optimum cycle numbers and template volumes determined from the PCR tests above (steps 9 a-g). For our experiments, we set up 8 reactions as follows:

9  $\mu$ l 10X PCR buffer  
2.25  $\mu$ l 10 mM dNTP  
3X  $\mu$ l ligation product (determined in steps 9 a-g) (we add 3 $\mu$ l)  
1.5  $\mu$ l 6  $\mu$ M primer Btitn  
1.5  $\mu$ l 6  $\mu$ M primer Btitn OR Btitn+halfswitch OR Btitn+TrsaC  
1.8  $\mu$ l Advantage2 polymerase  
H<sub>2</sub>O to final volume of 90  $\mu$ l

We add the primers such that  $\frac{1}{2}$  of the primer mix is Btitn+halfswitch,  $\frac{1}{4}$  is Btitn and remaining  $\frac{1}{2}$  is Btitn+TrsaC.

- b. Amplify the reactions using the following profile:  
94°C for 5 minutes;  
(94°C for 40 seconds, 65°C 1 minute, 72°C for 1 minute) X N cycles,  
where N is the optimum cycle number determined (9 a-g)
- c. Load an aliquot on a gel to verify that the reaction amplified as expected.
- d. Chase using the following profile:  
78°C 1 minute, 65°C 1 minute, 68°C 1 minute
- e. Purify the PCR using a column (e.g., Qiagen Qiaquick), and elute in 100 µl EB.
- f. Run the eluate through 'Small fragment removal' steps as listed above (step 8 a-g).
- g. The 30 µl supernatant separated from the AMPure beads is now ready for gel purification.

## 11. Gel purification of final samples

- a. Make 1% MetaPhor Agarose gel with 1X TAE.
- b. Load your DNA with loading dye of choice and run at 61 volts for 4 hours for fragment purification.
- c. At the end of the run, the 500-700bp swath is cut and purified using QiaQuick Gel Extraction Kit (Qiagen) with the following exceptions:
  - Do not use heat while melting the gel.
  - Follow the optional extra QG buffer spin as stated in the instructions.
  - After adding the PE to the column, incubate for 2-5 minutes before spinning.
  - After the PE dry spin, rotate the column 180° and spin an additional 1 minute to ensure complete removal of the ethanol.
  - Elute the column with 24 µl of Buffer EB (room temperature).

This DNA is now ready for emulsion PCR for 454. From here on, we begin at step 3.6 (pg 24) the GS FLX Titanium General Library Preparation Method Manual, October 2008 by Roche.