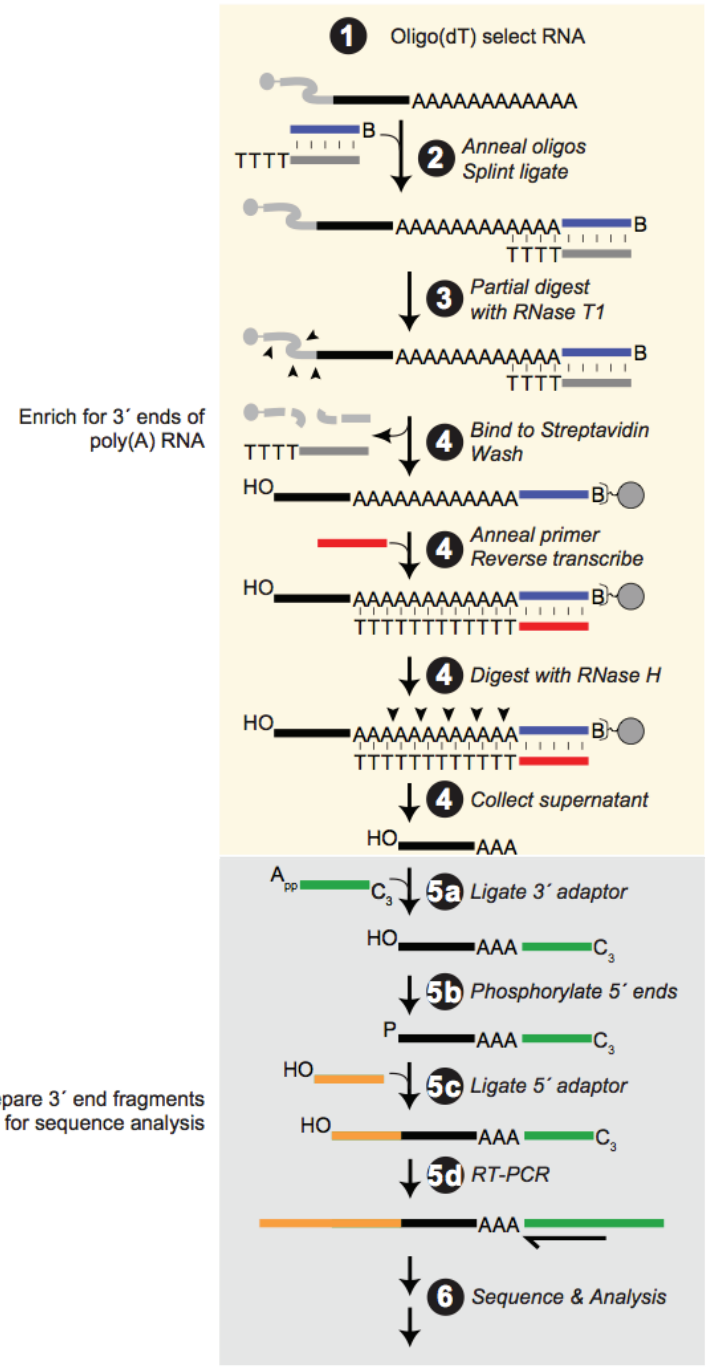
3P-Seq

Calvin Jan, David Koppstein, Noah Spies, and David Bartel

October 10, 2013



Poly(A)-Position Profiling by Sequencing (3P-Seq)

Outline

The protocol below is designed to generate cDNA libraries representing the 3′ ends of polyadenylated RNAs for Illumina sequencing. The libraries should be sequenced with the Illumina gDNA primer, which will sequence from the 3′ end of the RNA tag.

Oligonucleotides

- 3′ biotin adaptor - a 5′ phosphorylated, 3′ biotinylated RNA/DNA chimera (DNA bases are denoted by dN)
  - 5′-Phos.AGCGUGUAGGGCACCAUdGdCdAdCdAdTdAdC.3′-Bi
- 3′ biotin bridge, a DNA oligo
  - 5′-ATGGTGCCCTACACGCTTTTTTTTT-3′
- 3′ biotin RT primer, a DNA oligo
  - 5′-GTATGTGCATGGTGCCCTACACGCT-3′
- adenylated 3′ solexa adaptor ordered as a DNA oligo, then chemically adenylated.[1]
  - 5′-AppAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3′-C3spacer
- 5′ solexa adaptor - an RNA oligo
  - 5′-GUUCAGAGUUCUACAGUCCGACGAUC-3′
- solexa RT Primer - a DNA oligo
  - 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG-3′
- solexa PCR rev - a DNA oligo
  - 5′-AATGATACGGCGACCACCGA-3′
- solexa PCR fwd - a DNA oligo
  - 5′-CAAGCAGAAGACGGCATACGAGTTCAGAGTTCTACAGTCCGA-3′

Considerations

- Siliconized tubes should be used for all steps to reduce loss of nucleic acids from binding to the sides of the tube.
- Precautions should be taken to ensure that the work environment and all solutions are RNAse-free.
- All precipitations are done for at least 2 hrs at -20°C with 0.3 M final NaCl, 1.5 µl glycogen, and 3 volumes of EtOH, then spun at max speed for 20 minutes at 4°C, after which the supernatant is removed, the tube is pulsed again, residual ethanol is removed, and the pellet is dried for 5 minutes.
- We usually do Steps 1 and 2 on the first day, steps 3 and 4 on the second day, steps 5.1, 5.2, and 5.3 on the third day, and step 5.4 on the fourth day.

[1] The adenylation procedure is described in Lau et al., Science 2001 294: 858-862. Details can be found here.

Citations

- Jan, C.H., Friedman R.C., Ruby J.G., and Bartel D.P. Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. Nature. doi:10.1038/nature09616.
- Ulitsky I., Shkumatava A., Jan C.H., Subtelny A.O., Koppstein D., Bell G., Sive H., Bartel D.P. Extensive alternative polyadenylation during zebrafish development. Genome Research. doi:10.1101/gr.139733.112

Acknowledgments

Calvin Jan invented the protocol. David Koppstein and Noah Spies modified the protocol. We thank Ben Kleaveland and Wendy Johnston for assisting with optimization experiments, and David Weinberg, Alex Subtelny, and Josien van Wolfswinkel for helpful advice.

1. Enrich for polyA$^+$ RNA

Materials required

- 200 µl oligo dT_{25} magnetic beads (Invitrogen)
- 75 µg Total RNA (from TRI Reagent Extraction; can use 20-75 µg, but more is better)
- 300 µl Binding Buffer [20 mM Tris-Cl pH 7.5, 1 M LiCl, 2 mM EDTA]
- 400 µl Buffer B [10 mM Tris-Cl pH 7.5, 0.15 M LiCl, 1 mM EDTA]
- 100 µl 10 mM Tris-Cl pH 7.5 (pre-warm to 75°C)
- 1.5 µl Glycoblue (Life Technologies)
- 6.4 µl 5 M NaCl

Procedure

1. Magnetize beads for 30-60 seconds and discard supernatant.
2. Wash beads by resuspending in 100 µl binding buffer by pipetting or vortexing, magnetize and discard supernatant as before.
3. Resuspend in 100 µl binding buffer; the beads are now ready for binding.
4. Dilute 75 µg of total RNA to a final volume of 100 µl in water.
5. Add 100 µl binding buffer to diluted RNA and heat at 65°C for 2 minutes, then place on ice for 2 minutes to denature.
6. Add denatured RNA to beads, and rotate tubes at room temperature for 5 minutes to anneal.
7. Magnetize beads for 30-60 seconds and discard the supernatant. Resuspend beads in 200 µl buffer B and wash 1 minute rotating at room temperature.
8. Repeat wash in buffer B.
9. Resuspend beads in 100 µl pre-warmed 10 mM Tris-Cl pH 7.5.
10. Incubate at 75°C for 2 minutes, then magnetize 30 seconds. Collect the supernatant, which contains the polyA$^+$ RNA.
11. Add 1 µl glycoblue, 6.4 µl 5M NaCl and 300 µl ethanol to the supernatant and precipitate at -20°C for at least 2 hours.

2. Attach biotin linker to polyadenylate 3′ ends

Materials required

- PolyA$^+$ RNA from step 1, resuspended in 7 µl ddH_2O
- 1 µl 25 µM 3′ biotin adaptor
- 1 µl 25 µM 3′ biotin bridge
- 10x annealing buffer [100 mM Tris-Cl pH 7.5, 500 mM NaCl, 10 mM EDTA]
- 1 µl 10 mM $MgCl_2$
- 2 µl 10x ligase buffer (-ATP) [500 mM Tris-Cl pH 7.5, 100 mM $MgCl_2$, 100 mM DTT]
- 1 µl 5 mM ATP
- 1 µl T4 RNL2 (wt, NEB)
- 0.5 µl T4 RNL2 (truncated K227Q, NEB) - optional

Procedure

1. Mix the following in the PCR tube:
   - 7 µl RNA
   - 1 µl 25 µM 3′ biotin adaptor
   - 1 µl 25 µM 3′ biotin bridge
   - 1 µl 10x annealing buffer
2. Heat at 75°C for 5 minutes, then slowly ramp down to 18°C (in a PCR machine, 0.1°C/s).
3. Once the tube has reached room temperature, combine the remaining reagents:
   - 1 µl 10 mM $MgCl_2$
   - 2 µl 10x ligase buffer (-ATP)
   - 1 µl 5 mM ATP
   - 5 µl ddH_2O (4.5 µl if also using RNL2 mutant)
   - 1 µl T4 RNL2
   - Optional 0.5 µl T4 RNL2 mutant
4. Incubate at 18°C overnight, ~16hrs.
5. Freeze at -20°C or proceed directly to step 3 below.

3. RNAse T1 Digestion

The RNAse T1 digestions may require optimization. To ensure a partial digestion, we recommend cordycepin labeling of an *in vitro* transcribed RNA $\geq$ 200 nt in length. This RNA can be spiked into the reaction to monitor the extent of digestion by T1. Optimize such that the majority of the RNA has been cut and a ladder of products is visible. We have had good results with the biochemistry grade T1 nuclease from Ambion using the volume and time described below.

Materials required

- Ambion RNAse T1 (biochemistry grade)
- 0.1 M EDTA
- Glycoblue
- Tracer RNA - optional, for monitoring the T1 digest

Procedure

1. Add the following directly to the ligation reaction:
   - 4 µl 0.1 M EDTA
   - 80 µl sequence buffer (Ambion)
   - 1.5 µl glycoblue
   - optional - 0.75 µl radiolabeled RNA
2. Heat to 50°C in PCR machine with heated lid for 5 minutes.
3. Ramp down to 22°C and hold.
4. Add 1 µl RNAse T1.
5. Incubate at 22°C for 20 minutes.
6. Add 250 µl stop solution,[2] precipitate at -20°C for at least 2 hours.

[2]Remember to add ethanol to stop solution according to manufacturer's instructions!

4. Biotin capture and 3′ end tag release

Materials required

- 100 µl Streptavidin M-280 or MyONE C1 Dynabeads
- 200 µl Bead prep buffer [0.1 M NaOH, 50 mM NaCl] (add 20 µl 1 M NaOH to 180 µl 55.5 mM NaCl)
- 200 µl 0.1 M NaCl
- 600 µl 2x B&W buffer [10 mM Tris-Cl pH 7.5, 1 mM EDTA, 2 M NaCl]
- 800 µl Wash buffer [10 mM Tris-Cl pH 7.5, 1 mM EDTA, 50 mM NaCl], pre-warmed to 50°C.
- 200 µl 1x SuperScript III buffer [40 µl 5x buffer + 12 µl 0.1 M DTT + 148 µl ddH_2O
- 1 µl 100 µM 3′ biotin RT primer
- 1 µl 10 mM dTTP
- 5 µl 5x SSIII buffer
- 1.5 µl 0.1 M DTT
- 1 µl SuperScript III (Invitrogen)
- 1 µl RNAse H (Invitrogen)
- P-30 Micro-spin column (Bio-Rad)
- 75 µl PCA, pH 8, with 1 mM EDTA (Sigma Aldrich P3803)
- 75 µl chloroform

Procedure

1. Prep beads
    - Aliquot 100 µl beads and magnetize for 30 seconds.
    - Discard supernatant and resuspend in 100 µl bead prep buffer to wash (rotate for 2 minutes at room temperature)
    - Repeat wash with 100 µl bead prep buffer.
    - Wash with 100 µl 0.1 M NaCl.
    - Repeat wash with 100 µl 0.1 M NaCl.
    - Wash beads in 200 µl 2x B&W buffer.
    - Resuspend beads in 200 µl 2x B&W buffer.
2. Bind and wash RNA
    - Resuspend RNA from step 3 above in 200 µl water.
    - Add to prepared beads, rotate at room temperature for 15 minutes.
    - Very briefly spin down tube and magnetize for 1 minute.
    - Collect and precipitate supernatant if evaluating T1 with tracer RNA.
    - Wash once in 400 µl 1x B&W buffer.
    - Wash in 400 µl wash buffer at 50°C for 2 minutes, occasionally inverting the tube to mix.
    - Repeat wash in 400 µl wash buffer at 50°C.

3. Release tags into supernatant
    - Wash beads in 200 µl 1x SuperScript III buffer at room temperature for 2 minutes.
    - Make RT mix:
        - 1 µl 100 µM 3′ biotin RT primer
        - 1 µl 10 mM dTTP
        - 5 µl 5x buffer
        - 1.5 µl 0.1 M DTT
        - 15.5 µl water
        - 1 µl SSIII
    - Magnetize beads and remove supernatant
    - Add RT mix to beads and transfer to a PCR tube
    - Incubate at 48°C for 20 minutes in PCR machine.
    - Ramp down to 37°C, then add 1 µl RNAse H.
    - Incubate for 25 minutes; vigorously vortex every 5 minutes.
    - Add 50 µl ddH2O and transfer to a 1.5 mL tube.
    - Add 75 µl PCA, vortex briefly, then centrifuge for 1 mL. Transfer the supernatant to a new tube.
    - Add 75 µl chloroform, vortex briefly, then centrifuge for 1 mL. Transfer the supernatant to a packed P-30 column (see manufacturer's instructions).
    - Spin for 4 minutes at 650 x g, add 1.5 µl glycoblue, 4.8 µl 5 M NaCl, and 256 µl EtOH, and precipitate at -20°C for at least two hours.[3]

[3]We usually precipitate overnight at this point.

5. Library Preparation

5.1. 3′ ligation

Materials required

- 2 µl PEG 8000 (50%; NEB)
- 0.8 µl 100 µM adenylated 3′ solexa adaptor (3′ ligation adaptor)
- 0.8 µl 10x ligase buffer (-ATP) [500 mM Tris-Cl pH 7.5, 100 mM $MgCl_2$, 100 mM DTT]
- 0.5 µl T4 RNL1
- 3.9 µl ddH_2O
- 7 µl 2x urea gel loading buffer
- Radiolabeled Decade and Century markers (Life Technologies)

Procedure

1. Combine the above reagents and resuspend RNA pellet from previous step in mixture. Incubate at 22°C for 2 hours.
2. Add 7 µl 2x urea gel loading buffer and heat for 1 minute at 90°C, then pulse and snap cool on ice.
3. Cast a 1 mm[4], 6% denaturing urea polyacrylamide gel with 1cm x 1cm lanes.
4. Run gel for 1 hour at 5W, or until dye has migrated 5 cm, using radiolabeled Decade and Century as markers.
5. Excise gel slice ranging from 75 to 500 nucleotides on a 1mm, 6% denaturing urea polyacrylamide gel, using radiolabeled Decade and Century as markers.
6. Pierce a 0.5 mL tube with a 20 gauge needle and put it inside a 1.5 mL tube.[5]
7. Place gel slice in 0.5 mL tube and spin for 5 min at max speed.
8. Resuspend crushed gel in 800 µl 0.3 M NaCl and freeze at -80°C for 20 minutes.[6]
9. Heat at 50°C for 20 minutes with occasional vortexing, then rotate at room temperature for 20 minutes.
10. Divide into two tubes, add 1.5 µl glycogen to each, and precipitate with 1 mL EtOH at -20°C for at least 2 hours.

[4] A thin gel is used to minimize acrylamide and maximize elution.
[5] Adapted from Ingolia et al., Cell, 2011
[6] Adapted from the Szostak lab.

5.2. Phosphorylate 3′ ends

Materials required

- 21 µl Ligated RNAs (from step 5a, resuspended in 21 µl water)
- 2.5 µl 10x T4 DNA ligase buffer[7] (NEB)
- 0.5 µl Superase-IN
- 1 µl T4 PNK

Procedure

1. Combine the above reagents and incubate at 37°C for 1 hour.
2. Bring volume to 250 µl, then PCA and chloroform extract with equal volumes as before, and precipitate for at least 2 hours at -20°C.

5.3. 5′ ligation

Materials required

- 1 µl 100 µM 26.71 (5′ ligation adaptor)
- 0.5 µl 10x ligase buffer (-ATP) [500 mM Tris-Cl pH 7.5, 100 mM $MgCl_2$, 100 mM DTT]
- 0.5 µl RNL1 (NEB)
- 0.1 µl Superase-IN
- 2.9 µl water

Procedure

1. Combine the water, buffer, Superase-IN, and 5′ adaptors in a PCR tube.
2. Use this mix to resuspend the RNA pellet from step 5.2.
3. Add 0.5 µl T4 RNL1, and incubate overnight at 22°C.
4. Gel purify, crush slices, elute, and precipitate RNA from 100-500 nucleotides as in step 5.1.

[7] Contains ATP.

5.4. RT-PCR

Materials Required

- 14.6 µl RNA from step 5.3, resuspended in water
- 1 µl 100 µM solexa RT primer
- 6.4 µl 5x RT buffer (SuperScript)
- 3 µl 0.1 M DTT
- 17 µl 10x dNTPs
- 1 µl SSIII
- 5 µl 1 M NaOH
- 25 µl 1 M HEPES pH 7
- 1 µl 20 µM solexa PCR fwd primer
- 1 µl 20 µM solexa PCR rev primer
- 20 µl 5x PCR buffer (Phusion)
- 1 µl Phusion HF DNA polymerase (NEB/Finnzymes)
- P-30 Micro-Bio Spin Column (Bio-Rad)
- SYBR Gold (Invitrogen)
- MinElute Gel Purification Kit (Qiagen)

Procedure

1. Mix the following in a PCR tube:
   - 14.6 µl RNA
   - 1 µl 100 µM solexa RT primer
2. Heat at 65°C for 5 minutes, then place on ice for 2 minutes.
3. Add the following:
   - 6.4 µl 5x RT buffer
   - 3 µl 0.1 M DTT
   - 7 µl 10x dNTPs
4. Incubate at 48°C for 3 minutes.
5. Remove 3 µl and freeze for a no RT control (for PCR).
6. Add 1 µl SSIII and incubate at 48°C for 1 hour.
7. Add 5 µl 1 M NaOH and incubate at 90°C for 10 minutes.
8. Place on ice for 1 minute to cool, then add 25 µl 1 M HEPES pH 7.
9. Desalt 30 µl of the RT reaction over a P-30 column, freeze the rest.
10. Setup the PCR by mixing the following in a PCR tube:
    - 10 µl desalted RT reaction (or 3 µl of no RT control)
    - 20 µl 5x PCR buffer
    - 10 µl 10x dNTPs
    - 57 µl water (64 for no RT control)
    - 1 µl 20 µM solexa PCR fwd primer
    - 1 µl 20 µM solexa PCR rev primer
    - 1 µl Phusion
11. Amplify for 16-20 cycles with the following program:

(a) 98°C for 1 min
(b) 98°C for 15 seconds
(c) 60°C for 30 seconds
(d) 72°C for 30 seconds
(e) Goto (b) for 15-19 extra cycles.
(f) 72°C for 10 minutes.
(g) Hold at 10°C.

12. Ethanol precipitate for at least 1 hour.
13. Gel purify 150-550 bp range.
    - Run a 2% agarose gel and stain with SYBR gold for 20 minutes.
    - Purify using Qiagen's minelute gel extraction kit.
    - Elute in EB.

6. Library Sequencing and Analysis

Libraries should be sequenced with Illumina's gDNA primer. The resulting sequences are antisense to the sample RNA. 3P-Seq libraries will be, by design, heavily biased towards T bases for the first several cycles of sequencing. This homogeneity can lead to difficulties in cluster definition and base-calling, as neighboring clusters are less likely to have contrasting bases in the first several cycles. When possible, we recommend that users cyclically permute the flow cell images such that base calling is initiated with images starting at cycle 6 and images from cycles 15 are appended to the end of the run. The data are subsequently de-convoluted back to the actual sequence.

There are a number of tools available to align reads to a reference genome or transcriptome. For simplicity, we reverse complement reads before alignment. It is imperative to allow mismatches for the oligo(A) tails at the end of reads. This is accomplished by either relaxing alignment stringency at the 3′ end of the read, or stripping the 3′ A's and checking that one or more of these A's was non-genomic after the read has been aligned. We then consider only those reads that terminated in 2 or more consecutive adenylates, at least one of which did not originate from the genomic template. The remaining reads may be treated as strand-specific RNA-Seq if desired.

We cluster reads based on the coordinates of the alignment of their 3′-most non-A nucleotide. Reads are iteratively clustered by identifying the coordinate at which the greatest number of reads terminate then grouping all reads which terminate within 10 bases. Reads falling in this cluster are removed and the process repeated until all reads have been accounted for.

Because of the high depth of library sampling attainable by high-throughput sequencing, we require that clusters be definitively derived from multiple independent RNA molecules. This is achieved by requiring that clusters meet at least one of the following criteria:

1. Contain reads with different 3′ termini after removing consecutive 3′ adenylates.
2. Contain reads with different numbers of consecutive 3′ adenylates (remnants of the poly(A) tail).
3. Contain reads sequenced from independent libraries

Pairing 3P-Seq clusters to transcript models depends greatly on the quality of the genome assembly and gene annotations. Due to the limited read lengths, parsimony is invoked to relate 3P-clusters to neighboring transcript models (ie, for 3′UTRs, clusters most likely represent the 3′ ends of transcripts from the nearest annotated gene). If available, RNA-Seq data can improve the robustness of the pairing between transcript models and 3′ ends, or can be used for *ab initio* construction of transcript models. The pertinent issues for correct 3′ end assignment depend on the organism being studied, but can relate to gene density (are genes sparse or are they close to each other? do they overlap?), gene structure (operons, trans-splicing, genome rearrangements?), uniqueness (are there many recent gene duplications?), etc.