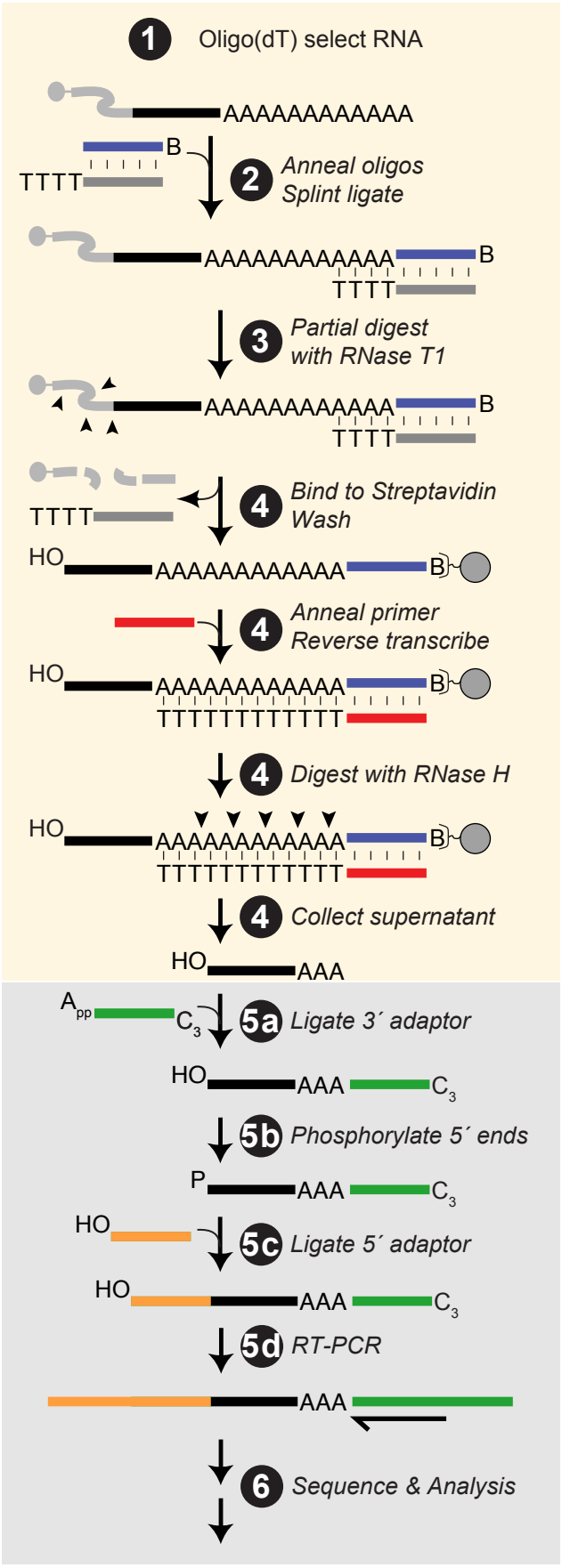
Poly(A)-position profiling by sequencing (3P-Seq)

Jan, C.H., R.C. Friedman, J.G. Ruby, and D.P. Bartel. Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. Nature, 10.1038/nature09616, published online Nov 17, 2010.

Overview



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

Jan, C.H., R.C. Friedman, J.G. Ruby, and D.P. Bartel. Formation, regulation and evolution of *Caenorhabditis elegans* 3´UTRs. Nature, 10.1038/nature09616, published online Nov 17, 2010.

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:

Oligos should be gel purified prior to use.

3´ biotin adaptor – a 5´ phosphorylated, 3´ biotinylated RNA/DNA chimera (dna bases = 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*
5' AppAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT.3'-C3spacer
5´ solexa adaptor – a 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'

* n.b. – The adenylation procedure is described in Lau et al., Science 2001 294: 858-862. Details can be found at http://web.wi.mit.edu/bartel/pub/protocols.html in the microRNA and siRNA cloning protocol.

Step 1: Enrich for polyA$^+$ RNA

Materials required (per sample):

160 µl Dynal oligo dT_{25} magnetic beads (Invitrogen)
60 µg* Total RNA (From trizol extraction)
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, .15 M LiCl, 1 mM EDTA]
100 µl 10 mM Tris-Cl pH 7.5 (pre-warm to 75^o C)
1 µl Glycoblue (Ambion)
6.4 µl 5M NaCl
270 µl 100% Ethanol

Procedure:

- Magnetize beads for 30–60 seconds and discard supernatant.
- Wash beads by resuspending in 100 µl binding buffer by pipetting or vortexing, magnetize and discard supernatant as before.
- Resuspend in 100 µl binding buffer; the beads are now ready for binding.
- Dilute 60 µg of total RNA to a final volume of 100 µl in water.
- Add 100 µl binding buffer to diluted RNA and heat at 65^o C for 2 minutes, then place on ice for 2 minutes to denature.
- Add denatured RNA to beads, and rotate tubes at room temperature for 5 minutes to anneal.
- 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.
- Repeat wash in buffer B.
- Resuspend beads in 100 µl pre-warmed 10 mM Tris-Cl pH 7.5.
- Incubate at 75^o C for 2 minutes, then magnetize 30 seconds. Collect the supernatant, which contains the polyA$^+$ RNA.
- Add 1 µl glycoblue, 6.4 µl 5M NaCl and 270 µl ethanol to the supernatant and precipitate at –20^o C for at least 2 hours.

* n.b. – We have had success starting with 20–60 µg total RNA.

Step 2: Biotin attachment to polyadenylate 3´ ends

Materials required (per sample):

5 µl *PolyA+ RNA from step 1, resuspended in 5 µl water*
1 µl 25 µM 3´ biotin adaptor
1 µl 25 µM 3´ biotin bridge
1 µl 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 7.8, 100 mM $MgCl_2$, 100 mM DTT]
1 µl 4 mM ATP
1 µl T4 RNL2 (wt, NEB)
.5 µl T4 RNL2 (truncated K227Q, NEB) – optional

Procedure:

- Mix the following in a pcr tube:
  - 5 µl RNA
  - 1 µl 25 µM 3´ biotin adaptor
  - 1 µl 25 µM 3´ biotin bridge
  - 1 µl 10x annealing buffer
  - 2 µl water
- Heat at 75^{o} C for 5 minutes, then slowly ramp down to 22^{o} C (In a pcr machine, 0.1^{o}C/sec).
- 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 4 mM ATP
  - 5 µl water (4.5 µl if also using RNL2)
  - 1 µl T4 RNL2
  - optional .5 µl T4 RNL2 mutant
- Incubate at 22^{o} C overnight, ~16 hours.
- Freeze at –20^{o} C or proceed directly to step 3 below,

Step 3: RNase T1 Digestion

The RNase T1 digestions may require optimization. To ensure a partial digestion, we recommend cordecypin labeling of an in vitro transcribed RNA ≥ 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 (biochemsitry grade)
0.1 M EDTA
Glycoblue
Tracer rna – optional, for monitoring the T1 digest

Procedure:

- Add the following directly to the ligation reaction:
  4 µl 0.1 M EDTA
  80 µl Sequence buffer (ambion)
  .75 µl glycoblue
  optional - 0.75 µl radiolabeled rna
- Heat to 50^{o} C in pcr machine with heated lid for 5 minutes.
- Ramp down to 22^{o} C and hold.
- Add 3 µl RNase T1.
- Incubate at 22^{o} C for 20 minutes.
- Add 220 µl stop solution, precipitate at –20^{o} C for at least 2 hours.

Step 4: Biotin capture and 3´ end tag release

Materials required (per sample):
100 µl Dynabeads M-280 streptavidin
200 µl Bead prep buffer [0.1M NaOH, 50mM NaCl]
200 µl 0.1 M NaCl
600 µl 2x B&W buffer [10mM Tris-Cl pH 7.5, 1mM EDTA, 2M NaCl]
800 µl Wash buffer [10mM Tris-Cl pH 7.5, 1mM EDTA, 50 mM NaCl], pre-warmed to 50^{o} C
200 µl 1x superscript III buffer [40 µl 5x buffer + 12 µl 0.1 M DTT + 148 µl water]
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)

Proceedure:
- Prep beads
    - Aliquot 100 µl beads and magnetize for 30~60 seconds.
    - Discard supernatant and resuspend in 100 µl bead prep buffer to wash (rotate for 2' at room temp).
    - 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 B&W buffer.
    - Resuspend beads in 200 µl B&W buffer.
- Bind and wash RNA
    - Resuspend RNA from step 3 above in 200 µl water.
    - Add to prepared beads, rotate at room temp 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^{o} C for 2 minutes, occasionally inverting or gently vortexing the tube to mix.
    - Repeat wash in 400 µl wash buffer at 50^{o} C.
- Release tags into supernatant
    - Wash beads in 200 µl 1x superscript III buffer at room temp 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^{o} C for 20 minutes in PCR machine.
    - Ramp down to 37^{o} C, then add 1 µl RNase H.
    - Incubate for 25 minutes, gently vortex occasionally to mix beads.
    - Add 25 µl water to beads and transfer to a 1.5 ml tube.
    - Magnetize for 1 minute and recover the supernatant (contains end tags).
    - Ethanol precipitate supernatant with glycoblue at –20^{o} C for at least 2 hours.

Step 5a: Library preparation - 3´ ligation

Materials required (per sample):
7.5 µl RNA (tags from step 4, resuspended in 7.5 µl water)
1 µl 100 µM adenylated 3´ solexa adaptor (3´ ligation adaptor)
1 µl 10x ligase buffer (–ATP) [500 mM Tris-Cl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT]
0.5 µl T4 RNL1 (neb)

Procedure:
- Combine the above reagents and incubate at 22° C (room temp) for 2 hours.
- Add 0.75 µl glycoblue, .63 µl 5M NaCl, 30 µl Ethanol. Precipitate at –20° C for at least 2 hours.
- Gel purify RNA ranging from 75 nt ~ 300 nt* on a denaturing 6% acrylamide gel. The gel is run short, such that the bromophenol blue dye migrates ~5 cm. This will limit the size of the gel slab excised.
    - Elute overnight in 800 µl 0.3 M NaCl.
    - Split eluate in half and precipitate each half with 1 µl Glycoblue and 2.5 volumes ethanol at least 2 hours.
    - I use radiolabeled RNA markers run in separate lanes to estimate sizes. Decade and century markers from Ambion cover the desired range.
    - I run thin gels to minimize acrylamide and maximize elution.

* n.b. – a larger range is more inclusive from a tag diversity perspective, but will make estimating the library molar concentration less accurate.

Step 5b: Library preparation – phosphorylate 5´ ends

Materials required (per sample):
21 µl Ligated RNAs (from step 5a resuspended in 21 µl water)
2.5 µl 10x T4 DNA ligase buffer (NEB)
.5 µl Rnase inhibitor
1 µl T4 PNK

Procedure:
- Combine the above reagents and incubate at 37° C for 1 hour.
- Phenol/chloroform extract and precipitate for at least 2 hours at –20° C.

Step 5c: Library preparation - 5´ ligation

Materials required (per sample):
pellet Kinased RNA (from step 5b, pelleted)
1 µl 100 µM 26.71 (5´ ligation adaptor)
0.5 µl 10x T4 RNL1 buffer (includes ATP, from NEB)
0.5 µl T4 RNL1 (NEB)
0.1 µl RNasin
2.9 µl Water

Procedure:
- Combine the water, buffer, RNase inhibitor and 5´ adaptors in a PCR tube.
- Use this mix to resuspend the RNA pellet from step 5b.
- Add 0.5 µl T4 RNL1, and incubate overnight (~16 hrs) at 22° C.
- Gel purify RNA ranging from ~100 nt ~ 325 nt on a denaturing 6% acrylamide gel as in step 5a.
    - The reaction volume is low so you can load the entire reaction without precipitation.
    - Elute overnight in 800 µl 0.3 M NaCl.
    - Split eluate in half and precipitate each half with 1 µl Glycoblue and 2.5 volumes ethanol at least 2 hours.

Step 5d: Library preparation – RT-PCR

Materials required (per sample):

14.6 µl RNA from step 5c resuspended in 14.6 µl 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 Superscript III
5 µl 1M NaOH
25 µl 1M HEPES pH 7
1 µl 20 µM solexa pcr fwd primer
1 µl 20 µM solexa pcr rev primer
20 µl 5x pcr buffer
1 µl Phusion hot start DNA polymerase (NEB/Finnzymes)
1 G-25 desalting column
sybr gold stain (Invitrogen)
minelute gel purification kit (Qiagen)

Procedure:

- Mix the following in a PCR tube:
  - 14.6 µl RNA
  - 1 µl 100 µM solexa rt primer
- Heat at 65° C for 5 minutes, then place on ice for 2 minutes.
- Add the following:
  - 6.4 µl 5x RT buffer
  - 3 µl 0.1 M DTT
  - 7 µl 10x dNTPs
- Incubate at 48° C for 3 minutes.
- Remove 3 µl and freeze for a no RT control (for PCR).
- Add 1 µl Superscript III and incubate at 48° C for 1 hour.
- Add 5 µl 1 M NaOH and incubate at 90° C for 10 minutes.
- Place on ice for 1 minute to cool, then add 25 µl 1M HEPES pH 7.
- Desalt 30 µl of the RT reaction over G-25, freeze the rest.
- 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 µl for no RT control)
  - 1 µl 20 µM solexa pcr fwd primer
  - 1 µl 20 µM solexa pcr rev primer
  - 1 µl phusion
- Amplify for 16–20 cycles with the following program:
  - 1 - 98° C x 1 min
  - 2 - 98° C x 15 sec |
  - 3 - 60° C x 30 sec | x 15–19 extra cycles
  - 4 - 72° C x 30 sec |
  - 5 - 72° C x 10 minutes
  - 6 - hold at 4° C
- Ethanol precipitate for at least 1 hour.
- Gel purify 150 ~ 375 bp range.
  - Run a 2% agarose gel and stain with sybr gold.
  - Purify using Qiagen's minelute gel extraction kit.
  - Elute in EB.

Step 6: Library sequencing and analysis considerations

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 1–5 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.