Chapter 35

Genome Wide Full-Length Transcript Analysis Using 5′ and 3′ Paired-End-Tag Next Generation Sequencing (RNA-PET)

Xiaoan Ruan and Yijun Ruan

Abstract

RNA-PET is a paired end tag (PET) sequencing method for full-length mRNA transcripts analysis using the next generation sequencer platforms such as Illumina GA and SOLiD. Unlike RNA-Seq method that sequences randomly sheared shotgun RNA short fragments, RNA-PET captures and sequences the 5′ and 3′ end tags of full-length cDNA fragments of all expressed genes in a biological sample. When mapped to reference genome, RNA-PET sequences can demarcate the boundaries of transcription units genome-wide, in addition to its ability to quantify the transcription level of each expression genes. Furthermore, the unique feature of RNA-PET is to identify fusion transcripts. Therefore, RNA-PET has been regarded as the best PET for genome annotation (1). Here in this chapter, we describe the details of the RNA-PET protocol and discuss the critical issues.

Key words: NGS, PET sequencing, Full length mRNA, Gene expression, Illumina GA, SOLiD, Transcriptome, RNA boundary, Splice variants, Fusion transcripts, EcoP15I

1. Introduction

Genomics holds much promise for huge improvements in human healthcare, and the next-generation sequencing technologies are becoming a driving force that penetrates the entire field of genomic science. As the current sequencing technologies are limited by short sequencing reads, an important part of the sequencing strategy is to use the PET sequencing approaches to analyze nucleic acid templates such as RNA for transcriptome (1) and DNA for genome (2).

To fully understand gene transcription regulation in the whole genome content, it is important to define where precisely gene transcription starts and terminates. To obtain such information, we developed an efficient strategy to demarcate the boundaries of

Ales Vancura (ed.), *Transcriptional Regulation: Methods and Protocols*, Methods in Molecular Biology, vol. 809,
DOI 10.1007/978-1-61779-376-9_35,

transcription units for the whole genome (1). The core concept is to obtain only linked 5′ and 3′ short tag sequences for each transcript, map these terminal "signatures" to the genome and thereby infer the complete transcription units by the genome sequence encompassed between these 5′ and 3′ signatures. As an intermediate step, we first developed the 5′ and 3′ LongSAGE protocols to capture the 5′ and 3′ tag sequences separately of expressed genes (3). With this ability, we then combined these two separate protocols into one for extracting the paired end 5′ and 3′ tags for sequencing and mapping analysis (4).

In the early version of the PET sequencing for full-length mRNA transcripts, short tag fragments of 20 bp (5′) and 20 bp (3′) were extracted through bacterial cloning-based process, and the paired ditag fragments were concatenated into longer DNA fragments for Snager capillary sequencing (AB3730xl). Later, we adapted the Roche 454 GS-FLX pyrosequencer for such analysis (5). However, the bacterial-based cloning method for PET extraction was a long and laborious process, and the 20-bp tag information is a limitation for high mapping specificity. We now further improved the PET analysis protocol by developing a complete in vitro cloning-free protocol and adapting the enzyme EcoP5I for extracting longer PET fragments (27–27 bp).

This new version of PET analysis for full-length mRNA (we called RNA-PET) consists of six major steps (Fig. 1): (1) Capture and synthesize full-length cDNAs from mRNA transcripts using Cap-Trapper approach and a specifically designed GsuI-dT (16) oligo for reverse transcription (RT). (2) Ligate the captured full-length cDNAs with specific DNA linkers. (3) Circularize the linker-ligated full-length cDNAs and excise 27 bp of terminal tags from each of the 3′ and 5′ end. (4) Isolate and purify the PETs and ligate with next generation sequencer adaptors. (5) Paired End (PE) sequencing the captured ends with either Illumina GA or SOLiD v4, and mapping both sequence tags back to the reference genome. (6) Cluster and annotate the tags onto existing reference database and visualize the annotated sequence data on genome browser, and produce digital gene expressions as well as detect unusual gene transcripts caused by a variety of genome rearrangements or abnormal transcription events.

To start RNA-PET, total RNA sample is used as starting material and poly (T) Oligos are used to enrich and purify mRNAs. Approximately 1–5 μg poly (A) mRNA is used in the RNA-PET library construction. The Cap-Trapper approach (6–8) is combined with GsuI-poly(T) Oligos to capture full-length cDNAs. After full-length cDNAs are obtained, the full-length cDNA are methylated to block EcoP15I recognition site at the fifth residue A. The cDNA is then ligated to the specifically designed linker sequences and circularized at larger ligation volume (~0.1 ng DNA/μl). Un-circularized molecules are further cleaved out by using a plasmid-save treatment and the remaining circular cDNAs are gone through EcoP15I

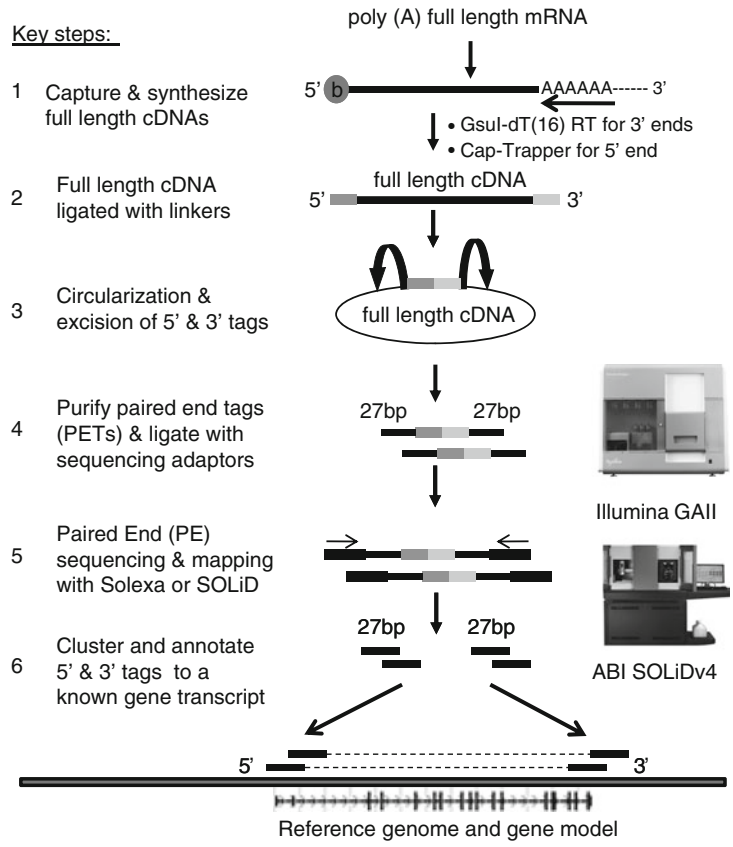

Fig. 1. Schematic overview of the RNA-PET analysis for full-length whole transcriptome. Six sections of steps are involved in the RNA-PET analysis as illustrated and include full-length cDNA capturing, PET excision, sequencing, mapping, and annotation analysis.

digestion to release the PETs at 27 bp from each end. The resulting PETs are modified with specific paired end (PE) sequencing adaptors compatible to either Illumina GAII or SOLiD4. After ligated to the sequencing adaptors, the PET template are further PCR amplified and sequenced through PE sequencing format. Approximately 20–30 million sequence PETs are generated through HT sequencing. After filter out redundant and noise tags, the unique PETs proceed to mapping on the specific reference genome. A approximate 90% of PETs, mapped to known transcripts or splicing variants, are named as *concordant* PETs, however, a small portion of the misaligned PETs (named as *discordant* PETs) are mapped either to wrong orientations in the same chromosome, or to different strands, or on different chromosomes. For *concordant* PETs, the digital expression level can be easily obtained from mapped sequence counts. Even though a majority of the *discordant* PETs are derived from noise ligations, a collection of these PETs serves as a valuable pool for identification of many novel transcripts and its associated splice variants.

It should be noted that as RNA-PET only characterizes the 5′ and 3′ ends of transcripts whereas RNA-seq is robust for tagging internal exons but poor on transcript terminal regions, the combination of RNA-PET and RNA-seq should be viewed as the ultimate solution for comprehensive transcriptome characterization.

2. Materials

2.1. Poly (A) mRNA and GsuI-dT (16) Oligo for RT

1. 1–5 μg of poly (A) mRNA isolated from total RNA using Miltenyi Biotec μMACs mRNA Isolation Kit.
2. GsuI-dT(16)oligo:5′-GAGCTAGTTCTGGAGTTTTTTTTT TTTTTTTVN-3′ in 1 μg/μl, stored in −20°C.
3. DNA low-bind tube used throughout DNA or RNA procedures (Eppendorf, see Note 1).
4. 3 M sodium acetate (NaOAc), pH 5.5.
5. Isopropanol.
6. 75% Ethanol (EtOH, see Note 2).
7. RNasin-Plus RNase inhibitor (Promega).
8. 2× GC-I Buffer (Takara).
9. dNTP mix (with 5-Me-dCTP in replace of dCTP): 10 mM dATP, 10 mM dTTP, 10 mM dGTP, 5 mM 5-Me-dCTP in 10 mM Tris–HCl, pH 8.0.
10. 4.9 M D-sorbitol (Sigma).
11. Superscript II and III reverse transcriptases (Invitrogen).
12. Saturated trehalose (RNase-free) (Sigma): Heat water in a 1.7-ml tube to 42°C in a heat block. Slowly add trehalose powder to the tube and dissolve it by vortex. Maintain temperature at around 42°C and continue add trehalose until saturation is reached. Upon cooling down the solution to room temperature, trehalose will form crystals and saturated solution is obtained (see Note 3). Aliquot the solution and store in −20°C.
13. Proteinase K 20 mg/ml (Ambion).
14. Phenol:Chloroform:IAA solution, 25:24:1, pH 6.6 (Ambion).
15. Water (nuclease-free).

2.2. Reagents for Oxidation of Diol Structures and Biotinylation of 5′ End

1. Sodium periodate ($NaIO_4$) (Sigma).
2. 1.1 M NaOAc, pH 4.5: Diluted from 3 M NaOAc, and adjusted to pH 4.5.
3. 10% SDS solution.
4. 5 M NaCl.

5. 10 mM biotin hydrazide (long arm) (Vector Laboratories).
6. 1 M NaOAc, pH 6.1: Diluted from 3 M NaOAc and adjusted to pH 6.1.

2.3. RNase One Selection

1. 5 M NaCl.
2. 1 M NaOAc, pH 6.1.
3. 10× RNase ONE buffer (Promega).
4. RNase ONE Ribonuclease 10 U/μl (Promega).
5. Yeast tRNA 10 mg/ml (Ambion).

2.4. Prepare Dynalbead M-280 Streptavidin

1. Dynal MPC-S (now replaced by DynalMag-2) (Invitrogen).
2. Dynalbead M-280 Streptavidin (Invitrogen).
3. 1× Binding buffer (BB): 2 M NaCl, 50 mM EDTA, pH 8.0.
4. 1× BB + Yeast tRNA: 2 M NaCl, 50 mM EDTA, pH 8.0, Yeast tRNA 0.25 μg/μl.
5. Eppendorf thermomixer.

2.5. Binding of Full-Length Biotinylated (–)DNA/RNA Heteroduplex

1. Intelli Mixer (Elmi).
2. Dynal MPC-S (now replaced by DynalMag-2) (Invitrogen).
3. 1× BB: 2 M NaCl, 50 mM EDTA, pH 8.0.
4. 1× Blocking buffer: 0.4% SDS, 50 μg/ml Yeast tRNA.
5. 1× Washing buffer: 10 mM Tris–HCl, pH 7.5, 0.2 mM EDTA, 10 mM NaCl, 20% glycerol, 40 μg/ml Yeast tRNA.

2.6. Hydrolytic Degradation of Bound RNA to Release Full Length (–) cDNA Strand

1. Alkaline hydrolysis buffer: 50 mM NaOH, 5 mM EDTA, pH 8.0. Prepare fresh each time.
2. 2 ml MaXtract High Density Tube (Qiagen).
3. Phenol:Chloroform:IAA, 25:24:1, pH 7.9 (Ambion).
4. Buffer EB (Qiagen).
5. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.

2.7. Synthesis of Double-Stranded Full-Length cDNA

1. 1× Tris–NaCl–EDTA (TNE) buffer: 10 mM Tris–HCl, pH 8.0, 50 mM NaCl, 0.1 mM EDTA. Sterilize the buffer by syringe filter or autoclave.
2. DNA linkers for Cap-Trapper of 5′-end of the full-length cDNAs, and double-stranded cDNA synthesis. Linker E-E2-GsuI-N5 and linker E-E2-GsuI-N6 (see Table 1 for sequence detail and Subheading 3.9.1 for oligo annealing).
3. Oligos are synthesized by HPLC purification from Integrated DNA Technologies (IDT). Oligos are annealed to dsDNA and stored in aliquots in –20°C at 0.4 μg/μl for use.
4. 4–20% TBE gels (Invitrogen).

Table 1
Sequences for DNA linkers and adptors used in RNA–PET construction

Oligos and DNA adaptors	Sequences	Length (nt)
Linkers for capture of 3′-end of full length poly A mRNA and RT synthesis:		
GsuI-dT (16) oligo	5′-GAGCTAGTTCTGGAGTTTTTTTTTTTTTTTTVN-3′	33
Linkers for Cap-Trapper of 5′-end of full length cDNA and second strand cDNA synthesis:		
Linker E-E2-GsuI-N5 (top)	5′-CTACCTGGAGAACATGAGGCAGCCAGGNNNNN-3′	32
Linker E-E2-GsuI-N5 (bottom)	5′-Phos-CTGGCTGCCTCATGTTCTCCAGGTAG-3′	26
Linker E-E2-GsuI-N6 (top)	5′-CTACCTGGAGAACATGAGGCAGCCAGNNNNNN-3′	32
Linker E-E2-GsuI-N6 (bottom)	5′-Phos-CTGGCTGCCTCATGTTCTCCAGGTAG-3′	26
Linkers for 5′ and 3′ paired end tag (PET) capture:		
Illumina E-E2-h-AGG linker (top)	5′-GGCCGCGA*T*ATCGGACAGCAG-3′ (Internal Biotin at the ninth d*T*)	21
Illumina E-E2-h-AGG linker (bottom)	5′-Phos-GCTGTCCGATATCGC-3′	15
Illumina E-E2-t-AGG linker (top)	5′-Phos-CTGCTGTCCGATATCGC-3′	17
Illumina E-E2-t-AGG linker (bottom)	5′-GGCCGCGA*T*ATCGGACAGCAGTT-3′ (Internal Biotin at the ninth d*T*)	23
SOLiD linker E5v4 (top)	5′-CCGCC*T*TGGCCGTACAGCAG-3′ (Internal Biotin at the sixth d*T*)	20
SOLiD linker E5v4 (bottom)	5′-Phos-GCTGTACGGCCAAG-3′	14
SOLiD linker E3v3 (top)	5′-GCGGA*T*GTACGGTACAGCAGTT-3′ (Internal Biotin at the sixth d*T*)	22
SOLiD linker E3v3 (bottom)	5′-Phos-CTGCTGTACCGTACAT-3′	16
Sequencing adaptors and PCR amplification primers for Solexa GA, or SOLiD V4:		
Illumina 454-adaptor E/A (top)	5′-CCATCTCATCCCTGCGTGTCCCATCTGTTCCCTCCCTGTCTCAG-3′	44

Illumina 454-adaptor E/A (bottom)	5-NNCTGAGACAGGGAGGGAACAGATGGGACACGCAGGG-ATGAGATGG-3′	46
Illumina 454-adaptor E/B (top)	5′-GACTCTGTGCGTTGTCCCCTATCCGTTCCGTGTGTCCC-CTATCC-3′	44
Illumina 454-adaptor E/B (bottom)	5′-GGATAGGGGACACACGGAACGGATAGGGGACAACGCAC-AGAGTCNN-3′	46
SOLiD P1 adaptor (dsDNA bought from ABI)	5-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3′ 3′-TTGGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA-5′	41/43
SOLiD P2 adaptor (dsDNA bought from ABI)	5′-AGAGAATGAGGAACCCGGGGCAGTT-3′ 3′-TCTCTTACTCCTTGGGCCCCGTC-5′	25/23
RNA-PET library PCR primers:		
Illumina-454 PCR primer-1	5′-AATGATACGGCGACCACCGAGATCTACACCCTATCCCC-TGTGTGCCTTG-3′	49
Illumina-454 PCR primer-2	5′-CAAGCAGAAGACGGCATACGAGATCGGTCCATCTCAT-CCCTGCGTGTC-3′	48
SOLiD library PCR primer-1 (bought from ABI)	5′-CCACTACGCCTCCGCTTTCCTCTCTATG-3′	28
SOLiD library PCR primer-2 (bought from ABI)	5′-CTGCCCCGGGTTCCTCATTCT-3′	21

5. Takara Solution I (Takara).
6. Takara Solution II (Takara).
7. 10× ExTaq buffer (Takara).
8. Takara ExTaq (Takara).
9. dNTP mix at 2.5 mM each (Takara).
10. Proteinase K at 20 mg/ml (Ambion).
11. 2 ml MaXtract High Density Tube (Qiagen).
12. Phenol:Chloroform:IAA, 25:24:1, pH 7.9 (Ambion).
13. Glycoblue, 15 mg/ml (Ambion).
14. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.
15. 70% EtOH in nuclease-free water.
16. GsuI, 5 U/μl (Fermentas).
17. cDNA size fractionation columns (Invitrogen).
18. 6× loading dye (Fermentas).
19. TEN buffer: 10 mM Tris–HCl, pH 8.0, 0.1 mM EDTA, pH 8.0, 25 mM NaCl.
20. Molecular probes SYBR Green, 10,000× in DMSO (Invitrogen).
21. Gel staining buffer: 1× TBE, Molecular probes SYBR Green 1×.
22. Gel handler, gel support (Sigma).
23. Molecular probes Quant-iT PicoGreen dsDNA reagent (Invitrogen).
24. Dynalbead M-280 Streptavidin beads (Invitrogen).
25. Calf thymus DNA (Sigma).
26. S-adenosylmethionine (SAM), 0.5 mM (NEB).

2.8. Double Strand cDNA Methylation by EcoP15I Enzyme (Recognition:CAGCAG)

1. EcoP15I enzyme (NEB).
2. 10× Buffer 3 (NEB).
3. 100× BSA (NEB).
4. SAM, 32 mM (NEB).
5. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.

2.9. Linker Ligation for 5′ and 3′ Paired End Tag Capture of ds Full-Length cDNAs

1. 5× T4 DNA ligase buffer with PEG (Invitrogen).
2. T4 DNA ligase, 30 U/μl (Fermentas).
3. Solexa E-E2-h-AGG linker and Solexa E-E2-t-AGG linker each at 200 ng/μl (see Table 1 for sequence detail and Subheading 3.9.1 for oligo annealing).
4. SOLiD linker E5v4 and SOLiD linker E3v3 each at 200 ng/μl (see Table 1 for sequence detail and Subheading 3.9.1 for oligo annealing).
5. 2 ml MaXtract High Density Tube (Qiagen).

6. Phenol:Chloroform:IAA, 25:24:1, pH 7.9 (Ambion).
7. Glycoblue, 15 mg/ml (Ambion).
8. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.

2.10. Add Phosphate Group at 5′-Ends of Linker-Ligated DNA Fragments

1. 10× T4 DNA ligase (NEB).
2. T4 DNA polynucleotide kinase, 10 U/μl (NEB).
3. Fermentas T4 DNA ligase (30 U/μl).
4. Isopropanol precipitation: 3 M NaOAc, pH 5.5, Isopropanol.

2.11. Circularization of Linker-Ligated cDNA Fragments

1. 10× T4 DNA ligase (NEB).
2. T4 DNA ligase, 30 U/μl (Fermentas).
3. 15 ml MaXtract High Density Tube (Qiagen).
4. Phenol:Chloroform:IAA, 25:24:1, pH 7.9 (Ambion).
5. Glycoblue, 15 mg/ml (Ambion).
6. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.
7. Buffer EB (Qiagen).

2.12. Nick Translation Repair of DNA

1. 10× *Escherichia coli* DNA ligation buffer (Qiagen).
2. 10 mM dNTP mix (Eppendorf).
3. *E. coli* ligase (NEB).
4. *E. coli* DNA polymerase I (NEB).
5. 2 ml MaXtract High Density Tube (Qiagen).
6. Phenol:Chloroform:IAA, 25:24:1, pH 7.9 (Ambion).
7. Glycoblue, 15 mg/ml (Ambion).
8. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.

2.13. Plasmid-Safe Treatment to Cleave Linear DNA Molecules

1. Plasmid-safe reaction: 25 mM ATP, 10× reaction buffer, Plasmid-safe DNase, 10 U/μl (Epicentre).
2. 2 ml MaXtract High Density Tube (Qiagen).
3. Phenol:Chloroform:IAA, 25:24:1, pH 7.9.
4. Glycoblue, 15 mg/ml.
5. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.

2.14. EcoP15I Digestion to Release 5′ and 3′ Paired End Tags from Full Length cDNAs

1. 10× Buffer 3 (NEB).
2. 100× BSA (NEB).
3. 10 mM Sinefungin (Calbiochem).
4. Adenosine 5′-triphosphate, (ATP) (NEB).
5. EcoP15I enzyme 10 U/μl (NEB).

2.15. Binding of 5′ and 3′ Pair End Tags to Dynalbeads M280 Streptavidin

1. Intelli Mixer (Elmi).
2. Dynal MPC-S (now DynalMag-2): Magnetic stand (Invitrogen).
3. Dynalbeads M-280 Streptavidin (Invitrogen).
4. 2× B&W buffer: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 2 M NaCl.
5. 1× B&W buffer: 5 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl.

2.16. Ligation of Sequencing Adaptors to 5′ and 3′ PETs

1. Solexa 454-adaptor E/A and Solexa 454-adaptor E/B each at 200 ng/μl.
2. SOLiD P1 and SOLiD P2 adaptor each at 200 ng/μl (see Table 1 for sequence detail).
3. T4 DNA ligase (30 U/μl) (Fermentas).
4. 10× T4 DNA ligase buffer (NEB).
5. Intelli Mixer (Elmi).
6. 1× B&W buffer: 5 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl.

2.17. Nick Translation for DNA Repair

1. 10× Buffer 2 (Qiagen).
2. 10 mM dNTP mix (Eppendorf).
3. *E. coli* DNA polymerase I (NEB).
4. Nuclease-free water (Ambion).
5. Buffer EB (Qiagen).
6. Intelli Mixer (Elmi).
7. 1× B&W buffer: 5 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl.

2.18. PCR Amplification for PET Sequencing Templates

1. 2× Phusion High-Fidelity PCR Master Mix with HF Buffer (Finnzymes).
2. Molecular probes SYBR Green 1, 10,000× in DMSO (Invitrogen).
3. Gel staining buffer: 1× TBE, Molecular probes SYBR Green 1.
4. Gel handler gel support (Sigma).
5. Solexa-454 PCR primer-1 and Solexa-454 PCR primer-2, each at 25 μM.
6. SOLiD library PCR primer-1 and SOLiD library PCR primer-2, each at 2 μM (see Table 1 for sequence detail).
7. 4–20% TBE gels (Invitrogen).
8. 25-bp DNA ladder at 1 μg/μl (Invitrogen).
9. 1× TBE buffer (First Base Inc.) diluted from 10× stock.

2.19. Scale-Up PCR Amplification for Sequencing Templates

1. 2× Phusion High-Fidelity PCR Master Mix with HF Buffer (Finnzymes).
2. Molecular probes SYBR Green 1, 10,000× in DMSO (Invitrogen).
3. Gel staining buffer: 1× TBE, Molecular probes SYBR Green 1.
4. Gel handler gel support (Sigma).
5. Solexa-454 PCR primer – 1 and 2.
6. SOLiD library PCR primer – 1 and 2.
7. 6% TBE gel with 5-wells (Invitrogen).
8. 25-bp DNA ladder at 1 μg/μl (Invitrogen).
9. 1× TBE buffer (First Base Inc.) diluted from 10× stock.
10. QIAquick PCR Purification Kit (Qiagen).

2.20. Purification of PCR Templates Using Gel-Crush Elution Method

1. Spin-X centrifuge tube filters, CA membrane, 0.22 pore size (Costar).
2. TE buffer, pH 8.0 (Ambion).
3. 21-G needle (Becton Dickinson).
4. 0.6-ml microtube (Axygen).
5. 1.5-ml screw cap tube (Axygen).
6. Glycoblue, 15 mg/ml (Ambion).
7. 70% Ethanol in nuclease-free water.
8. Nuclease-free water (Ambion).
9. Isopropanol precipitation: 3 M NaOAc, pH 5.5, isopropanol.

3. Methods

3.1. Mix Poly (A) mRNA and GsuI-dT (16) Oligo

1. Mix following reagents in order in a 0.2-ml PCR tube on ice (see Notes 4–6):

Poly (A) mRNA (1–5 μg)	10 μl containg 1 μl poly (A) mRNA
GsuI-dT oligo (1 μg/μl)	0.7 μg oligo
3 M NaOAc, pH 5.5	1/10 volume
Isopropanol	Equal volume

2. Keep tube at –80°C for 30 min.
3. Centrifuge at 24,000×*g* for 30 min at 4°C in a microcentrifuge.
4. Wash 2× with 500 μl of cold 75% EtOH.

5. Air dry pellet, resuspend pellet in 19-μl of nuclease-free water, and add 1 μl of RNase-Plus inhibitor to the mixture and transfer solution into a 0.2-ml PCR tube.

3.2. Set Up Reverse Transcription (RT)

1. Heat the reaction mixture at 65°C for 10 min and cool to 37°C for 1 min, then hold at 42°C on a thermal cycler while waiting for the other components to be prepared.
2. Set up the RT mix on ice in a 0.2-ml thin-walled PCR tube:

2× GC-I buffer	75 μl
RNasin-Plus RNase inhibitor	1 μl
10 mM dNTP (with 5-Me-dCTP in place of dCTP)	4 μl
4.9 M sorbitol	26 μl
Superscript II reverse transcriptase	8 μl
Superscript III reverse transcriptase	4 μl

3. Put 10 μl of saturated trehalose into another 0.2-ml PCR tube and leave warming at 42°C in a thermal cycler.
4. When the oligo-dT/mRNA annealing step is complete, place the RT mix into the thermal cycler that was preset at 42°C for at least 2 min.
5. Mix the warm trehalose together with the RT mix (volume = 128 μl), and quickly transfer the entire reaction mix into the tube containing the annealed oligo/mRNA (volume now = 148 μl), and immediately start the incubation:

42°C for 50 min
50°C for 25 min
55°C for 25 min
Hold at 4°C

6. Add 2 μl of proteinase K to digest all enzymes by incubation at 45°C for 15 min.
7. Transfer the solution into a lo-bind tube and add equal volume of Phenol:Chloroform:IAA, 25:24:1, pH 6.6 directly into the tube.
8. Mix the organic and aqueous phases thoroughly for about 1 min.
9. Centrifuge at 24,000×*g* for 3 min in a microcentrifuge to separate the phases.
10. Remove the upper, nucleic acid-containing phase by carefully pipetting into a new tube.
11. Repeat the extraction with another 150 μl of nuclease-free water.
12. Combine two aqueous phases together.

13. Isopropanol precipitates the RNA/DNA heteroduplex.
14. Resuspend the pellet in 44.5 µl of nuclease-free water.

3.3. Oxidation of Diol Structures

1. Prepare the following stocks fresh each time, using 1.7-ml tubes:

10 mM biotin hydrazide (long arm)
200 mM $NaIO_4$

2. In a 1.7 ml siliconized tube, add the following and mix well:

RNA/DNA heteroduplex	44.5 µl
1.1 M NaOAc, pH 4.5	3 µl
Fresh 200 mM $NaIO_4$	2.5 µl

3. Incubate reaction mix on ice for 45 min in the dark.
4. Isopropanol precipitates the DNA/RNA heteroduplex.
5. Resuspend the pellet in 50 µl of nuclease-free water.

3.4. Biotinylation of 5′-End of mRNA

1. To the 50 µl oxidized (−) cDNA/RNA heteroduplex, add the following:

1 M NaOAc pH 6.1	5 µl
10% SDS	5 µl
Fresh-made 10 mM biotin hydrazide	150 µl

2. Incubate at room temperature (RT) for overnight and keep in the dark.
3. Isopropanol precipitates the 210 µl biotinylated (−) cDNA/RNA heteroduplex.
4. Resuspend the pellet in 170 µl of nuclease-free water.
5. Preset the Eppendorf shaking incubator to cool down to 4°C.

3.5. RNase ONE Selection

1. Set up the following in a 1.7-ml lo-bind tube (see Note 7):

Biotinylated (−) DNA/RNA sample:	170 µl
10× RNase ONE buffer	20 µl
RNase ONE Ribonuclease (10 U/µl)	xx µl
Nuclease-free water	Add up to 200 µl

2. Incubate at 37°C for 30 min (see Note 8).
3. Quench the reaction by adding:

10 mg/ml Yeast tRNA	4 µl
5 M NaCl	50 µl

4. Leave the RNase ONE-treated sample on ice.

3.6. Preparation of Dynalbeads M-280 Streptavidin

1. Use the magnetic stand and lo-binding tubes for all steps involving M-280 beads. Use 200 µl of the Dynalbeads M-280 Streptavidin suspension per RNA sample (see Note 9).
2. Wash Dynalbead to remove preservatives (see Note 10).
3. Wash Dynalbeads 3× with 200 µl of 1× Binding buffer at room temperature (RT).
4. Add 200 µl of 1× BB + Yeast tRNA to the beads and incubate at 4°C for 30 min using Eppendorf shaker at 800 rpm (pre-cooled to 4°C, see Subheading 3.5, step 5).
5. Wash beads 3× with 1× BB at RT.

3.7. Binding of Full-Length Biotinylated (–) DNA/RNA Heteroduplex

1. Remove the supernatant (S/N) from beads and add the chilled RNase ONE-treated sample to the beads.
2. Rotate 30 min at RT on an Intelli-Mixer for binding to occur (Program F8: U = 50, u = 60, 30 rpm) (see Notes 11 and 12).
3. Wash the heteroduplex-bound beads at RT as follows:

2× with 200 µl of 1× BB
1× with 200 µl of 1× Block
1× with 200 µl of 1× Wash
1× with 200 µl of 50 µg/ml Yeast tRNA.

3.8. Hydrolytic Degradation of Bound RNA to Release Full-Length (–) cDNA Strand

1. Prepare alkaline hydrolysis buffer (see Note 13).
2. Prepare a tube containing 150 µl of 1 M Tris–HCl, pH 7.5 (for neutralization).
3. Remove S/N (50 µg/ml Yeast tRNA) from the beads and add 50 µl of alkaline hydrolysis buffer.
4. Shake the mixture at 65°C for 10 min using the Eppendorf shaker at 2,400×*g*.
5. Collect the S/N containing full-length (–) cDNA into the tube containing 150 µl of 1 M Tris–HCl, pH 7.5, for neutralization.
6. Repeat the hydrolysis and collection steps twice, collecting all fractions into the same tube to a final volume of 300 µl.
7. Immediately before use, pellet MaXtract High Density Tube by centrifugation at 19,300×*g* for 30 s in a microcentrifuge (see Note 14).
8. Transfer the DNA into the MaXtract tube and directly add equal volume of Phenol:Chloroform:IAA, 25:24:1, pH 7.9, to the MaXtract tube.
9. Mix the organic and aqueous phases thoroughly for about 1 min.
10. Centrifuge at 19,300×*g* for 3 min in a microcentrifuge to separate the phases.

11. Remove the upper, nucleic acid-containing phase by carefully pipetting into a new tube.
12. Isopropanol precipitates the DNA and resuspend the pellet in 5 μl of Buffer EB.

3.9. Synthesis of Double-Stranded cDNA

3.9.1. Annealing Oligos for Double Strand DNA Linkers (see Note 15)

1. Thaw single strand oligos at room temperature for 15 min.
2. Spin at maximum speed at 4°C for 1 min to collect any dislodged oligos.
3. Add 1× TNE buffer to thawed oligos to make it 100 μM.
4. Vortex for approximately 1 min to resuspend oligos and brief spin down to collect the resuspended oligos to the bottom of tubes.
5. Perform Nanodrop and use OD constant value to calculate concentration. Check if the measured concentration by Nanodrop falls within the expected range.
6. In a 0.2-ml PCR tube, mix together the Oligos (see Table 1 for sequences), e.g.,

Oligonucleotide A (top strand) (100 μM)	40 μl
Oligonucleotide B (bottom strand) (100 μM)	40 μl

7. Heat 95°C for 10 min, then, turn off the program with lid closed, and let the tube cool down to room temperature slowly to allow Oligos annealed together. It may take 90 min to complete the naturally cool-down process. For long term storage, keep the annealed Oligos (or called dsDNA linkers or adaptors) in −80°C.
8. Keep all annealed dsDNA linkers or adaptors on ice before use.
9. Measure the DNA concentration by Nanodrop and dilute it to a concentration of 200 ng/μl with 1× TNE.
10. Run 200 ng each of the single strand oligos together with 200 ng of annealed dsDNA linkers on the 4–20% PAGE gel to ensure the annealing result satisfactory.

3.9.2. 5′-End Cap-Trapper Ligation and Second Strand cDNA Synthesis

1. Set following reagents on ice in a 1.7-ml lo-bind tube:

Full-length single strand (−) cDNA	5 μl
0.4 μg/μl linker E-E2-GsuI-N5	4 μl
0.4 μg/μl linker E-E2-GsuI-N6	1 μl
Takara Solution II (A3101-1) (see Note 16)	10 μl
Takara Solution I (ligase, A201-1)	20 μl

2. Mix by flicking and short spin at 4°C.
3. Incubate at 16°C for overnight to allow degenerate oligo to anneal and ligate.

3.9.3. Primer Extension for Second Strand cDNA Synthesis

1. Set the following on ice in a 0.2-ml thin-walled PCR tube:

Overnight ligation mix	40 µl
Nuclease-free water	20 µl
10× ExTaq buffer with Mg^{2+}	8 µl
2.5 mM dNTP	8 µl
ExTaq polymerase	4 µl

2. Incubate the PCR reaction in the warmed up thermal cycler at 65°C for 5 min, 68°C for 30 min, 72°C for 10 min, and hold at 4°C.
3. Add 2 µl of proteinase K and mix by pipetting up and down and incubate at 45°C for 15 min to digest the enzyme.
4. Immediately before use, pellet MaXtract High Density Tube by centrifugation at 23,000×*g* for 30 s in a microcentrifuge.
5. Transfer the DNA into the MaXtract tube and adjust the reaction volume to 200 µl.
6. Purify DNA by Phenol:Chloroform:IAA as described in Subheading 3.8.
7. Precipitate the DNA with isopropanol and resuspend in 66.8 µl of nuclease-free water.

3.9.4. GsuI Digestion to Remove Poly (A) Tail and Produce 3′ Terminal Ends

1. Freshly dilute 32 mM SAM to 0.5 mM.
2. Set the following on ice in a 1.7-ml lo-bind tube:

Full-length double-stranded cDNA	66.8 µl
10× buffer TANGO with BSA (Fermentas)	8.6 µl
0.5 mM SAM (1 µl SAM + 63 µl dH_2O)	8.6 µl
GsuI (5U/µl, Fermentas, #ER0462)	2 µl

3. Mix by flicking and brief spin at 4°C.
4. Incubate at 30°C overnight.
5. Inactivate GsuI at 65°C for 20 min and transfer sample to ice.

3.9.5. Isolate Full-Length cDNAs by Size-Fractionation and Separation from Linkers

1. Prepare size fractionation column according to manufacturer's instructions while sample is held at 65°C for 20 min. Equilibrate column to room temperature before use (see Note 17). Remove top cap first, then bottom and then allow liquid to drain off completely. Add 0.8 ml TEN buffer and allow it to drain off. Repeat the washes three times (see Note 18).
2. Label twenty 1.7-ml tubes for fractionation purpose.
3. Add 2 µl of 6× loading dye to the cDNA sample and keep it on ice.

4. Transfer the overnight digested mixture into the prepared column and collect the entire flow out into the first collection tube.
5. Add 100 µl of TEN buffer and collect the entire flow through into the second collection tube.
6. Add another 100 µl of TEN buffer and start to collect cDNA by single drop per tube into the third tube and so on until the 20th tube.
7. During the collection process allow each drain-off completely before adding the next 100 µl of buffer. It may need adding a total seven times at each with 100 µl of TEN buffer.
8. After finishing the size fractionation, the collected tubes from 9 through 18 will be selected to run on a 4–20% TBE gel at 200 V for 15 min to assess cDNA collection results. Usually the tubes from 3 to 13 will be pooled and purified as full-length cDNA going to next step (see Note 19). The tubes beyond 14–20 will be discarded as they usually contain linker DNA added.
9. The quantity of the purified full-length cDNA is measured using Quant-iT PicoGreen method following manufacturer's instruction.

3.10. Methylation of Full-Length cDNA Using Enzyme EcoP15I

1. Set the following reaction mix on ice in a 1.7-ml lo-bind tube:

Full-length double strand cDNA	Top up to 100 µl with dH2O
10× Buffer 3	10 µl
100× BSA	1 µl
32 mM SAM (1 µl SAM + 63 µl dH2O)	25 µl
EcoP15I enzyme (see Note 20)	10 U/µg cDNA

2. Incubate reaction mix at 37°C for overnight.
3. Immediately before use, pellet MaXtract High Density Tube by centrifugation at 14,000 rpm for 30 s in a microcentrifuge.
4. Transfer DNA into MaXtract tube, adjust the reaction volume to 200 µl with water.
5. Purify DNA by Phenol:Chloroform:IAA as described in Subheading 3.8.
6. Precipitate with isopropanol and resuspend cDNA in 50 µl of water.

3.11. Ligation of Sequencing Linkers to Full-Length cDNAs

1. Estimate the amount of linkers (h/or t) to be used with a formula below:

$$[(\text{ng cDNA} \times 200 \times 20\text{bp}) / 2{,}500\ \text{bp}] / [200\text{ng/µllinkers}] = \text{µloflinkerstobeused}$$

(see Note 21).

2. Add the proper linkers and reagents on ice in order and incubate at 16°C for overnight.

Sequencing linker-1 (200 ng/μl) (see Note 15)	x μl (see formula above)
Sequencing linker-2 (200 ng/μl)	x μl (see formula above)
Full-length cDNA	50 μl
5× T4 DNA ligase buffer + PEG	40 μl
Nuclease-free water	Top up to 200 μl
T4 DNA ligase (see Note 22)	1 μl

3. Prepare MaXtract High Density Tube as previous procedures.
4. Transfer the DNA into the MaXtract tube.
5. Purify DNA by Phenol:Chloroform:IAA as described in Subheading 3.8.
6. Precipitate with isopropanol and resuspend DNA in 44 μl of water.

3.12. Addition of Phosphate Group to the 5′-Ends of Linker-Ligated Full-Length cDNAs

1. Set the following on ice in a 1.7-ml tube:

Linker-ligated full-length cDNAs	44 μl
10× T4 DNA ligase buffer	5 μl
T4 DNA polynucleotide kinase	1 μl (final conc. ~0.2 U/μl)

2. Incubate at 37°C for 30 min.

3.13. Circularization of Linker-Ligated cDNA by Ligation in a 5-ml Volume

1. Prepare the following enzyme reaction mix (5 ml) on ice into a 15-ml tube:

Nuclease-free water	4,425 μl
10× T4 DNA ligase buffer	495 μl
T4 DNA ligase	30 μl (final conc. ~0.18 U/μl)
Linker-ligated cDNA mix (from Subheading 3.12)	50 μl

   Transfer the premix (50 μl) described in Subheading 3.12 into the above ligase mix (15 μl tube).
2. Incubate the ligation at 16°C for overnight (see Note 23).
3. Purify DNA by Phenol:Chloroform:IAA as described in Subheading 3.8.
4. Precipitate the DNA with isopropanol (see Note 24) and resuspend it in 78 μl of Buffer EB.

3.14. DNA Nick Repair

1. Set the following reagents on ice in a 1.7-ml tube:

DNA in Buffer EB	78 µl
10× *E. coli* DNA ligation buffer	10 µl
10 mM dNTP	2 µl
E. coli DNA ligase	2 µl
E. coli DNA polymerase I	8 µl

2. Incubate at 16°C for 2 h.
3. Immediately before use, pellet a 2-ml MaXtract High Density Tube by centrifugation at 13,200 rpm for 1 min in a microcentrifuge.
4. Transfer DNA nick repair reaction mix into the MaXtract tube.
5. Purify DNAby Phenol:Chloroform:IAA as described in Subheading 3.8.
6. Precipitate DNA with isopropanol and resuspend in 84 µl of water.

3.15. Plasmid-Safe Treatment to Cleave Remaining Linear DNA Molecules

1. Set the following on ice in a 1.7-ml lo-bind tube:

Circularized DNA	84 µl
25 mM ATP (Epicentre)	4 µl
10× reaction buffer (Epicentre)	10 µl

2. Incubate at 37°C, 40 min (see Note 25).
3. Prepare the 2 ml MaXtract High Density Tube as described above.
4. Transfer the DNA into the MaXtract tube; adjust reaction volume to 200 µl.
5. Purify DNA by Phenol:Chloroform:IAA as described in Subheading 3.8.
6. Precipitate DNA with isopropanol and resuspend in 50 µl of water.

3.16. EcoP15I Digestion to Release 5′ and 3′ Tags from Circularized Full-Length cDNA

1. Set the following on ice in a 1.7-ml lo-bind tube:

Circularized full-length cDNA	50 µl
10× Buffer 3	10 µl
100× BSA	1 µl
10 mM Sinefugin	1 µl
10× ATP (NEB) (see Note 26):	2 µl
EcoP15I (NEB)	10 U/µg DNA
Nuclease-free water	Adjust volume to 100 µl

2. Incubate at 37°C, 2 h.

3.17. Binding of EcoP15I Digested DNA Tags to Dynalbeads M-280 Streptavidin

1. Swirl bottle of Dynalbeads M-280 streptavidin suspension thoroughly.
2. Transfer 50 μl of Dynalbeads M-280 Streptavidin suspension to a 1.7-ml tube.
3. Using the MPC, wash beads with 150 μl of 2× B&W buffer by pipetting up and down.
4. Resuspend beads in 100 μl of 2× B&W buffer.
5. Add 100 μl of EcoP15I-digested DNA to the resuspended Dynalbeads, mix well.
6. Incubate at room temperature with rotation on the Intelli-Mixer (Program F8, 30 rpm) for 30 min. During the incubation, biotinylated linkers associated with the captured DNA tags are to be bound and remained on the beads.
7. With the help of MPC, the reaction beads are washed twice with 150 μl of 1× B&W buffer by pipetting up and down, which will remove those DNA fragments with no linker attached.

3.18. Ligation of (Solexa or SOLiD) Sequencing Adaptors to DNA Template for HT Sequencing

1. For Solexa adaptor ligation, set the following on ice in a 1.7-ml tube:

Nuclease-free water	36 μl
Solexa 454-adaptor E/A (200 ng/μl)	4 μl
Solexa 454-adaptor E/B (200 ng/μl)	4 μl
10× T4 DNA ligase buffer	5 μl

2. For SOLiD adaptor ligation, set the following on ice in a 1.7-ml tube:

Nuclease-free water	36 μl
SOLiD P1 Adaptor (200 ng/μl)	4 μl
SOLiD P2 Adaptor (200 ng/μl)	4 μl
10× T4 DNA ligase buffer	5 μl

   Since the SOLiD sequencing adaptors (P1 and P2) are blunt-end, therefore, the EcoP15I released tags (5′-overhang) have to be blunt-end first by a "filled-in" reaction step before adaptor ligation (see Note 27 for detail).
3. Resuspend tag-bound beads with the above ligation mix based on specific sequencing platform chosen.
4. Add 1 μl of T4 DNA ligase to the beads suspension to make a final concentration at 0.6 U/μl to ligate adaptors to the captured PETs.
5. Incubate at RT for overnight with rotation on an Intelli-Mixer (Program F8, 30 rpm, U = 50, u = 60)
6. Wash the beads twice with 150 μl of 1× B&W buffer and go for next Nick Repair.

3.19. Nick Translation Repair

1. Set the following reagents on ice in a 1.7-ml tube:

Nuclease-free water	38.5 µl
10× Buffer 2	5 µl
10 mM dNTP	2.5 µl (final conc. 500 µM)
E. coli DNA polymerase I	4 µl

2. Resuspend Dynalbeads in the above reaction mix.
3. Incubate at RT with rotation for 2 h on an Intelli-Mixer (F8, 30 rpm).
4. Wash the beads twice with 150 µl of 1× B&W buffer using the MPC.
5. Resuspend the Dynalbeads in 50 µl of Buffer EB.

3.20. PCR Amplification to Assess Captured 5′ and 3′ cDNA Tags

1. Set the following on ice in 0.2-ml thin-walled PCR tube:

Nuclease-free water	21 µl
Dynalbeads suspension	2 µl
Solexa or SOLiD PCR primer-1	1 µl
Solexa or SOLiD PCR primer-2	1 µl
2× Phusion Master Mix with enzyme	25 µl

PCR cycling conditions:

98°C	30 s
98°C	10 s
65°C	30 s
72°C	30 s
72°C	5 min
4 C	hold

Repeat 20 times from step "98°C/10 s" to "step 72°C/30 s."

2. After PCR reaction is done, take 25 µl of PCR products and run on a 10-well 4–20% TBE gel at 200 V for 45 min. Stain the gel for 10 min in SYBR-TBE buffer before taking picture.
3. Load 500 ng of a 25-bp DNA ladder side by side for size determination.
4. As shown in the gel picture (Fig. 2) below, an expected 237 bp (Solexa) and 154 bp (SOLiD) bands are observed which are consistent with the structures of the captured cDNA tags associated with platform-specific linkers and adaptors.

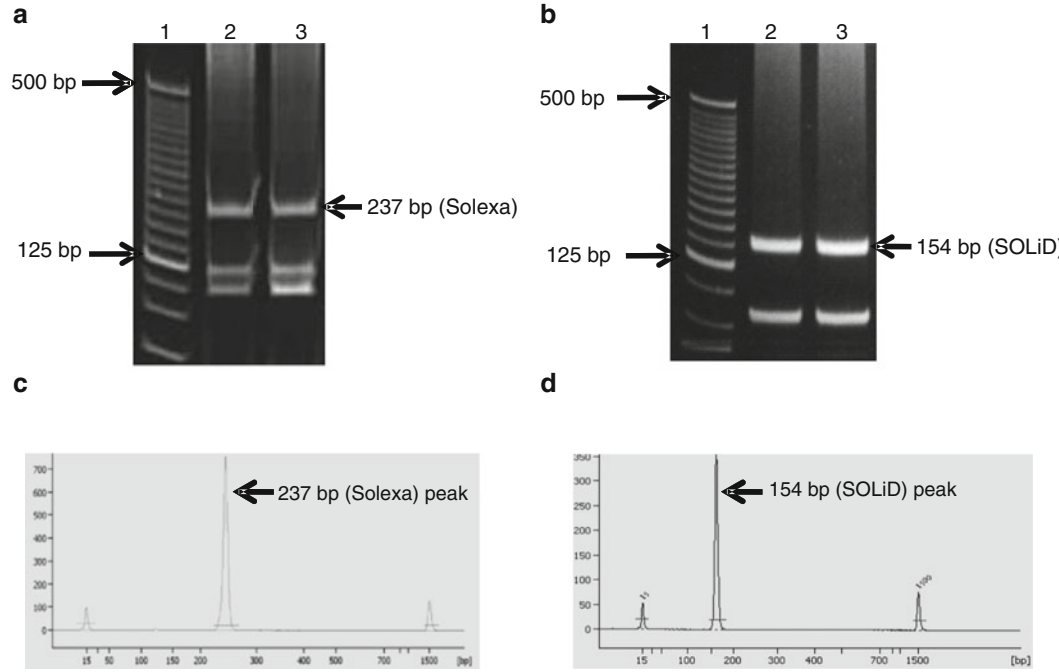

Fig. 2. RNA-PET sequence template QC assessments. Examples of RNA-PET libraries constructed for Illumina GAII (**a**) and ABI SOLiD4 (**b**) sequencing methods. The libraries were purified as specific PCR fragments (237 bp for Illumina GAII 154 bp for SOLiD), which are composed of 5′ and 3′ end tags, a linker sequence and two sequencer-specific adaptors ligated at each end. After gel purification, the library materials were further analyzed by Agilent Bioanalyzer profiles, showing unique DNA fragment peak at 237 bp for Illumina GAII (**c**) and 154 bp for SOLiD (**d**), respectively.

3.21. PCR Scale-Up for Preparation of Sequencing Templates

1. Set PCR scale-up reactions using all or at least half amount of available Dynalbeads as amplification template and collect expected PCR fragments separated on 6% TBE gel.
2. Depending on the number of PCR reactions set up, the PCR products need to be concentrated before loading into the gel.
3. A 25-bp DNA ladder is critical to be loaded side by side in the gel for size reference when harvesting the desired band.

3.22. Purify PCR Fragment from 6% TBE gel Using Gel-Crush Method

1. PCR fragment of the interest is carefully excised and collected into several 0.6-ml microtubes that have been pierced at the bottom with a 21-G needle. Two or more gel slices can be put into each 0.6-ml microtube according to the size of the gel slices. The pierced tube is placed inside a 1.5-ml screw-cap microtube and centrifuged at 13,200 rpm for 5 min in a microcentrifuge. The gel slices are thus conveniently shredded and collected in the bottom of each 1.5-ml tube.
2. Add 400 μl of TE buffer to each 1.5-ml screw-cap tube, stir the gel pieces with the pipette tip to ensure gel pieces are immersed in the TE buffer.

3. Transfer the 1.5-ml screw-cap tubes containing shredded to −80°C freezer for 1–2 h; then bring tubes back to a 37°C incubator for an overnight incubation. The DNA from the shredded gel will come out from the gel into TE buffer during the incubation.
4. After overnight incubation, transfer gel pieces together with the buffer to the filter cup of a Spin-X column and centrifuge at 13,200 rpm, 4°C for 10 min in a microcentrifuge.
5. After centrifugation, add 200 μl of TE to each filter cup and stir to loosen the gel pie with a pipette tip. Centrifuge again at 13,200 rpm for 10 min to recover remaining DNA.
6. Transfer the filter-through liquid to a new tube. Precipitate the DNA with isopropanol and resuspend in 20 μl of TE buffer.
7. Take 1 μl of DNA template to perform QC check with Agilent 2100 Bioanalyzer using DNA-1000 kit according to manufacturer's instruction. The Agilent profile should show a clean and strong DNA fragment peak at expected size and no noise background detected in the sample.
8. Sequence the RNA-PET library templates by the paired end (PE) sequencing method according to manufacturer's (Illumina GA or SOLiD v4) guidelines and instructions.

3.23. PET Sequencing and Mapping

3.23.1. Paired End Tag Sequencing with Illumina (Solexa) GA and SOLiD v4

1. Paired end (PE) sequencing run-type is performed with the respective library made using either Illumina (Solexa) GA or SOLiD v4 following manufacturer's guidelines and instructions. For Illumina GA, the PE sequencing generates two paired end reads with each at 36 bp length. SOLiD v4 PE sequencing generates two paired end reads at each 35 bp long.
2. In Solexa sequences, each 36 bp read consists of a 27-bp tag plus 9-bp linker sequence, which is part of the linker immediately close to the tag. For SOLiD PE sequence, each 35 bp read contains a 27-bp tag plus 8-bp linker sequence that is close to the tag. The PET structure is simply illustrated as 27 bp tag + 9/or 8 bp – linker – 9/or 8 bp + 27 bp tag.
3. In our experience, one lane of a Solexa flow cell (FC), after filter out the noise reads, can usually produce over 20–30 millions of pass filter PETs, or PETs. For SOLiD run, a spot equivalent to 1/8 of a slide can also produce approximately 30 millions of usable paired end reads.
4. Since the sequencing adaptors can be ligated to either end of a given transcript PET, to differentiate which end tag is come from 3′ or from 5′ end of a transcript, a specific signature sequence (*AA*CTGCTG), characteristic by a double *AA* residue in the beginning of the signature, and is also the two end residues at the 3′-end of a 27-bp tag, services as an indicator for identifying the 3′-end tag first.

5. To begin the PET analysis, the signature sequence is searched out first from both paired end sequence reads. As long as the *AA*CTGCTG (5→3) sequence is identified from one end of a PET, this tag is considered as the 3′-end tag and the other end is defined as the 5′-end tag. There is a small portion of the PETs which do not have any signature sequence being identified, or in rarely cases, the signature sequence could appear on both paired ends; these PETs are discarded from future analysis.

3.23.2. Mapping of the Paired End Tags

1. After PET orientation (5′>3′) is identified, regular mapping of 5′ and 3′ tags to reference genome is performed through Solexa (ELANE) or SOLiD (BioScope) analysis pipeline, followed by further analysis approaches specifically designed for RNA-PET.
2. In Solexa and SOLiD mappings a seed of 25 bp of tag is used and a maximum 2 bp mismatches are allowed within the seed sequence for each tag mapping.
3. For those PETs which have uniquely mapped both to the 5′ and 3′ ends of the reference genome are classified as uniquely mapped PETs.
4. Approximately 90% of PETs which are mapped on the same chromosome, in the same strand, and in the same correct orientation to the known transcripts or known transcript variants, are defined as *concordant* PETs.
5. Whereas, a small portion ~10% of PETs, which are mapped incorrectly to the reference genome, are referred as *discordant* PETs, that makes a class of the PETs which are mapped either in the wrong orientations in the same strand (e.g., 3′-end tag mapped before the 5′-end tag), or the paired tags mapped on two different strands, or two different chromosomes, e.g., one end mapped on chromosome 3, another end mapped on chromosome 8.
6. A majority of *discordant* PETs are derived from ligation noise; however, this class of the PETs serves as a valuable pool to identify those novel, fusion or transcriptional variants which might be caused by a variety of genome rearrangements such as deletions, inversions, insertions, tandem repeats, and translocations; or from transcriptional variations of *trans*-splicing or mutational events etc.

3.23.3. PET Clustering and Genome Annotation

1. After *concordant* PETs are identified, they are clustered each other with nearby PETs based on searching a 100-bp distance as a window to extend each PET at 5′- and 3′-end, respectively. Specifically, the mapping location of the 5′ and 3′ tag of a given PET is extended in a 100-bp wide window from both directions.
2. If the 5′ and 3′ tags of a second PET mapped within the 5′ and 3′ search window of the first PET then the two PETs are clustered

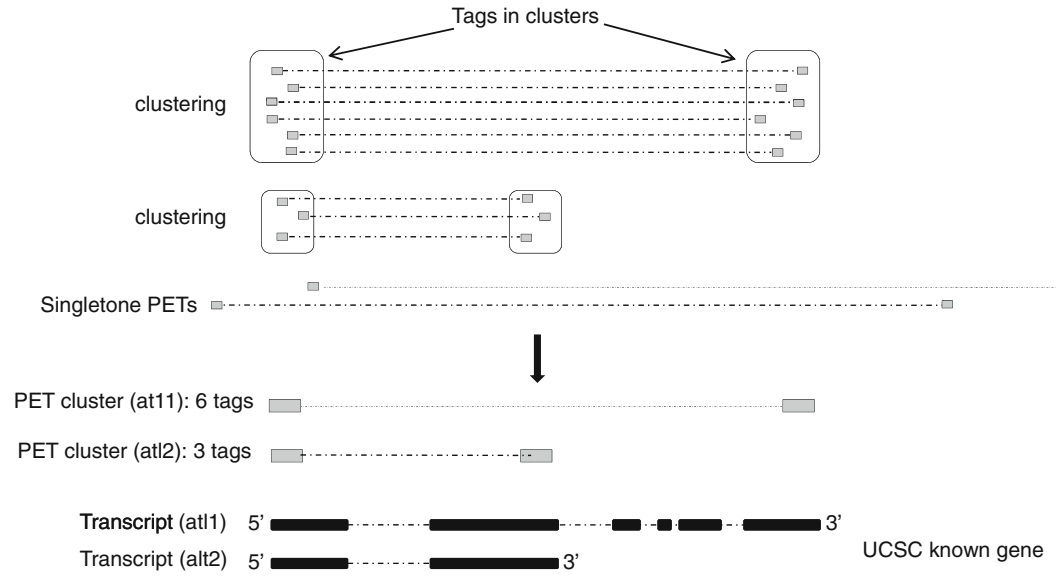

Fig. 3. Scheme of RNA-PET sequence clustering and annotation to reference genome. After mapping to reference genome, tag sequences aligned in defined span range are cluster together to represent molecules derived from the same transcripts. The illustrated example showing 6 PET sequences clustered together to represent one transcript of the gene encoded at this location, and three other PET sequences cluster together tp represent a short alternative of transcript of this gene. Singletons are discarded from further analysis.

and the search windows are re-adjusted and continue the expansion to acquire new PETs. This process is dynamic and iterative, and continues until no new PETs can be found within the allowed window distance.

3. At the end of clustering process, most related *concordant* PETs are clustered each other and those PETs which do not fall into this search window are classified as singletons (Fig. 3), and filtered out from the dataset and not go for further analysis. Furthermore, any PETs to be clustered, 5′-end of tags should be within 100-bp each other, and the same criteria applies to the 3′-end tags.
4. The following (Fig. 3) diagram illustrates the mapping and annotation concepts.
5. Figure 4 shows the sequence tags mapped on the RNA-PET browser from two libraries derived from breast cancer cell line samples MCF7 with Control vs. Estrogen diol treatment.

Two RNA-PET library datasets are shown in Fig. 4. PET sequence tags mapped on the browser are from two breast cancer cell samples i.e., MCF7-Control vs. Estrogen diol treated. As the concordant PET data shown on the browser, expression levels of two libraries

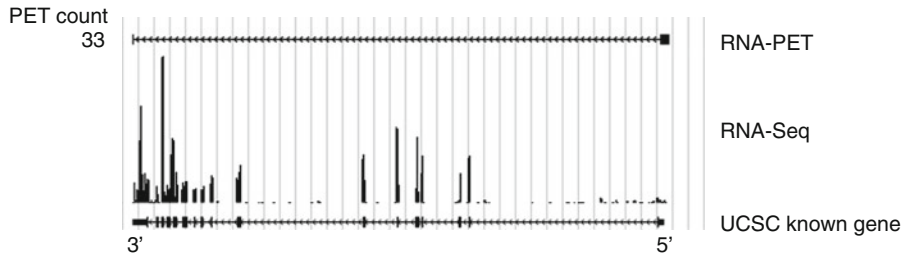

Fig. 4. An Example of RNA-PET data visualized in genome browser. RNA-PET sequence tags mapped in genome browser are from breast cancer cell MCF7. The PET count reflects the expression level and the tag mapping locations demarcate the boundaries of the transcription unit. RNA-Seq data from the same cells showed the marks of each exons. The combination of RNA-PET and RNA-Seq data provides a comprehensive view of transcript structure and expression abundance.

are clearly represented by the PET counts. In addition, novel splice transcript can be seen from the mapped PETs.

4. Notes

1. For all steps involving single-stranded DNA or RNA, it is preferable to use "lo-bind" microfuge tubes to avoid loss of nucleic acid.
2. Commercially available nuclease-free water (instead of DEPC-treated water) was used for all RNA-containing enzymatic reactions to avoid possible inhibition of enzymatic reactions by residual DEPC or ethanol.
3. Because the solubility of trehalose increases with temperature, it is important to maintain the liquid temperature at or around 42°C (in which temperature trehalose is used in the protocol) when making trehalose solution.
4. For all steps involving RNA manipulations, ensure RNase-free conditions are maintained, including all reaction buffers.
5. This step is necessary when combined volume of polyA mRNA and GsuI-dT oligo exceeds 9 μl.
6. Do not use glycogen at any stage where it is not specifically mentioned as glycogen will interfere with the cap-trapper selection process.
7. Use 2.5 U of RNase ONE (Promega) per μg of starting polyA mRNA.
8. During RNase One digestion, proceed to Subheading 3.6 to prepare the Dynalbeads.
9. Before using the Dynalbeads, resuspend the beads in suspension by strong shaking.
10. Cautions in Dynalbeads washing procedures:

(a) After transferring the required amount beads (e.g., 200 μl) to a fresh lo-bind tube, the tube should be placed on a MPC stand (magnet) for at least 1–2 min.

(b) Remove wash supernatant only when the tube is positioned on a MPC. Never remove S/N while the tube is taken off from the MPC.

(c) Add washing buffer along the inside wall of a tube.

(d) Resuspend and mix the beads only when the tube is taken off the MPC.

11. As Dynabeads are heavy and can easily settle down to the bottom of the tube, the immobilization of (−) cDNA/RNA heteroduplex onto the beads should be done on an Intellimixer with constant rotations.
12. The incubation time of immobilization can be increased if the samples are diluted.
13. Always use freshly prepared alkaline hydrolysis buffer.
14. RNase-free condition is not necessary from this step onward.
15. DNA linkers should be annealed beforehand and stored in aliquots at −20°C. Always keep the annealed linkers on ice and avoid warming them when thawing. This precaution could prevent possible denaturation and subsequent complications. Centrifugation of the annealed adaptors is also recommended in cold (4°C) condition.
16. Ensure that the cDNA and linkers are well mixed before adding solution II as the latter contains PEG, which could lead to precipitation of glycoblue.
17. It is important to use only columns that do not show any visible bubbles trapped within the matrix as these tubes are likely faulty in our experiences. Also, the chances of bubble formation can be reduced by allowing the columns to equilibrate to room temperature.
18. It is preferred that each draining of the cDNA columns does not take longer than 25 min.
19. Avoid collecting any fractionation tubes showing presence of small molecular weight linker DNA bands as the presence of linkers in following reactions will cause problems in quenching enzymatic reactions.
20. Use about 10 U of EcoP15I (NEB) per μg of starting double-stranded cDNA.
21. It is assumed that the average length of cDNA is 2.5 kbp.
22. Add T4 DNA ligase to the reaction at the last and keep the reaction in cold at all times.
23. Circularization can be performed for as long as >24 h at 16°C.

24. Precipitation can also be done in multiple 1.7-ml tubes.
25. The maximum incubation time for plasmid-safe treatment can be 2 h.
26. The ATP that is supplied with EcoP15I (NEB, B6101S) is at 10×, thus 20 μl is needed to obtain a 2× final concentration while the ATP sold separately (NEB, P0756) is usually at 100×, and hence 2 μl is used instead in the protocol.
27. The "filled-in" blunt-end reaction is conducted using Endit kit (Epicentre Inc.) following manufacturer's instruction. After "filled-in" reaction, the beads are washed following the steps described in Subheading 3.17 (7) and are ready for next blunt-end ligation for SOLiD adaptors.

Acknowledgments

The authors would like to thank GOH, Yufen; Ho, Andrea; QUEK, Kelly, KHNG, Alexis, CHOY, Yeen Hui, Atif, Shahab; ONG, WaiLoon; POH, WanTing; Lavanya Veeravalli; Thoreau, Herve; Ong, ChinThing; Dawn, Sum; Leong, SeeTing and all other members of the GTB research and sequence group. This work was supported by NIH ENCODE grant and A-STAR, Singapore.

References

1. Peters, B. A and Velculescu, V. E. (2005). Transcriptome PETs: A genome's best friends. *Nature Methods* 2, 93–94.
2. Fullwood, M. J., CL. Wei, E.T., Liu, and Y. Ruan (2009) Next-Generation DNA Sequencing of Paired End diTags for Transcriptome and Genome Analysis (Review). *Genome Research* 19, 521–532.
3. Wei, C.L., Ng Patrick W.P., Chiu, K.P., Wong, C.K., Ang, C.C., Lipovich L., Liu, E.T., and Y. Ruan (2004). 5′ Long serial analysis of gene expression (LongSAGE) and 3′ LongSAGE for transcriptome characterization and genome annotation. *PNAS*, USA. 101, 11701–11706.
4. Ng P, Wei CL, Sung WK, Chiu KP, Lipovich L, Ang CC, Gupta S, Shahab A, Ridwan A, Wong CH, et al. (2005) Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nature Methods.* 2, 105–111.
5. Ng P, Tan JJ, Ooi HS, Lee YL, Chiu KP, Fullwood MJ, Srinivasan KG, Perbost C, Du L, Sung WK, et al. (2006) Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes. *Nucleic Acids Res.* 34, e84.
6. Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., Muramatsu, M., Hayashizaki, Y., Schneider, C. (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics.* 37, 327–336.
7. Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., Schneider, C., Hayashizaki, Y. (1997) High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Res.* 4, 61–66.
8. Carninci P, Hayashizaki Y. (1999) High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19–44.