Published in final edited form as:

Methods. 2019 February 15; 155: 20–29. doi:10.1016/j.ymeth.2019.01.002.

Development of Poly(A)-ClickSeq as a Tool Enabling Simultaneous Genome-wide Poly(A)-site identification and Differential Expression Analysis

Nathan R. Elrod^{*,1}, Elizabeth A. Jaworski^{*,1}, Ping Ji¹, Eric J. Wagner^{1,2}, Andrew Routh^{1,2}

¹Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA

²Sealy Centre for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas, USA.

Abstract

The use of RNA-seq as a generalized tool to measure the differential expression of genes has essentially replaced the use of the microarray. Despite the acknowledged technical advantages to this approach, RNA-seq library preparation remains mostly conducted by core facilities rather than in the laboratory due to the infrastructure, expertise and time required per sample. We have recently described two ‘click-chemistry’ based library construction methods termed ClickSeq and poly(A)-click-seq (PAC-seq) as alternatives to conventional RNA-seq that are both cost effective and rely on straightforward reagents readily available to most labs. ClickSeq is random-primed and can sequence any (unfragmented) RNA template, while PAC-seq is targeted to poly(A) tails of mRNAs. Here, we further develop PAC-seq as a platform that allows for simultaneous mapping of poly(A) sites and the measurement of differential expression of genes. We provide a detailed protocol, descriptions of appropriate computational pipelines, and a proof-of-principle dataset to illustrate the technique. PAC-seq offers a unique advantage over other 3’ end mapping protocols in that it does not require additional purification, selection, or fragmentation steps allowing sample preparation directly from crude total cellular RNA. We have shown that PAC-seq is able to accurately and sensitively count transcripts for differential gene expression analysis, as well as identify alternative poly(A) sites and determine the precise nucleotides of the poly(A) tail boundaries.

Keywords

Polyadenylation; ClickSeq; Differential Gene Expression; Poly(A) sites; Click-Chemistry

To whom correspondence should be sent: ejwagner@utmb.edu, alrouth@utmb.edu.

*equally contributed

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest Statement

E.J and A.R. are co-founders and owners of ‘ClickSeq Technologies LLC’, which is a Texas-based Next-Generation Sequencing solution provider offering ClickSeq protocols including PAC-seq.

1. Introduction

RNA-seq is regularly used as a tool to detect subtle changes in gene expression in samples varying by two or more conditions. The primary advantage of RNA-seq over other approaches such as microarray, is that RNA-seq will detect any template, regardless of sequence identity or prior knowledge. This is achieved through the use of random fragmentation and ligation approaches or by random-primers that can generate cDNA libraries from any template. These approaches, whilst ‘agnostic’, will thus sequence the entire RNA sample and return sequencing data across the entirety of an mRNA. This yields valuable information about the mRNA transcript, such as the identity of differential RNA splicing events, intron inclusion, and single nucleotide variants. Differential Expression (DE) is typically performed simply by extracting counts per mRNA from normalized RNA-seq data coverage over entire mRNA transcripts [1].

However, read coverage across the entire mRNA transcript is not necessarily required to accurately measure RNA abundance. Instead, targeting sequencing to just the 3’ ends of eukaryotic RNAs can provide a highly sensitive and cost-effective method for differential gene expression [2]. Only one read is obtained per mRNA molecule and so transcript abundance can be assigned directly from the number of reads mapping to a single transcript. Without having to map reads across the entire RNA, the number of sequencing reads required to achieve equivalent information is this dramatically reduced, allowing multiple replicates of multiple samples to be analyzed on a single flowcell and thus saving considerably on cost. Furthermore, common random RNA-seq pipelines require enrichment of mRNAs and/or depletion of ribosomal RNAs which would otherwise dominate the output sequencing data. Therefore, pipelines that focus mRNA quantification just to the 3’ ends of mRNAs have recently become popular as a means of quickly yet accurately performing differential gene expression studies [3–6].

Poly(A) tails are present at the 3’ termini of almost all eukaryotic mRNAs, with only a handful of exceptions [7]. A number of approaches have been described for the 3’ seq of poly(A) tailed mRNAs (recently reviewed by Szkop and Nobeli [8]). While many of these techniques are primarily designed for the specific identification of poly(A) sites in mRNAs, methods like *QuantSeq* [5] use a simple and high-throughput poly(A)-directed approach to allow the rapid assessment of mRNA transcript abundance.

We recently developed a method called “*ClickSeq*” [9] that utilizes copper-catalyzed azide-alkyne cycloaddition (‘click-ligation’) in place of enzymatic ligation as a simple and efficient chemistry based approach for next-generation sequencing library synthesis. Briefly, 3’-Azido-2’,3’-dideoxynucleotides (AzNTPs) are spiked-in to reverse transcription reactions generating random azido-blocked cDNA fragments that can be ‘click-ligated’ to alkyne-functionalized sequencing primers [10]. ClickSeq has advantages over other RNA-seq approaches including the removal of template fragmentation and reduced template switching, in addition to practical advantages including its simplicity, cost-effectiveness and fast turnaround [11]. ClickSeq is a flexible platform on which to build a number of applications.

We recently reported Poly(A)-ClickSeq (PAC-seq) [2], that specifically sequences the 3' ends of poly(A)-tailed mRNAs allowing quantification of the positions and abundance of poly(A) sites (PASs) at the ends of eukaryotic mRNAs. For PAC-seq, we initiate reverse transcription using poly(T) primers, without a non-T anchor. Importantly, to specifically generate sequence reads spanning the junctions of mRNA 3'UTRs and poly(A) tails, AzTTP is omitted from the RT-PCR reaction (AzVTPs). As a result, reverse-transcription must continue through to beginning of the poly(A)-tail and into the 3'UTR before chain termination can occur, thus 'homing-in' on the 3'UTR/poly(A) junction. Thereafter, we can purify the azido-terminated cDNA, click-ligate the 5' Illumina adaptor and generate an NGS library enriched with 3'UTR/poly(A) junctions. A schematic is illustrated in Figure 1. By virtue of priming from the poly(A) tails of mRNAs and terminating shortly upstream, no fragmentation, enrichment of poly(A)-tailed RNAs, nor depletion of ribosomal RNA is required. We previously demonstrated that we could robustly generate PAC-seq libraries from as little as 125ng of crude total cellular RNA [2].

The primary goal of PAC-seq was to identify and characterize the frequency and positions of poly(A) sites (PASs) within the transcriptome. In this manuscript, we describe how we can also use PAC-seq to perform differential gene expression (DE) analysis. To illustrate the utility of PAC-seq, we depleted a component of the *Drosophila* Integrator complex in DL1 cells using RNAi and compared the gene expression changes relative to control dsRNA-treated cells. RNA isolated from these cells was then subject to standard RNA-seq or PAC-seq and DE analyses were performed using each library generation method. In addition to providing information on poly(A) tail position, we demonstrate that PAC-seq also reveals global changes in the mRNA transcript abundance and that these results closely match the changes observed by canonical RNA-seq methods. PAC-seq therefore provides a robust and cost-effective method for DE analysis that can simultaneously reveal the positions and frequencies of poly(A) sites.

2 Materials and Methods

2.1 Cell culture and RNA samples

For the RNAi knockdown of target genes, dsRNA of 500bp in length was prepared using the New England Biological T7 HiScribe kit (NEB) and purified using RNazol (Sigma). Four wells in a 6-well plate were spotted with 15µg of dsRNA for each dsRNA target and 1×10^6 DL1 cells [12] from *Drosophila melanogaster* were added in 1 mL of serum free media (Gibco) and incubated for 1 hour at 27°C before the addition of 2mL media containing 10% FBS. These cells were then incubated for 60 hours at 27°C before harvesting. Total RNA was extracted from cells in three replicates of each series using the standard TRIzol extraction protocol (Invitrogen) and resuspended in water to a concentration of 500 µg/µL. The final well in each replicate was harvested using RIPA buffer for protein Western analysis of target protein knockdown verification.

2.2 Library Preparation Protocol

Our protocol for randomly-primed ClickSeq (CS) for RNA-seq libraries has been described in detail in Jaworski et al. 2018 [11] and was used to generate the ClickSeq libraries

described in this manuscript. Here, we describe the steps required for PAC-seq that contain a number of specific modifications compared to the ClickSeq protocol, most importantly during the reverse transcription step.

2.2.1 Reverse Transcription

For PAC-seq, we followed a modified version of the SuperScript-III (Invitrogen) first-strand reverse transcription protocol by supplementing 3'-Azido-2',3'-dideoxynucleotides (AzVTPs) (Trilink Biotechnologies). First, a 1:5 mix of 5mM AzVTP:dNTPs was prepared by mixing 2μL 10mM AzATP, 2μL 10mM AzCTP, 2μL 10mM AzGTP, 10μL 10mM dNTPs each in water, and 4μL H₂O. The reverse transcription reaction was setup by combining:

- 2μL 5mM 1:5 AzVTP:dNTPs
- 1μL Illumina_4N_21T primer (100μM in water)
(GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNTTTTTTTTTTTT
TTTTTTTTTT),
- 0.5–2μg of RNA total cell
- H₂O to a final volume of 13μL

Input RNA can be extracted using any standard RNA extraction procedure (e.g. TRIzol) but no further purification, rRNA depletion, poly(A) selection, or fragmentation steps are required. This mixture was incubated at 65°C for 5 mins to melt RNA and immediately cooled on ice for >1 min to anneal the poly(T) primer. Subsequently, following the standard RT-PCR protocol, we combined the following for a final reaction volume of 20μL:

- 4μL 5X Superscript First Strand Buffer
- 1μL 0.1M DTT
- 1μL RNase OUT Recombinant Ribonuclease Inhibitor(40 U/μL) (Invitrogen)
- 1μL Superscript III Reverse Transcriptase(200 U/μL)

The RT reaction was incubated at 50°C for 40 mins, 75°C for 15 mins, and then held at 4°C. To remove template RNA, 2U of RNase H (NEB) was added and incubated at 37°C for 20 mins, 80°C for 10 mins, and held at 4°C.

2.2.2 Azido-terminated cDNA purification—After single strand synthesis and RNA digestion, the azido-terminated cDNA fragments must be purified, removing any excess AzVTPs which may compete with the subsequent click reactions. This can be achieved in many ways including any silica column based extraction (i.e. Zymo Research DNA Clean and Concentrator-5, D4013), phenol/chloroform, or SPRI (Solid Phase Reversible Immobilization) Magnetic Beads such as AMPure beads (Beckman Coulter) or homemade [13]. We prefer to use SPRI beads due to their simplicity of use and high throughput ability. Here, following the standard procedure, we mixed 1.8X (38μL) SPRI beads into the RT-PCR reaction and incubated for 5 min at room temperature. Beads were pelleted using a magnetic rack, discarding the supernatant. We washed beads twice with 200μL 80% ethanol taking care not to disturb the bead pellet and air-drying until no excess ethanol was visible. Bead

were finally re-suspended in 10 μ L 50mM HEPES pH 7.2, re-pelleted on a magnetic rack and the supernatant containing the eluted cDNA fragments were transferred to a new tube.

2.2.3 Click-ligation to attach sequencing adapter—After purification, a click-ligation reaction is used to chemically attach a 5' alkyne-modified click-adapter onto the 3' end of the azido terminated cDNA fragments. To proceed with the reaction, we diluted the 10 μ L of cDNA (from the purification step 2.2.2) with 20 μ L 100% DMSO and added 3 μ L of 5 μ M Click-Adapter. (5'Hexynyl-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCCCGTATCATT), 5 μ M in water (IDT, with HPLC purification).

Next, we separately prepared the catalyst/accelerant mixture by combining 0.4 μ L Vitamin C (Sigma) at 50mM with 2 μ L Cu(II)-TBTA (10 mM in 55% aq. DMSO, Lumiprobe). This initiates the reduction of the Cu(II) to Cu(I). We immediately added 2.4 μ L of the copper mixture to the 23 μ L of cDNA, initiating the click-reaction and incubated for 30min at room temperature. We repeated the addition of the copper catalyst mixture for a total of two times. This reaction yields a cDNA fragment that is linked to the p5 sequencing adapter through a triazole-ring in place of the phosphate backbone. The cDNA fragments were purified away from the excess components of the click-ligation reaction following the SPRI magnetic bead protocol: 1.8X (68 μ L) SPRI beads were added to 37.8 μ L of the click-reaction as washed as in step 2.2.2, eluting in 20 μ L 10mM Tris pH 7.4 or water.

2.2.4 Final PCR Amplification: The cDNA fragments were then PCR amplified to add on the remainder of the p7 Illumina sequencing adapter and to generate sufficient enough material for sequencing. The PCR reaction was set up by combining the following reagents for a 50 μ L reaction:

- 10 μ L Clean Click-ligated DNA
- 2 μ L 5 μ M Universal Primer: (AATGATACGGCGACCACCGAG),
- 2 μ L 5 μ M Indexing Primer (1 barcode/sample):
(CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGT) underlined portion corresponds to the index sequence, any index sequence may be used here
- 25 μ L 2X One Taq Standard Buffer Master Mix (NEB)
- 11 μ L H₂O

PCR cycling was as follows:

- 1m at 94°C; 30s at 54°C; 10m at 68°C;
- [30s at 94°C; 30s at 54°C; 2m at 68°C] $\times$ 15–20 cycles;
- 5m at 68°C; hold at 4°C.

Finally, the PCR products were cleaned and isolated, again following the SPRI magnetic bead protocol: 1X (50 μ L) SPRI beads were added to 50 μ L of the PCR reaction, following the same washing procedure as in step 2.2.2, eluting in 20 μ L 10mM Tris pH 7.4 or water.

2.2.5 Gel extraction and size selection—The final sequencing library needs to be size selected. We have found that the most accurate way is by running the amplified cDNA library on an electrophoresis gel and cutting the appropriate band size based off a molecular weight ladder. Size selection is critical for PAC-seq since fragments that are too short will not yield map-able cDNA fragments and long fragments will not cluster properly on the sequencing platform or produce sequence reads that are distal from the poly(A) tail. Therefore, gel size excision should be 200–300bp for a 1×150bp Illumina run or 200–400 for a 1×300bp Illumina run. Libraries can be extracted from the agarose following any standard protocol (i.e. Zymoclean Gel DNA Recovery Kit, D4001). Final sequencing libraries are quantified and pooled accordingly.

2.3 Data processing

2.3.1 RNA-seq—The first step in our ClickSeq processing pipeline is canonical, with an Illumina adaptor trimming step and quality filtering. In ClickSeq, the first four nucleotides read by the Illumina platform are the four random nucleotides included in the click-adaptor, which are designed to aid cluster generation and ensure diversity in these first nucleotides. We have noticed that the fifth and sixth nucleotides tend to be read as either an “A” or “G”. We believe that this may be due to the Taq polymerase inserting a thymine opposite the triazole linkage that is present at this site in the click-linked cDNA as it may resemble an abasic site. Therefore, we additionally trim the first 6 nucleotides from the beginning of each read. We perform all of these raw-read processing steps with *fastp* [14] that can perform all these required actions and trim oligo-G tracts, as are commonly seen when using two-color sequencing systems such as the Illumina NextSeq using the following command-line entry:

```
fastp -i <input fastq> -a AGATCGGAAGAGC -f 6 -g -l 40 -Q -o <output fastq>
```

2.3.2 PAC-seq—PAC-seq reads require careful preprocessing due to the frequent presence of long homo-A tracts derived from the poly(A) tails of the targeted mRNAs. In our original description of PAC-seq [2], our pipeline for preprocessing was similar to that for randomly-primed ClickSeq reads, except in addition to Illumina adaptor trimming we performed an additional adaptor trimming step to remove the poly-A tracts. In that pipeline [2], we used a custom script (available at <https://sourceforge.net/projects/dPAC-seq/>) that compared the final trimmed reads to the raw data in order to measure how many ‘A’s had been removed from the read and added this information into the read name. In effect, this measures the length of the poly(A) tail that was successfully sequenced by the Illumina platform so that this information can later be used as a quality filtering metric when assigning the position of poly(A) sites. This step is important as the reverse-transcription oligo-dT primer may often mis-prime from non-cognate templates which would generate artifactual poly(A) sites. To control for this, PASs are filtered by only accepting those with reads mapping that contain a greater number of ‘A’s than the number of ‘T’s in the oligo-dT primer (>21) as these must be derived from the native and authentic poly(A) tail, rather than the RT-primer.

These filtering steps, while providing an accurate report of the position and frequency of PASs in mRNAs, results in the removal of a large number of mapped reads. In our original

manuscript[2], we reported that approximately half of the raw reads were finally used to assign PASs after all filtering steps. For differential gene expression analysis, our goal is simply to determine how many reads map to mRNAs rather than find PASs. Therefore, these stringent filtering steps may remove a swath of data, and may also bias the output data to templates with longer poly(A) tails or to templates with a low 'A' content in the 3'UTR, as these would be more likely to terminate adjacent to the poly(A) tail. This has the further advantage in that shorter read lengths are suitable when selecting sequencing reagent kits. For PASs characterization, we recommend at least 150nt reads to obtain reads that are of sufficient length to capture a the 3'UTR as well as a sufficiently long portion of the poly(A) tail to allow quality filtering as described above. However, for DE-Seq, only short reads are required for an unambiguous mapping and so it is suitable to perform only 1×75 or shorter single-end sequencing. This therefore saves considerably on sequencing cost and time.

Therefore, we present a much simpler read preprocessing steps for DE analysis. First, we perform the same processing steps as before (*fastp* filtering followed by poly(A) tail trimming using the FASTX toolkit) and then trim all the reads to a fixed length of 50–60 nucleotides without requiring reads to contain a poly(A) tail using the following command-line entries:

1. `fastp -i <input fastq> -a AGATCGGAAGAGC -f 6 -g -l 40 -Q -o <output fastq>`
2. `fastp -i <input fastq> --disable_quality_filtering --trim_poly_x -a
AAAAAAAAAAAAAAAAAAAA -l 50`
3. `fastx_trimmer -Q33 -l 60 -i <input fastq> -o <output fastq>`

2.4 Read alignment

For both random-primed ClickSeq and PAC-seq, short reads can be mapped to reference genomes using a number of available alignment software packages. When targeting eukaryotic mRNAs, we recommend using splice-aware short-reads aligners such as STAR[15] or HiSat2[16]. The STAR aligner will account for annotated splice-sites that may be present in terminal exons and may be preferred when mapping very short reads (<75nts in length). Similarly, HiSat2 will find *de novo* splice sites and so may be preferred when mapping longer reads (>75nts) that can reliably map over splice events. Furthermore, these aligners will tolerate short soft-pads in the 3' ends of the mapped reads in case they over-run the end of reference transcript sequence and contain non-reference fragments of the poly(A) tail or the Illumina sequencing adaptors.

2.5 Transcript count assignment

Random-primed ClickSeq generates even coverage over entire mRNA transcripts. Therefore, to extract transcript abundance, read distribution must be normalized over the length of transcript, including factors such as GC content and bias in order to obtain a final count-per-transcript. For PAC-seq, transcript counts are obtained simply by counting the number of reads mapped over a reference transcript. Annotated transcripts are obtained from the UCSC refseq [17] databases and output as BED files for the full-length transcript. We next use the *featureCounts* command from the *Subread* suite [18] to return the number of reads mapping over individual transcripts. This returns a matrix of reads counts per annotated transcript for

each RNA sample (example in Table 1), which can be used as an input for downstream DE-Seq pipelines.

2.6 DEseq2 pipeline

DEseq2 analysis was performed following the vignette on the *Bioconductor* website [19]. In short, the data from the *featureCounts* output was read into a matrix in *R* and used to create a metadata file containing treatment groups. Sample counts for each gene were then normalized and Wald tests were performed to determine differential expression. Results were tabulated and output to determine changes in gene expression for each sample preparation method.

2.7 Data availability

All batch scripts and associated scripts have been uploaded to *sourceforge* and are freely available and regularly maintained at: (<https://sourceforge.net/projects/DPAC-seq/>). Raw data associated with this manuscript have been deposited into the NCBI SRA archive with the project number PRJNA498335.

3 Results

3.1 Samples and NGS libraries

To compare the effect of knockdown of Integrator subunit 11 upon global transcript abundance, we extracted total cellular RNA from DL1 cells that had Integrator 11 (INTS11) knocked-down using dsRNA and compared these to a control DL1 cells with a control knock-down to LacZ. This control was chosen to avoid observing changes in expression profiles that may be due to the delivery of dsRNA (e.g. innate immunity induction). To establish whether PAC-seq produces accurate and reproducible results in a differential gene expression analysis, we prepared sequencing libraries using PAC-seq directly from crude cellular RNA extracted from the cultured DL1 cells [2]. In parallel using the same samples, we performed a poly(A) enrichment (NEBNext) to remove non-poly(A) tailed RNAs such as ribosomal RNAs and generated randomly-primed RNA ClickSeq libraries. This is a well-established method of generating RNA-seq libraries that provides coverage across the length of the mRNA transcripts [20, 21] and will allow us to directly compare 3' end poly(A)-tail targeted methods for making NGS libraries versus random untargeted methods. Twelve libraries (2 methods $\times$ 3 replicates $\times$ 2 conditions) were submitted for sequencing on a NextSeq 550 for SE reads (1 $\times$ 75 for ClickSeq, 1 $\times$ 150 for PAC-Seq). We obtained in the range of 10–30 million raw reads per library. Each of these libraries were trimmed and quality filtered as described in the methods (schematic illustrated in Figure 2) to remove adaptor sequences, poly(A) tails, fragments of the click-adaptor, and any poor-quality base called reads.

3.2 Alignment of NGS data and extraction of gene counts

Reads for each replicate (~10–22 million reads each) were mapped to the *Drosophila* reference genome (*Dm3*) using the STAR aligner[15], as described in *materials and methods*. The breakdown of the locations of read mapping is shown in Figure 3A for each condition (SFig 1 for each replicate individually). In the random-primed ClickSeq (CS)

libraries for both LacZ control and INTS11 knockdown samples, over 80% of all the mapped reads were located within or 200nts downstream of annotated mRNAs. As illustrated in Figure 3B with an example mRNA, RpS4, RNA-seq reads are distributed along the length of the mRNA transcripts, revealing exon splicing junctions, and occasionally mapping over known introns.

In Poly(A)-ClickSeq (PAC-seq), over 80% of all the mapped reads were similarly found within or 200nts downstream of annotated mRNA Figure 3A for each condition (SFig 1 for each replicate individually). However, the majority of these reads were found in the terminal exon, as would be expected when priming from the poly(A) tail of mRNAs. This is illustrated in Figure 3B showing a large peak of read coverage extended from the annotated PAS into the 3' UTR, but not further in the gene body. This illustrates how PAC-seq can effectively and specifically sequence the 3' ends of mRNAs from crude cellular RNA extracts.

3.3 Differential gene expression pipeline and results

To determine whether PAC-seq can reliably reveal changes in gene expression, we compared PAC-seq libraries coverage profiles between the control LacZ dsRNA treated cell and the INTS11 dsRNA cells. As expected, coverage of reads over the 3' UTR of INTS11 was substantially reduced (Figure 3B and Table 1). Inspection of other genes also revealed numerous instances of gene down regulation upon INTS11 KD (Figure 3B) as well as up regulation (Figure 3C and Table 1).

To systematically measure changes in gene expression and perform DE, we employed the DESeq2 pipeline for both our RNA-seq and PAC-seq data (Supplementary Datasets 1 and 2 respectively). For the RNA-seq data, replicates from each condition were normalized by calling the *estimateSizeFactors* command in the DESeq2 pipeline which uses a negative binomial distribution, linking variance and mean by local regression [22]. This resulted in normalization values that were in line with the differences in total number of read counts sequenced for each sample.

For RNA-seq, assignment of transcript abundance/count was performed using the *featureCounts* command. For PAC-seq, transcript abundance/count was equivalent to read coverage. Next, read counts were compared between samples within each preparation to ensure reproducibility of the techniques. Both PAC-seq and RNA-seq exhibited a high level of correlation between raw read counts within each preparation method with a Pearson's correlation coefficient of R^2 greater than 0.9 for all comparisons of samples within methods (Figure 4A). However, although there is good overall concordance, when examining the read counts between preparation methods there is greater variance (Figure 4B). This is likely due to the differences in preparation techniques. While PAC-seq specifically produces only one read per mRNA, traditional RNA-seq produces multiple fragments for each transcript that can be biased due to a number of factors including transcript length, GC content, and primer annealing conditions. Nonetheless, the high concordance between the individual replicates of each method indicates that these variances are likely to be reproducible and systematic.

Transcript counts were used as input for the DESeq2 pipeline to measure differential expression. Principle Component Analysis of the four datasets (PAC vs Click and LacZ vs INT11) (Figure 5A) show clear clustering of replicate experiments. As was expected, the CS libraries are strongly separated relative to the PAC-seq libraries along the first component (79% of variance), while the KD libraries are both separated along the second component (17% of variance). The tight clustering of sample types indicates a high level of confidence in reproducibility by both methods and the similar distance of knockdown effect gives confidence in the changes detected. DE analysis revealed a global upregulation of mRNAs upon INTS11 KD with both techniques (Figures 5B and C and Supplementary Datasets 1 and 2). When requiring Bonferroni-adjusted p-values >0.05 and fold change in read count >2 , 510 and 540 genes were upregulated and 110 and 141 were down-regulated for CS and PAC-seq respectively. At these statistic thresholds, it therefore appears that PAC-seq is capable of detecting a greater number of DE events, suggesting a greater sensitivity.

An example of the output data is shown in Table 2 for genes that were upregulated, unchanged, and down-regulated upon INTS11 knockdown. This table demonstrates the high level of correlation between PAC-seq and RNA-seq with respect to the fold changes in each gene. The fold change for mRNAs with adjusted p-values <0.05 are shown in the scatter plot in Figure 5D. The Fold Change trends along the $x=y$ axis with a Pearson correlation coefficient of $R^2=0.8961$ ($p<0.0001$) indicating that PAC-seq and RNA-seq perform similarly in estimation of overall changes in gene expression between two conditions and there is general agreement between the two sequencing techniques. Interestingly these fold changes correspond well irrespective of variance in the number of read counts detected by each method. There was a minor difference in the fold change of INTS11 expression due to a number of reads detected in RNA-seq data for the region targeted by the INTS11 dsRNA for knockdown even though the total RNA was purified using Oligo(dT) beads. This indicated contaminating presence of these RNA molecules that were able to carry over with the beads and a possible shortcoming of the RNA-seq method.

4. Discussion

Differential gene expression analysis is routine and widespread in the biomedical sphere. In typical RNA-seq approaches for DE analyses, to achieve sufficient depth and sensitivity while obtaining a sufficient number of biological and technical replicates, a large volume of next-generation sequencing data must be acquired often at a considerable cost. Therefore, 3'end targeted methods of estimating gene abundance are becoming viewed as cost-effective alternatives to standard RNA-seq. These approaches generate a simplified snapshot of the transcriptome and so miss many important factors in understanding transcriptome diversity, such as the presence of transcript isoforms and splice variants.

Nonetheless, there have recently been a number of approaches developed for the identification of 3'ends of poly-adenylated transcripts (reviewed in [23]). However, among the simplest strategies employing the fewest manipulations are the QuantSeq[5] and the Poly(A)-ClickSeq (PAC-seq) pipeline[2]. We have previously demonstrated how PAC-seq can be used to locate and quantify the poly(A) sites of mRNAs and measure changes in poly(A) tail positions upon manipulation of the cleavage and poly-adenylation machinery

and described how this reveals the regulation of poly(A) site selection by CFI25m [2]. Alternative poly-adenylation (APA) is emerging as a critical factor in the regulation of mRNA translation efficiency and transcript stability[24]. Furthermore, 3'UTR shortening has been clearly associated with a number of disease states such as glioblastoma [25].

Here we have demonstrated that PAC-seq simultaneously provides a simple and cost-effective strategy for measuring absolute levels and relative changes in gene expression, thus providing a dual-use tool. When compared side-by-side to random-primed RNA-seq of poly(A)-enriched mRNA, we demonstrated that PAC-seq can capture changes in gene expression in good agreement with the RNA-seq data, despite only sequencing a fraction of the mRNA terminal exon. Additionally, PAC-seq detects a greater number of genes that display significant (adjusted $p < 0.05$) changes in expression level. This is achieved without the need for any sample preparation or target enrichment. Rather, crude cellular RNA is used as an input for the library preparation. This simplifies the process, and also removes the biases and cost that may be imparted by these sample treatments.

In general, there are numerous advantages to employing a 3' end targeted sequencing methods such as PAC-seq for characterizing quantitative changes in the transcriptome by NGS:

1. By only sequencing the 3' end of an mRNA transcript, sequencing depth is limited to one read per transcript. This saves on the amount of sequencing that must be performed when compared to standard RNA-seq where reads may be found across the length of the RNA transcript. It is possible with our method that a poly(A) tail is in fact primed twice, and that a lagging reverse transcriptase strand-displaces the upstream cDNA [26]. This would generate two cDNAs (or more) from a single mRNA. However, this would only be likely for very long poly(A) tails, which are scarce.
2. As all transcripts only have one poly(A) tail, this negates the need for computational normalization of read counts assigned to a mRNA as a function of their length. This process is normally complicated by the presence of alternative splice variants as well as incomplete annotation in some reference genomes.
3. Very small transcripts that would otherwise receive very low sequence coverage in standard RNA-seq approaches and may be missed can be accurately quantified in an equivalent manner to longer transcripts when using PAC-seq.
4. Long poly(A) tails are generally found only on mature mRNA transcripts rather than transcripts undergoing nascent transcription, which may contain introns or arise as a result of abortive transcription, which would not represent translating mRNAs.
5. If ribo-depletion were employed rather than poly(A) tail selection, RNA-seq would sequence a large amount of non-mRNA material including transposable elements and ncRNAs. Unless these transcripts were poly-adenylated, they would not be detected by PAC-seq.

PAC-seq strength lies in its property of focusing solely on the 3' end of mRNA transcripts in order to estimate RNA abundance. However, for some transcripts, this may expose a vulnerability in scenarios where the 3'UTR of a specific mRNA is highly structured, has unusual nucleotide composition, or contains some other intrinsic property that reduces or inhibits reverse transcription. In these cases, the abundance of these transcripts may be under-estimated relative to other transcripts. This may be ameliorated by varying RT conditions and/or using highly processive RT enzymes such as Superscript IV or TGIRT [27]. However, for the purposes of DE analysis, small biases in 3' end sequencing would likely be similar between multiple replicates of the same terminal exon and therefore would be accounted for when measuring DE.

By sequencing only the 3'ends of transcripts, computational pipelines for estimating gene abundance and determining changes in gene expression are greatly simplified and streamlined, as demonstrated in this manuscript. However, there remain two key disadvantages. Firstly; there is a small probability that the 3'end of some transcripts are identical to one another (e.g. between paralogs, duplicated genes and pseudo-genes). This would result in ambiguous mapping of the PAC-seq reads where standard RNA-seq would take advantage of SNPs found throughout the transcript in order to assign isoform/paralog abundance [28]. Secondly, due to the PCR cycles required for RNA-seq library construction, PCR duplication is a common artifact in RNA-seq libraries that may return erroneous and aberrant estimates of gene abundance. This artifact can be computationally limited using de-duplication methods that collapse multiple reads into a single read if they share the same start and stop sites in the reference genome. This may lose some reads that coincidentally had the same mapping positions, but these cases are infrequent when read coverage is not excessive. However, in PAC-seq, all reads are enriched in a small portion of the mRNA transcript (usually the terminal exon) and so the probability of two reads having the same mapping coordinates becomes very high. Therefore, standard computational de-duplication cannot be employed. However, unique molecular identifiers [29] can easily be appended to the beginning of the click-adaptor providing a more robust method for collapsing identical reads.

In addition to priming from the poly(A) tail during RT-PCR, priming from A-rich sequences within mRNAs is also possible when using an oligo-dT primer. This may result in absolute read counts being elevated for particular transcripts. However, we would expect the frequency of internal priming from A-rich tracts to correlate with gene abundance and be conserved among multiple replicates. Therefore, this factor would not perturb differential gene expression analysis. In PAC-Seq, as priming is not anchored to the 3'UTR/poly(A) tail junction, small portions of the poly(A)-tail are copied yielding non-primer derived A's in the final read data. These extra A's allow us to control for internal and/or mis-priming by requiring a greater number of A's in a mapped read than were T's used in the oligo-dT primer thus allowing confident annotation of poly(A)-sites. Moreover, in the absence of this control (e.g. when using short reads), internal priming can also be ameliorated by only counting reads that map to the expected 3' terminal exons of mRNA transcripts when determining transcript abundance.

Despite some limitations of 3' end sequencing approaches, we envisage the unique features of techniques such as PAC-seq will lend themselves to be critical tools in high-volume investigations (large sample numbers with multiple repeats) requiring analysis of gene abundance or changes in expression. Additionally, PAC-seq may be broadly applicable for the detection and titering of other poly(A) tailed species such as positive sense RNA viruses. While there is scope for improvement in the basic PAC-seq pipeline, such as adjustments to the specific click-chemistry used in adaptor ligation [30] to improve read-through efficiency, PAC-seq provides a robust and highly scalable method for both PAS annotation and gene expression analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

A.R. is supported by start-up funds from the University of Texas Medical Branch, Galveston and a University of Texas System Rising STARs Award. E.J.W. is supported by The Welch Foundation [H1189] and the Cancer Prevention Research Institute of Texas [RP140800]. P.J. is supported by the National Institutes of Health [R03-CA223893-01]. The authors thank the UTMB Next-Generation Sequencing core, particularly Dr. Tom Wood, Dr. Steve Widen and Jill Thompson, for providing critical support and advice.

Abbreviations

PAC-seq	Poly(A)-ClickSeq
CS	ClickSeq
NGS	Next-Generation Sequencing
PAS	Poly(A) Site
APA	Alternative Polyadenylation
DE	Differential Expression
AzNTPs	3'-azido deoxynucleotides
INTS11	Integrator Subunit 11

References

- [1]. Lin Y, Golovkina K, Chen ZX, Lee HN, Negron YL, Sultana H, et al. Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. BMC Genomics. 2016;17:28. [PubMed: 26732976]
- [2]. Routh A, Ji P, Jaworski E, Xia Z, Li W, Wagner EJ. Poly(A)-ClickSeq: click-chemistry for next-generation 3'-end sequencing without RNA enrichment or fragmentation. Nucleic Acids Res. 2017.
- [3]. Xiong Y, Soumillon M, Wu J, Hansen J, Hu B, van Hasselt JGC, et al. A Comparison of mRNA Sequencing with Random Primed and 3'-Directed Libraries. Scientific reports. 2017;7:14626. [PubMed: 29116112]
- [4]. Routh A, Ji P, Jaworski E, Xia Z, Li W, Wagner EJ. Poly(A)-ClickSeq: click-chemistry for next-generation 3'-end sequencing without RNA enrichment or fragmentation. bioRxiv. 2017.

- [5]. Moll P, Ante M, Seitz A, Reda T. QuantSeq 3' mRNA sequencing for RNA quantification. *Nature methods*. 2014;11:i–iii.
- [6]. Galatro TF, Holtman IR, Lerario AM, Vainchtein ID, Brouwer N, Sola PR, et al. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat Neurosci*. 2017;20:1162–71. [PubMed: 28671693]
- [7]. Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes Dev*. 2011;25:1770–82. [PubMed: 21896654]
- [8]. Szkop KJ, Nobeli I. Untranslated Parts of Genes Interpreted: Making Heads or Tails of High-Throughput Transcriptomic Data via Computational Methods: Computational methods to discover and quantify isoforms with alternative untranslated regions. *Bioessays*. 2017;39.
- [9]. Routh A, Head SR, Ordoukhanian P, Johnson JE. ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3'-Azido cDNAs. *Journal of molecular biology*. 2015.
- [10]. El-Sagheer AH, Sanzone AP, Gao R, Tavassoli A, Brown T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2011;108:11338–43. [PubMed: 21709264]
- [11]. Jaworski E, Routh A. ClickSeq: Replacing Fragmentation and Enzymatic Ligation with Click-Chemistry to Prevent Sequence Chimeras. *Methods Mol Biol*. 2018;1712:71–85. [PubMed: 29224069]
- [12]. Schneider I Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *Journal of embryology and experimental morphology*. 1972;27:353–65. [PubMed: 4625067]
- [13]. Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res*. 2012;22:939–46. [PubMed: 22267522]
- [14]. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–i90. [PubMed: 30423086]
- [15]. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21. [PubMed: 23104886]
- [16]. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature methods*. 2015;12:357–60. [PubMed: 25751142]
- [17]. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44:D733–45. [PubMed: 26553804]
- [18]. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*. 2014;30:923–30.
- [19]. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. [PubMed: 25516281]
- [20]. Routh A, Head SR, Ordoukhanian P, Johnson JE. ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3'-Azido cDNAs. *J Mol Biol*. 2015;427:2610–6. [PubMed: 26116762]
- [21]. Jaworski E, Routh A. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLoS pathogens*. 2017;13:e1006365. [PubMed: 28475646]
- [22]. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology*. 2010;11:R106. [PubMed: 20979621]
- [23]. Zhang Y, Carrion SA, Zhang Y, Zhang X, Zinski AL, Michal JJ, et al. Alternative polyadenylation analysis in animals and plants: newly developed strategies for profiling, processing and validation. *International Journal of Biological Sciences*. 2018;14:1709–14. [PubMed: 30416385]
- [24]. Masamha CP, Wagner EJ. The contribution of alternative polyadenylation to the cancer phenotype. *Carcinogenesis*. 2018;39:2–10. [PubMed: 28968750]
- [25]. Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, et al. CFI_{m25} links alternative polyadenylation to glioblastoma tumour suppression. *Nature*. 2014;510:412–6. [PubMed: 24814343]

- [26]. Acevedo A, Brodsky L, Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*. 2014;505:686–90. [PubMed: 24284629]
- [27]. Qin Y, Yao J, Wu DC, Nottingham RM, Mohr S, Hunicke-Smith S, et al. High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA*. 2016;22:111–28. [PubMed: 26554030]
- [28]. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*. 2012;7:562–78. [PubMed: 22383036]
- [29]. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A*. 2011;108:20166–71. [PubMed: 22135472]
- [30]. Miura F, Fujino T, Kogashi K, Shibata Y, Miura M, Isobe H, et al. Triazole linking for preparation of a next-generation sequencing library from single-stranded DNA. *Nucleic Acids Res*. 2018;46:e95. [PubMed: 29846671]

- Poly(A)-ClickSeq is a click-chemistry based method for making 3'-end NGS libraries.
- Poly(A) sites and transcript abundance can be simultaneously measured in a facile manner.
- We performed gene expression analysis in Integrator S11 knockdown experiments.
- Changes in transcript abundance measured by RNAseq and PAC-Seq correlate well.
- Poly(A)-ClickSeq is a simple and cost-effective method for gene expression analysis.

Fig. 1. Schematic of PAC-seq library preparation.

(1) A non-anchored 21-dT primer with a partial p7 adapter sequence is used to reverse synthesize mRNA containing a poly(A) tail. 3'-azidodeoxynucleotides (AzCTP, AzATP, and AzGTPs) are added to terminate the RT-reaction randomly depending on dNTP:AzVTP ratio upon reaching the 3'UTR. (2) Click chemistry is used to attach a p5 universal adapter to the azido-terminated 3' end of the cDNA. (3) PCR is then performed using the p5 Universal primer and Indexing primers completing the p7 end. (4) The PCR product is electrophoresed on an agarose gel and the band corresponding to products with a 3'UTR segment of approximately 230nt is selected for purification. (5) Representation of final PAC-seq preparation that is used for downstream sequencing.

Fig. 2. Flowchart of data analysis.

Raw reads in fastq format are filtered and trimmed before alignment to the reference genome using STAR aligner. Sorted reads are then input to the DESeq2 pipeline following a matrix input format. This allows comparison of DE and normalized counts both through visualization by creating bedgraphs and fold change analysis.

Fig. 3. Example bedgraphs of outputs from data analysis.

(A) Total number of mapped reads from each sample type to different annotated regions of mRNAs is illustrated (dark cyan: terminal exons; light cyan: all exons; white: introns; yellow: 200nts downstream of annotated 3' end; green: unannotated regions). Approximately 75–80% of all the reads map to exonic sequences (cyan), with a greater proportion of these mapping to the terminal exons (dark cyan) in the PAC-seq datasets. (B) Comparison of read coverage between the two sample preparation methods from the same source RNA. The upper light grey RNA-seq by ClickSeq method shows read coverage throughout the gene body due to the random nature of the RT-reaction steps. The lower dark grey PAC-seq coverage is specific for the 3'UTR. (C) Representative genes that showed down-regulation from LacZ control upon INTS11 knockdown, grey and red bars respectively. (D) Representative genes that showed up-regulation from LacZ control upon INTS11 knockdown, grey and red bars respectively.

Fig. 4. Comparison of read counts between samples and preparation methods.

(A) Non-normalized Log_{10} read counts between the three biological replicates of the INTS11 KD for the PAC-seq and RNA-seq methods, upper and lower rows respectively. The tight data patterns of both preparation methods indicate a high level of reproducibility. (B) Non-normalized Log_{10} read counts of the three biological replicates of INTS11 KD comparing the read counts for all transcripts between the PAC-seq and RNA-seq methods.

Fig. 5. Output visualization of DESeq2 results.

(A) Principal Component Analysis of gene read count distributions. The majority of the variance is between PAC-seq and RNA-seq preparation methods at 79%. The remaining variance along the PC2 component, at 17%, was due to the effects of INTS11 knockdown compared to LacZ. (B and C) Log₂ fold change versus the P-adjusted value for the DE detected in PAC-seq and RNA-seq respectively. Genes that had a fold change greater than two and a P-adjusted value less than 0.05 have been colored red for those upregulated and blue for those downregulated upon INTS11 knockdown. (D) Comparison of log₂ fold change between methods. The log₂ fold change of genes that had an adjusted p-value <0.05 was highly correlated between methods with an R₂ value of 0.8961 and a slightly lower fold change detected by RNA-seq at 0.8933 of the value for PAC-seq.

Table 1:

Read Counts in each RNA-seq and PACseq replicate associated with selected mRNAs (bold names shown in Figure 3)

	RNA-seq Libraries						PAC-seq Libraries					
Gene Name	CS Z-1	CS Z-2	CS Z-3	CS 11-1	CS 11-2	CS 11-3	PAC Z-1	PAC Z-2	PAC Z-3	PAC 11-1	PAC 11-2	PAC 11-3
<i>regucalcin</i>	66	174	58	1019	2163	1177	76	162	89	1743	2137	2761
<i>mgl</i>	114	158	134	2455	2503	1975	107	196	196	2756	2288	3858
<i>Col4a1</i>	207	305	321	1112	1527	1172	127	137	199	704	1000	1261
<i>Arc1</i>	7110	8664	8027	16028	23132	14556	15670	20101	24821	58040	71451	80399
<i>CG33926</i>	513	665	590	865	1840	869	2944	4347	5395	8583	10598	13487
<i>Adgf-A</i>	1503	1946	1417	3237	4246	3022	493	603	700	1472	1087	2109
<i>Cys</i>	76	54	94	148	219	170	1907	2377	2862	4898	5252	7897
<i>spir</i>	3799	3896	3512	7323	6881	5053	1778	2015	2237	4822	3793	5074
<i>Su(Tpl)</i>	17119	17954	17963	10355	11170	8605	9901	10225	14267	9139	7060	11429
<i>CtBP</i>	12593	16444	14870	7198	9241	7114	16885	20821	25103	15024	14442	20559
<i>Ncoa6</i>	12071	10237	9735	7376	6315	4814	4360	3797	4400	3875	2410	3979
<i>RpS4</i>	8893	10246	10045	4250	5175	4228	20592	23793	29169	16447	13365	21986
<i>HmgZ</i>	6815	9213	9714	1816	2812	2207	14205	20442	24310	5954	6337	9735
<i>Tctp</i>	10700	15929	12795	1601	2967	1628	19428	31234	32365	4564	6905	8026
<i>IntS11</i>	652	981	648	34	95	62	690	865	1172	15	30	47

Table 2:

DESeq2 output for RNA-seq and PAC-seq replicates (bold names shown in Figure 3)

Gene Name	RNA-seq Z mean	RNA-seq INT11 mean	RNA-seq log2 Fold Change	RNA-seq p-adjusted	PAC-seq Z mean	PAC-seq INT11 mean	PAC-seq log2 Fold Change	PAC-seq p-adjusted
<i>regucalcin</i>	55.68	1354.87	4.60	1.08E-41	124.67	3143.08	4.66	3.51E-43
<i>mg1</i>	78.04	2233.50	4.84	3.28E-170	185.38	4140.88	4.48	2.10E-150
<i>Col4a1</i>	160.37	1216.72	2.92	3.98E-40	172.92	1402.62	3.02	1.05E-40
<i>Arc1</i>	4593.57	16986.00	1.89	1.90E-35	22642.76	100741.45	2.15	2.40E-46
<i>CG33926</i>	340.56	1103.87	1.70	2.21E-16	4701.16	15483.25	1.72	1.50E-17
<i>Adgf-A</i>	935.20	3347.50	1.84	6.14E-48	675.81	2148.53	1.67	4.54E-37
<i>Cys</i>	44.00	171.21	1.97	9.13E-21	2679.45	8433.48	1.65	1.45E-21
<i>spir</i>	2171.39	6202.61	1.51	2.41E-18	2286.26	6525.54	1.51	3.27E-18
<i>Su(Tpl)</i>	10280.56	9682.88	-0.09	7.42E-01	12916.78	12933.92	0.00	9.95E-01
<i>CtBP</i>	8458.02	7533.67	-0.17	2.77E-01	23563.77	23577.10	0.00	9.97E-01
<i>Ncoa6</i>	6241.29	5984.12	-0.06	9.14E-01	4829.10	4829.54	0.00	1.00E+00
<i>RpS4</i>	5643.40	4383.64	-0.36	1.46E-03	27673.26	24171.62	-0.20	1.46E-01
<i>HmgZ</i>	4964.61	2177.84	-1.19	1.62E-13	21941.21	10271.61	-1.10	1.29E-11
<i>Tctp</i>	7560.20	1933.95	-1.97	8.22E-21	30985.18	9291.25	-1.74	1.79E-16
<i>IntS11</i>	436.22	59.52	-2.87	3.76E-20	1014.07	42.20	-4.58	3.50E-45