TAG-Seq

Dhivya Arasappan

(Thanks for Some Slides from Misha Matz, and Emily Tallman)
Whole RNA-Seq typically means small number of biological replicates

(A) RNA-seq publications by year in ecology and evolution

(B) Replicate usage by eco-evolutionary RNA-seq studies testing differential expression

Todd, E. et al, Molecular Ecology, 2016
3’ TAGSEQ- An Alternative to Whole RNA-Seq

Targeting the 3’ prime end of RNA

Fig. 1 Overview of the protocol used to prepare 3’ cDNA tag libraries from total RNA. RNA was fragmented at the beginning to eliminate biases resulting from differences in transcript lengths. First-strand cDNA was primed with a modified oligo-dT containing primer to target 3’ ends. Each sample was prepared with a sample-specific oligonucleotide barcode, then quantified and pooled prior to sequencing.

Meyer, E. et al, Molecular Ecology, 2011
WHY TAGSEQ?

• Cheaper to sequence 3’ end instead of the entire RNA.
• Fewer reads need to be sequenced.
• Lib prep is 10 times cheaper
• Amount of input RNA required is less.
• You can still identify differential expression.
WHY NOT TAGSEQ?

- If you want to assemble a transcriptome
- If you want to look at differential splicing
- If you want to identify polymorphisms in gene sequences
TAG-Seq Template Switching

mRNA

Fragmentation by heat

Reverse transcription

T-primer
Tag-Seq Template Switching

mRNA

5’ ---------------------- AAAAAAAAAAAAAAAAA 3’

Fragmentation by heat

Reverse transcription

RT

T-primer
Tag-Seq Template Switching

mRNA

5’ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAAAAA...

Fragmentation by heat

__________________________AAAAAA...AAAA

Reverse transcription

RT

T-primer
Tag-Seq Template Switching

mRNA

5’ __________________________________________________________________________ 3’

Fragmentation by heat

Reverse transcription

RT

CCC

T-primer
Tag-Seq Template Switching

mRNA

Fragmentation by heat

Reverse transcription

Switch oligo

T-primer
Tag-Seq Template Switching

mRNA

5'  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3'

Fragmentation by heat

Reverse transcription
Tag-Seq Template Switching

Amplification and barcoding

bc1

bc2
Tag-Seq Template Switching

Amplification and barcoding

Sequencing: R 50b

bc1

bc2
Amplification and barcoding

Sequencing:  R  

bc1  

bc2  

One tag per mRNA
Unique:
- NNMW (64-fold)+
- First 20 bases (Frag point)
Is Tag-Seq Accurate?

Rozenberg, A. et al, BioTechniques, 2018
Is Tag-Seq comparable to whole RNA-Seq?

Tag-Seq recovers known concentrations of mRNA (ERCC controls) with more accuracy than whole mRNASeq.

Fig. 1 Regression of observed vs. expected ERCC transcripts shows TagSeq has higher adjusted $R^2$ values for four different biological samples prepared with both methods (paired $t$-test, $t = 18.63$, d.f. = 3, $P < 0.001$).
Is Tag-Seq comparable to whole RNA-Seq?

Fig. 3 Breakdown of control mRNAs by abundance class shows that TagSeq recovers mRNAs better than TruSeq, especially at lower abundances. Light grey bars are TagSeq, dark grey bars are whole mRNASeq. Fences indicate standard error.

Lohman et al, Molecular Ecology Resources, 2016
How do we analyze RNA-Seq data?

• **STEP 1**: PREPROCESS RAW DATA
• **STEP 2**: EVALUATE QUALITY
• **STEP 3**: MAP TO REFERENCE, ASSESS RESULTS
• **STEP 4**: QUANTIFY TRANSCRIPTS
• **STEP 5**: TEST FOR DIFFERENTIAL EXPRESSION
• **STEP 6**: VISUALIZE AND PERFORM OTHER DOWNSTREAM ANALYSIS
STEP 1 - Preprocess Raw Data

https://github.com/z0on/tag-based_RNAseq

Tag-Seq Read Structure
STEP 1- Preprocess Tag-Seq Data

Trim Header

Degenerate Base: More than one base possibility

N = A, T, G, or C
M = A or C
W = A or T

Before
[NNMWGGGNN]AAAAAAAATTTTGGGCGCT

After
AAAAAATTTTGGGCGCT
STEP 1- Preprocess Tag-Seq Data

Remove Duplicates
Reads that share the same degenerate header and the first 20 bases of the sequence.

```
[NNMWGGGNN]AAAAAAAAATTGGGCCT
```

Barcode

```
N N M W
4 * 4 * 2 * 2 = 64 Possible Barcodes
```
STEP 1- Preprocess Tag-Seq Data

Filter polyA and low quality reads
Reads that share the same degenerate header and the first 20 bases of the sequence.
STEP 1- Preprocess Tag-Seq Data

Filter polyA and low quality reads
Reads that share the same degenerate header and the first 20 bases of the sequence.

PolyAs:
fastx_clipper -a AAAAAAAAA -l 20 -Q33

Low Quality Reads:
fastq_quality_filter -Q33 -q 20 -p 90
How do we analyze RNA-Seq data?

• **STEP 1**: PREPROCESS RAW DATA
• **STEP 2**: EVALUATE QUALITY
• **STEP 3**: MAP TO REFERENCE, ASSESS RESULTS
• **STEP 4**: QUANTIFY TRANSCRIPTS
• **STEP 5**: TEST FOR DIFFERENTIAL EXPRESSION
• **STEP 6**: VISUALIZE AND PERFORM OTHER DOWNSTREAM ANALYSIS
Pseudoaligners don’t seem to perform as well with Tag-Seq Data

<table>
<thead>
<tr>
<th>Sample</th>
<th>Kallisto Mapping %</th>
<th>BWA Mapping %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1-WLD1_S28</td>
<td>61.2%</td>
<td>77.75%</td>
</tr>
<tr>
<td>A2-WLD2_S36</td>
<td>60.4%</td>
<td>77.68%</td>
</tr>
<tr>
<td>A3-WLD3_S44</td>
<td>59.5%</td>
<td>77.10%</td>
</tr>
<tr>
<td>A4-WLD4_S52</td>
<td>68.4%</td>
<td>79.17%</td>
</tr>
</tbody>
</table>

Mapping Summary

- For Tag-Seq, unspliced and spliced mappers would both work fine.
- Most often, you will map to the transcriptome instead of genome.
- For Tag-Seq, Avoid pseudoaligners like Kallisto.
Where is Tag-Seq Being Used?

Identifying Gene Expression Signatures for Drug Compounds

Ye, C. et al, Nature Communications, 2018
Where is Tag-Seq Being Used?

Studying Gene Expression Responses of Coral Larvae to Temperature Changes

Meyer, E. et al, Molecular Ecology, 2011
Where is Tag-Seq Being Used?

- Whole RNA-Seq for assembling a transcriptome.
- Tag-Seq for large number of samples for gene expression profiling.