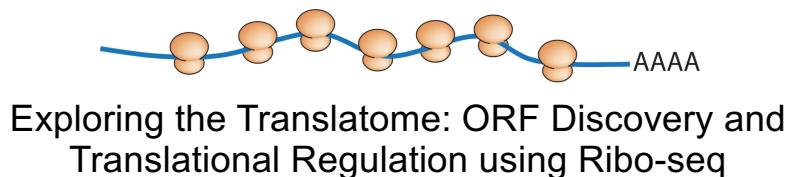


ASPB 2024



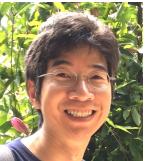
Survey link
[https://shorturl.at/
KbtyO](https://shorturl.at/KbtyO)



Polly Hsu



Isaiah Kaufman



Larry Wu



Michigan State University

@PollyHsuLab

1

How much do you know about Ribo-seq?

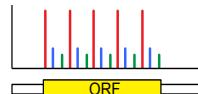
- A. I have never heard of it before this workshop
- B. I heard about it and want to know if I can apply it to my research
- C. I just started doing Ribo-seq or about to
- D. I have Ribo-seq data and want to learn its data analysis
- E. I am expert in Ribo-seq but want to hear the update or your way of doing it



We don't know everything, but we are happy to tell you what we know and use

3

Motivation for this workshop



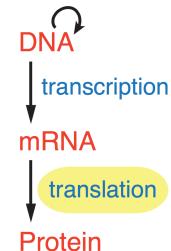
Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*

Polly Yingshan Hsu^a, Lorenzo Calviello^{b,c}, Hsin-Yen Larry Wu^{d,1}, Fay-Wei Li^{a,e,f,1}, Carl J. Rothfels^{e,f}, Uwe Ohler^{b,c}, and Philip N. Benfey^{a,g,2} 2016 PNAS, PMID: 27791167

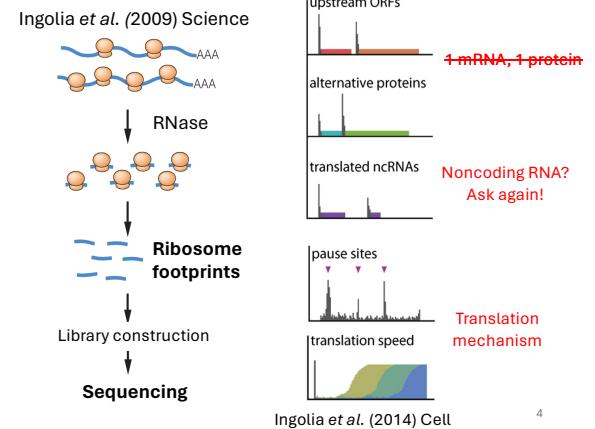
- We often received questions about Ribo-seq experiments & analysis
- Some commercial reagents have been discontinued. Now what?
- What works? What doesn't work? How to work around?

2

Caught in the act: studying translatomes using Ribo-seq

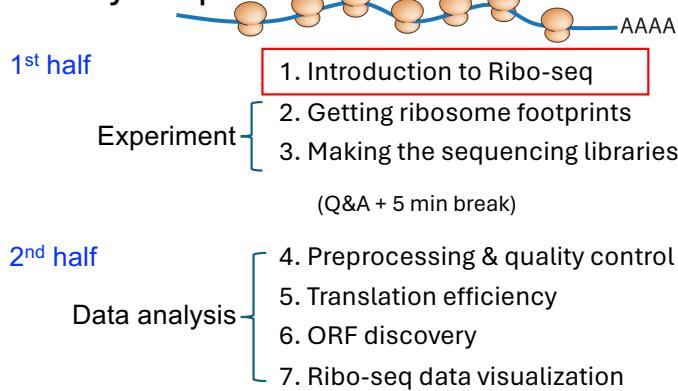


- A proxy of proteome
- Higher sensitivity
- Novel translation events & regulation



4

Today's topics

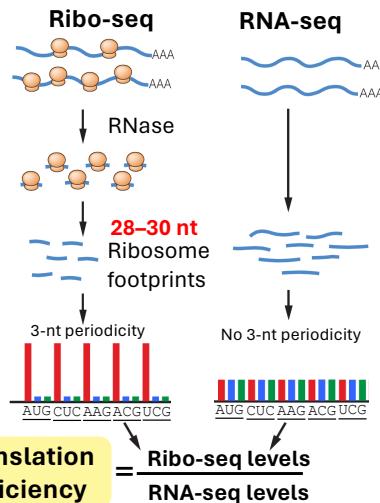


What is Ribo-seq?

- Ribo-seq = Ribosome footprints + sequencing
- A.k.a. ribosome profiling
- Established in budding yeast (Ingolia *et al.* 2009 Science)
- First plant Ribo-seq studies (Arabidopsis)
 - Liu *et al.* 2013 Plant Cell (Shu-Hsing Wu Lab)
 - Zoschke *et al.* 2013 Plant Cell (Alice Barkan Lab)

Two main applications of Ribo-seq

- 1) Identify translated ORFs
- 2) Quantify translation efficiency



Actively translating ribosomes display 3-nt periodicity

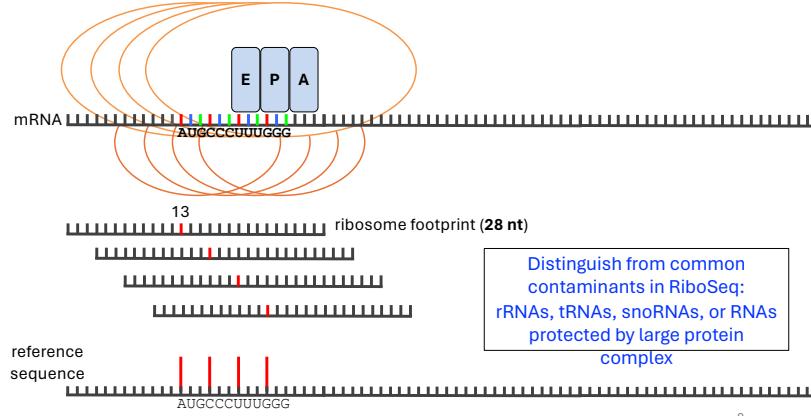
5

6

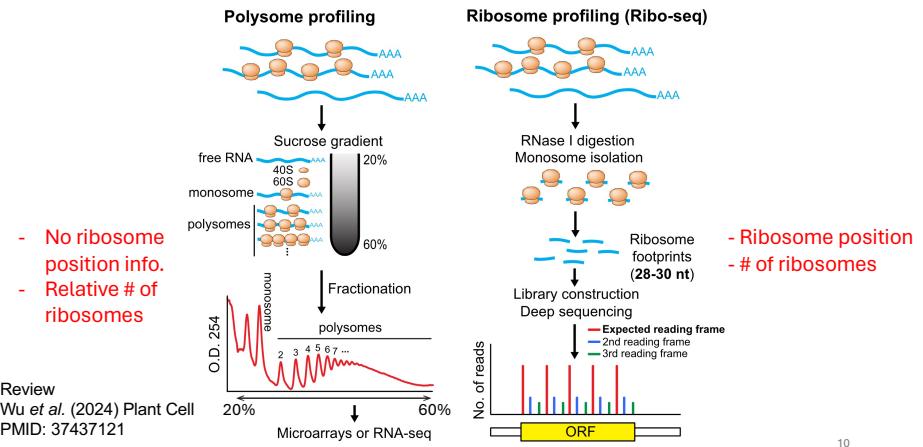
7

8

Actively translating ribosomes display 3-nt periodicity



Polysome profiling vs. ribosome profiling?



Today's topics

1st half

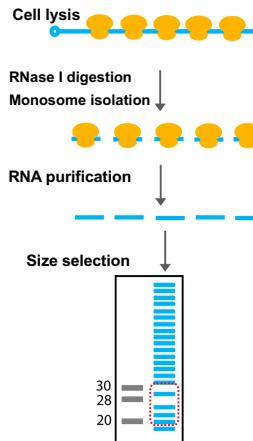
- Experiment
1. Introduction to Ribo-seq
 2. Getting ribosome footprints
 3. Making the sequencing libraries
- (Q&A + 5 min break)

2nd half

- Data analysis
4. Preprocessing & quality control
 5. Translation efficiency
 6. ORF discovery
 7. Ribo-seq data visualization

11

Getting ribosome footprints - overview

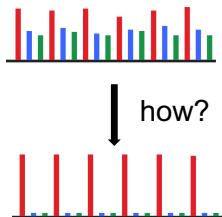


1. Immobilize ribosomes and prepare cell lysates
 - Flash freezing and grind with liquid nitrogen
 - Resuspend in lysis buffer (**critical**) containing **cycloheximide** (for cytosolic ribosomes) and/or **chloramphenicol** (for organelle ribosomes)
 2. RNase I digestion (**critical**)
 - RNase I works well for eukaryotes
 - RNase amount needs to be tested for different tissues
 3. Monosome isolation
 - Sucrose cushion, sucrose gradient
 - Size exclusion columns
 4. RNA isolation (including ribosome footprints)
 5. rRNA depletion (**recommend to do after size selection**)
 6. Size selection of ribosome footprints in the gel (**critical**)
- 12

Challenge 1: How to improve 3-nt periodicity?

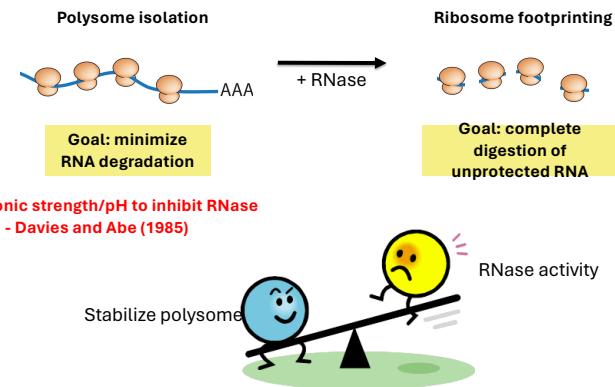
- Goal: Strong 3-nt periodicity → ORF discovery

In 2015, the best Ribo-seq reads in frame % = 62%



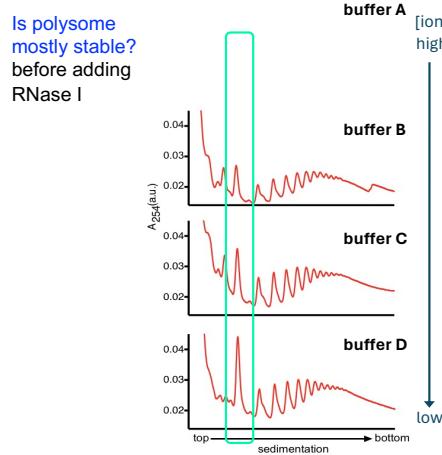
13

Why is it difficult to get good 3-nt periodicity?



14

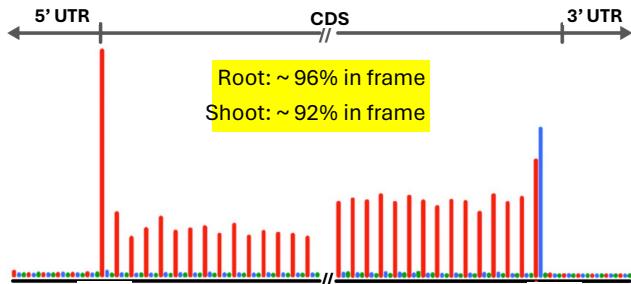
Which buffer is optimal for ribosome footprinting?



Hsu et al. (2016) PNAS
PMID: 27791167

15

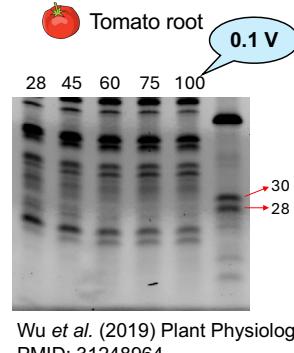
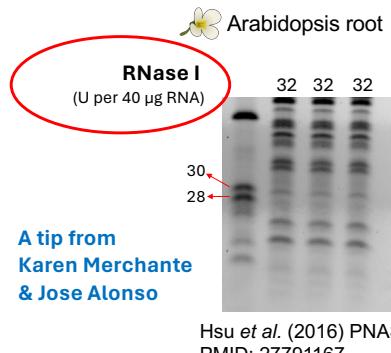
Super-resolution Ribo-seq in Arabidopsis



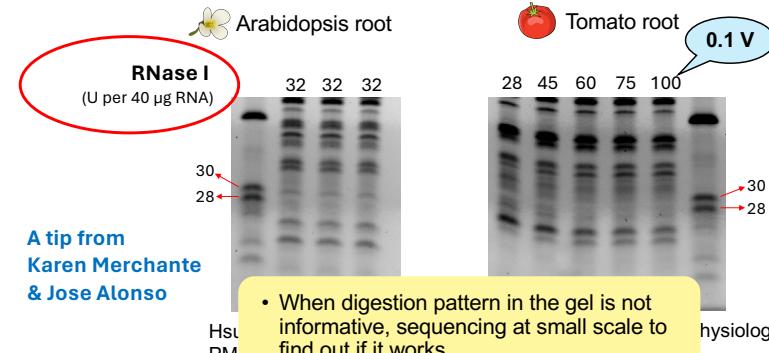
Hsu et al. (2016) PNAS
PMID: 27791167

16

Is there an easy way to tell the digestion is good?



Is there an easy way to tell the digestion is good?



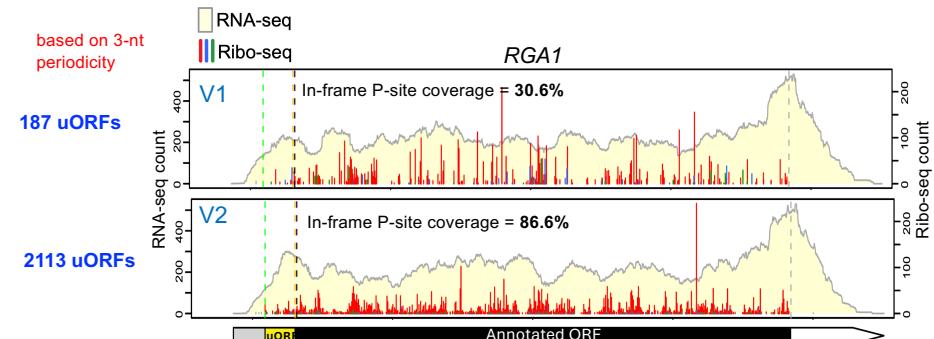
Challenge 2: How to improve Ribo-seq coverage?

- Goal: Better coverage → short ORF identification



Shorter ORFs, like uORFs, are often missed by conventional identification methods

Super-resolution Ribo-seq v2 improved uORF identification



Wu et al. (2024) Plant Cell PMID: 38000896

How did we improve the coverage?

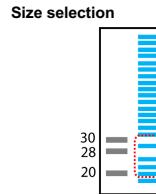
What could improve the coverage?

- Increase starting footprints (input)?
- Minimize ribosome footprint loss?



Getting ribosome footprints - overview

1. **Immobilize ribosomes and prepare cell lysates**
 - Flash freezing and grind with liquid nitrogen
 - Resuspend in **lysis buffer (critical)** containing **cycloheximide** (for cytosolic ribosomes) and/or **chloramphenicol** (for organelle ribosomes)
2. **RNase I digestion (critical)**
 - RNase I works well for eukaryotes
 - RNase amount needs to be tested for different tissues
3. **Monosome isolation**
 - Sucrose cushion, sucrose gradient
 - Size exclusion columns
4. **RNA isolation (including ribosome footprints)**
5. **rRNA depletion (recommend to do after size selection)**
6. **Size selection of ribosome footprints in the gel (critical)**



Minimize ribosome footprint loss

Increase starting materials

21

22

Summary: getting good ribosome footprints



- **Good 3-nt periodicity:**
 - Lower ionic strength & pH buffering in Lysis buffer
 - Check the digestion pattern in the gel could be useful
 - **Tissue: buffer ratio**
- **Good coverage:**
 - Increase input and reduce loss by doing size selection before rRNA depletion

23

Today's topics

1st half

- Experiment
1. Introduction to Ribo-seq
 2. Getting ribosome footprints
 3. Making the sequencing libraries
- (Q&A + 5 min break)

2nd half

Data analysis

4. Preprocessing & quality control
5. Translation efficiency
6. ORF discovery
7. Ribo-seq data visualization

24

Updated workflow and new challenges

- Obtain ribosome footprints

1. Immobilize ribosomes and prepare cell lysates
2. RNase I digestion (most critical)
3. Isolate monosome
4. Isolate RNA (ribosome footprints)
5. Size selection of ribosome footprints in the gel
- 6. rRNA depletion**
(illumina RiboZero has been discontinued)

- Generate sequencing library (illumina Ribo-seq kit) **(discontinued)**

25

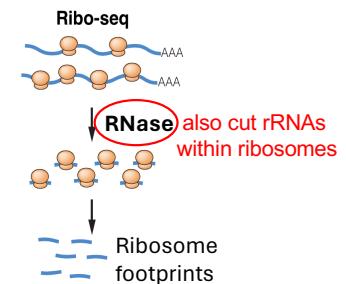
Topic 1. rRNA depletion in Ribo-seq

- In **RNA-seq**, ~80% of total RNAs are rRNAs
→ design probes that hybridize to individual (~intact) rRNAs

- In **Ribo-seq**, >90% of RNAs are rRNAs
→ the probes must recognize 'many' rRNA 'fragments'

- **Customized oligos** specific to your experimental condition
- **Illumina RiboZero:** (discontinued)
biotinylated/antisense to the entire rRNAs

Still available as part of
TruSeq RNA-seq kit

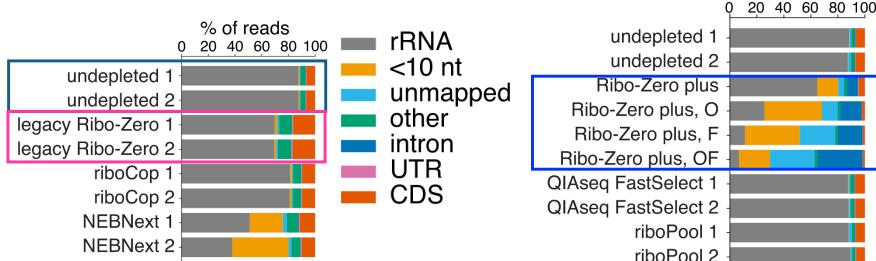


26

Alternatives to RiboZero for rRNA depletion?

- **Zinshteyn et al. (2020) RNA, PMID: 32503920**

- Tested available commercial kits at the time for Ribo-seq (human cells)
- Found the **original RiboZero** is the best, the alternative is the **customized oligos**
- **Warning:** 'RiboZero plus' is a nuclease-mediated method and shows off-target activity
- New kits keep emerging, so reach out for the update later!



27

Topic 2. Customized library construction

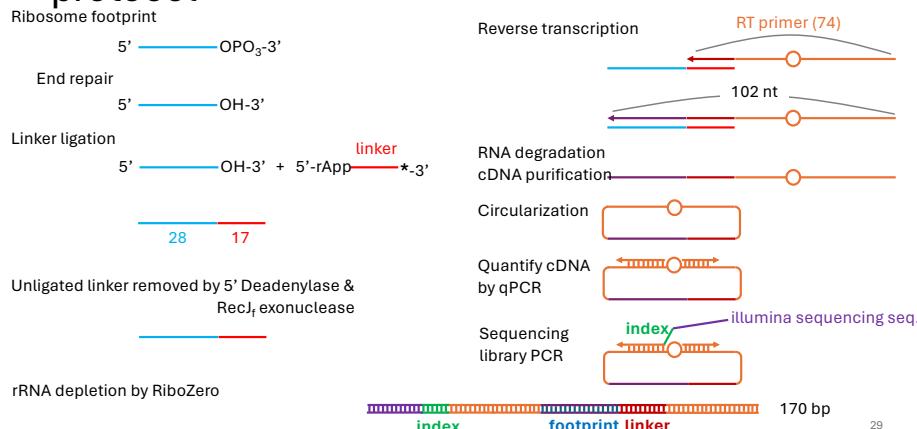
- **108 steps + 80 reagents/equipment** 🧬

- **Detailed recipes and catalog numbers:**

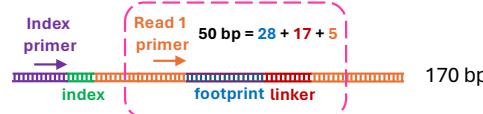
Wu and Hsu (2022) Plant Methods
PMID: 36195920

28

Workflow of our current library construction protocol



Q: Which illumina sequencing should you use for Ribo-seq?



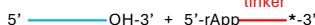
- A. Single-end 50 bp sequencing since ribosome footprints are ~28 nt
- B. Single-end 100 bp sequencing since I need to sequence both the ribosome footprint and the index
- C. Single-end 150 bp sequencing since the library is ~170 bp
- D. Pair-end 150 bp sequencing because the longer the better
- E. Please send help!!!

30

Topic 3. Other things to consider

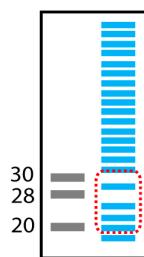
• Unique molecular identifier

- Reduce and correct bias introduced by ligation and PCR



• Size selection of ribosome footprints

- Focus on 28-30 nt?
- Cutting wider range (20-30 nt, or 20-40 nt) increases RF diversity but also increases contamination
- Shorter RFs have a poorer mapping rate



31

Summary: Ribo-seq library construction

• rRNA depletion:

- RiboZero (or RiboPOOL)
- Customized oligos

• Library construction

- Our detailed protocol is available: Wu & Hsu (2022) Plant Methods, PMID: 36195920
- Single-end 50 bp sequencing is enough!

• Other considerations

- UMI
- Size selection

32

Today's topics

1st half

- Experiment
 - 1. Introduction to Ribo-seq
 - 2. Getting ribosome footprints
 - 3. Making the sequencing libraries
- (Q&A + 5 min break)
- Survey link**
<https://shorturl.at/KbtyO>
- 2nd half
 - 4. Preprocessing & quality control
 - 5. Translation efficiency
 - 6. ORF discovery
 - 7. Ribo-seq data visualization

33

Today's topics

1st half

- Experiment
 - 1. Introduction to Ribo-seq
 - 2. Getting ribosome footprints
 - 3. Making the sequencing libraries

(Q&A + 5 min break)

2nd half

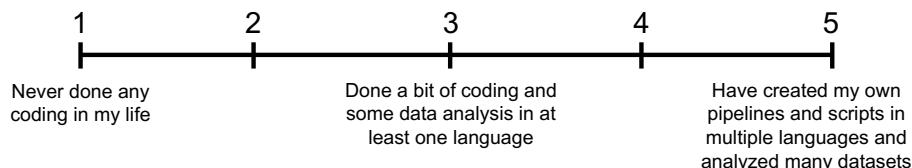
- Data analysis
 - 4. Preprocessing & quality control
 - 5. Translation efficiency
 - 6. ORF discovery
 - 7. Ribo-seq data visualization



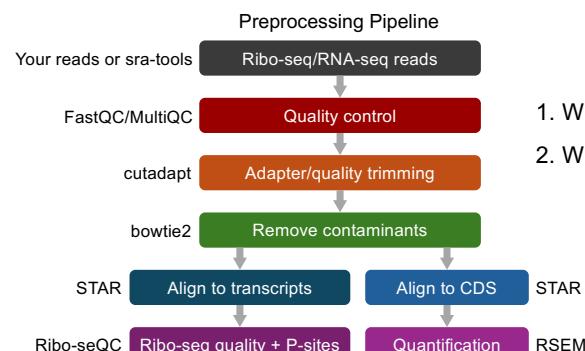
Isaiah Kaufman

34

How much coding and sequencing analysis experience do you have?



35



https://github.com/kaufm202/Riboseq_ASPB_2024

36

sra-tools – downloading reads if analyzing publicly available data

Inputs: SRA IDs found in publication / NCBI Gene Expression Omnibus (GEO) / NCBI Short Read Archive (SRA)

Outputs: raw read files in FASTQ format (sometimes read from NCBI will already be adapter/quality trimmed)

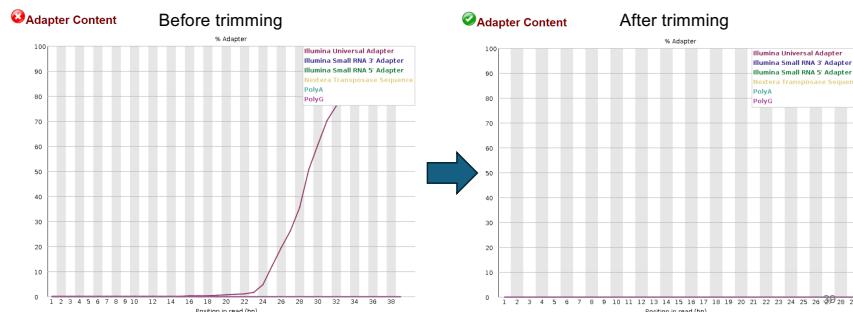
```
Read ID → @SRR3498209.1 SN638:766:HC5HBCXX:1:1101:1200:2077 length=50
Sequence → ATGGNTGATTTAGCTTCCAAGAACAAAGGAGATCGGAAGGCACACGTCTG
+SRR3498209.1 SN638:766:HC5HBCXX:1:1101:1200:2077 length=50
Quality scores → DDDD#<DGHHIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
@SRR3498209.2 SN638:766:HC5HBCXX:1:1101:1222:2081 length=50
AGAGNGGGTGAGAGCCCCGTCGTGCCCGAGATCGGAAGAGCACACGTCT
+SRR3498209.2 SN638:766:HC5HBCXX:1:1101:1222:2081 length=50
DDDD#<DGHHIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

53

`cutadapt` – remove adapters and low quality reads

Inputs:

Outputs: FASTQ files with trimmed reads



Ribo-seq reads **MUST**
have adapter!

FastQC – quality report

Inputs: reads in FASTQ format

Outputs: HTML quality report

MultiQC – combine multiple FastQC reports

Inputs: directory containing FastQC report

Outputs: a single HTML quality report summarizing all the samples

Will run FastQC 3 times, once for raw reads, once for cutadapt trimmed reads, once for bowtie2 filtered reads

20

bowtie2 – remove rRNA, tRNA, snRNA, snoRNA

Inputs:

- reads from cutadapt
 - FASTA file containing contaminant sequences
 - (from genome annotation, NCBI search, or from closely-related species)

Outputs: FASTQ files with cleaned reads

can BLAST search to identify more contaminants, but don't stress too much about removing all overrepresented sequences

4

STAR – splice aware genome alignment

Inputs:

- reads from bowtie2
- genome FASTA file
- genome annotation in GTF format
- max read length – 1 (check FastQC)

Outputs: BAM files (Aligned.sortedByCoord.out.bam and Aligned.toTranscriptome.out.bam) and log files

Requires separate index for Ribo-seq and RNA-seq

For Ribo-seQC, align to full transcripts

For quantification (RSEM), align to only CDS

Polyploidy may cause high rate of multimapping for short reads

Log.final.out	
Started job on	Jun 11 18:48:28
Started mapping on	Jun 11 18:48:29
Finished on	Jun 11 18:49:40
Mapping speed, Million of reads per hour	1427.76
Number of input reads	28158672
Average input read length	28
UNIQUE READS:	
Uniquely mapped reads number	22439186
Uniquely mapped reads %	79.69% →
Average mapped length	28.75
Number of splices: Annotated (gtdb)	1783488
Number of splices: GT/AG	1767288
Number of splices: GC/AG	12961
Number of splices: AT/AC	2705
Number of splices: Non-canonical	526
Mismatch rate per base, %	0.06%
Deletion rate per base	0.00%
Deletion average length	1.25
Insertion rate per base	0.00%
Insertion average length	1.01
Number of reads mapped to one single loci	4871248
% of reads mapped to multiple loci	17.30% →
Number of reads mapped to too many loci	52431
% of reads mapped to too many loci	0.19%
UNMAPPED READS:	
Number of reads unmapped: too many mismatches	0
% of reads unmapped: too many mismatches	0.00%
Number of reads unmapped: too short	773854
% of reads unmapped: too short	2.75% →
Number of reads unmapped: other	21953
% of reads unmapped: other	0.08%
CHIMERIC READS:	
Number of chimeric reads	0
% of chimeric reads	0.00% 41

Ribo-seQC – Ribo-seq quality + P-sites

Inputs:

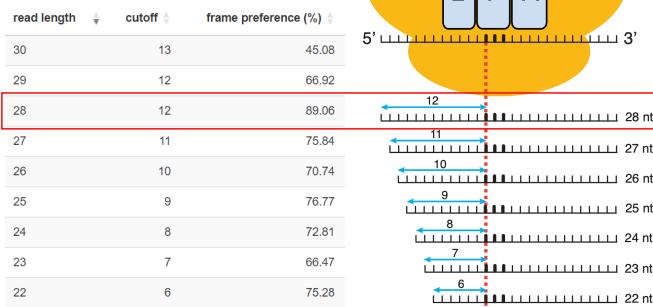
- Ribo-seq Aligned.sortedByCoord.out.bam files from full transcript-aligned STAR
 - Merge and subset as desired
- genome FASTA file
- genome annotation in GTF format (same one from STAR)
- list of circular sequences (names as they appear in genome FASTA)

Outputs:

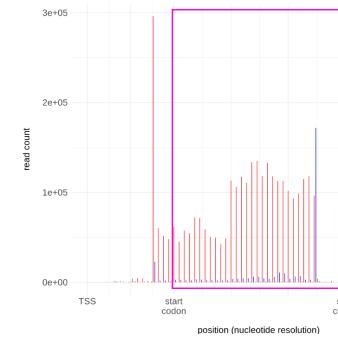
- HTML report
- P-site offset values (from HTML or _P_sites_calcs)
- P-site files (_P_sites_minus.bedgraph and _P_sites_plus.bedgraph)

42

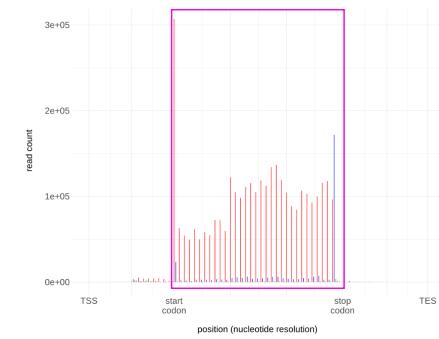
Ribo-seQC P-site offsets



Before offset (first nuc of read)

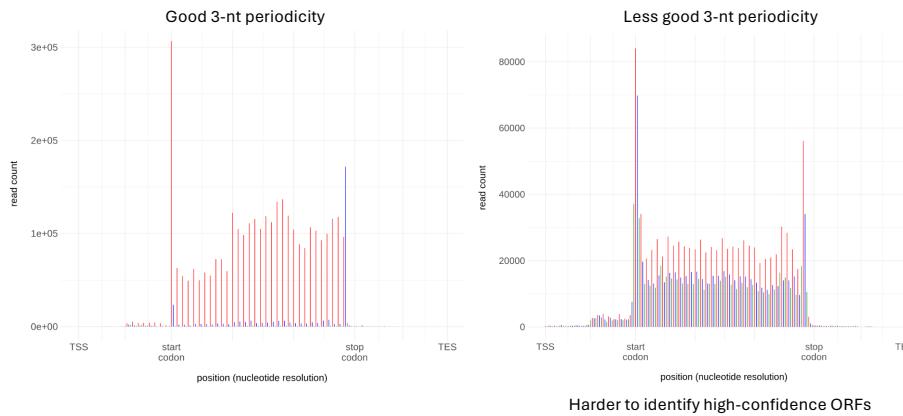


After offset (P-sites)



43

44



45

RSEM – Ribo-seq/RNA-seq quantification

Inputs:

- Aligned.toTranscriptome.out.bam from CDS-aligned STAR
- genome FASTA file
- genome annotation in GTF format (same one from STAR)
- Read length mean and SD for single-end Ribo-seq and RNA-seq
- Input same index directory used in STAR**

- Outputs:** two tab-delimited files with gene and transcript level quantification (.genes.results and .isoforms.results)

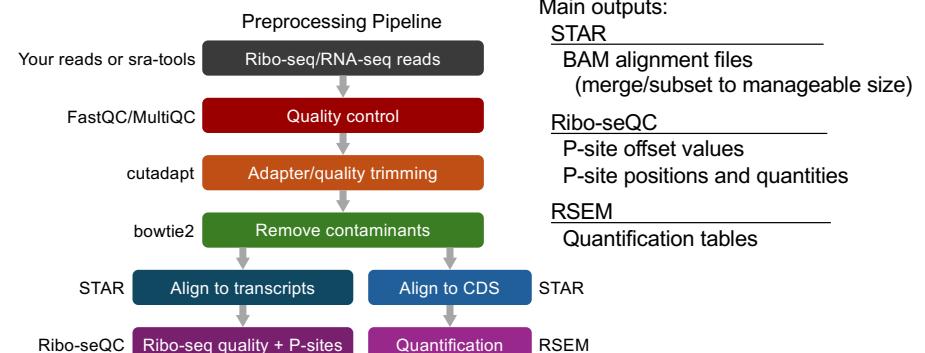
transcript_id	gene_id	length	effective_length	expected_count	TPM	FPKM	IsoPct
AT1G01010..1	AT1G01010	1290	1260.51	96.00	2.64	3.00	100.00
AT1G01020..1	AT1G01020	738	708.51	0.00	0.00	0.00	0.00
AT1G01020..2	AT1G01020	576	546.51	0.00	0.00	0.00	0.00
AT1G01020..3	AT1G01020	711	681.51	0.00	0.00	0.00	0.00
AT1G01020..4	AT1G01020	711	681.51	0.00	0.00	0.00	0.00
AT1G01020..5	AT1G01020	597	567.51	84.00	5.12	5.83	100.00
AT1G01020..6	AT1G01020	315	285.51	0.00	0.00	0.00	0.00
AT1G01030..1	AT1G01030	1077	1047.51	0.00	0.00	0.00	0.00
AT1G01030..2	AT1G01030	1008	978.51	72.00	2.55	2.90	100.00

46

If you encounter an error...

- Probably one of these issues
 - Syntax – missing punctuation, incorrect file paths
 - Memory – exceeded allowed RAM or disc space
 - GTF issue – program ran into problem while parsing GTF**
↳ ~95% of hard-to-resolve errors

Just because the code runs, does not mean the output is correct
be diligent about checking the outputs for any issues



https://github.com/kaufm202/Riboseq_ASMB_2024

47

48

Today's topics

1st half

- Experiment
- 1. Introduction to Ribo-seq
- 2. Getting ribosome footprints
- 3. Making the sequencing libraries

(Q&A + 5 min break)

2nd half

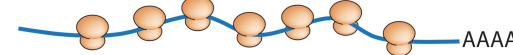
- Data analysis
- 4. Preprocessing & quality control
- 5. Translation efficiency
- 6. ORF discovery
- 7. Ribo-seq data visualization



Isaiah Kaufman

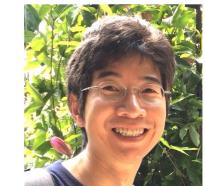
49

The 3rd Part of the Ribo-seq informatics workshop



Data analysis

- 5. Preprocessing & quality control
- 6. Translation efficiency
- 7. ORF discovery
- 8. Ribo-seq data visualization

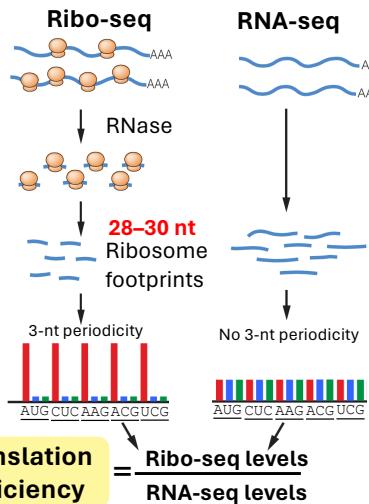


Larry Wu

50

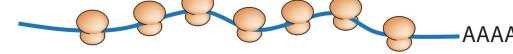
Two main applications of Ribo-seq

- 1) Identify translated ORFs
- 2) Quantify translation efficiency



51

Tools I will cover



1. Xtail (Differential translation)
2. RiboTaper (ORF identification)
3. RiboPlotR (visualization)
4. GWIPS and TAIR websites (visualization)

52

Analyzing Differential Translation

Analyzing differential translation (i.e., differential translation efficiency) means that we need to estimate the variations of both Ribo-seq and RNA-seq.

One important consideration is that you want more biological replicates than RNA-seq (differential expression analysis). Ribo-seq is expensive and time consuming, so people tend to use less replicates :)

53

Analyzing Differential Translation

Preparing for Xtail

Map RNA-seq & Ribo-seq with STAR Map to CDS regions

https://github.com/kaufm202/Riboseq_ASMB_2024/blob/main/Translation%20Efficiency.md
<https://rdrr.io/github/xryanglab/xtail/f/vignettes/xtail.Rmd>

55

Analyzing Differential Translation

Xtail as an [example](#). → DEseq2 (PMID: 25516281)
deltaTE (PMID: 31763789)

ARTICLE

Received 6 Jul 2015 | Accepted 29 Feb 2016 | Published 4 Apr 2016

DOI: 10.1038/ncomms11194

OPEN



Genome-wide assessment of differential translations with ribosome profiling data

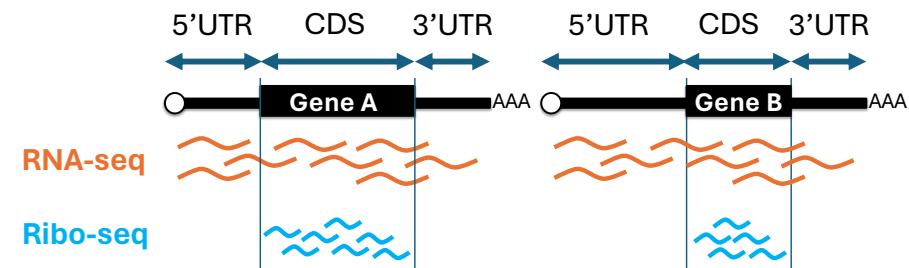
Zhengtao Xiao^{1,2,3,4}, Qin Zou^{1,3,4,5}, Yu Liu^{1,2,3,4} & Xuerui Yang^{1,2,3,4}

https://github.com/kaufm202/Riboseq_ASMB_2024/blob/main/Translation%20Efficiency.md (example code)
<https://rdrr.io/github/xryanglab/xtail/f/vignettes/xtail.Rmd> (xtail vignettes)

54

Analyzing Differential Translation

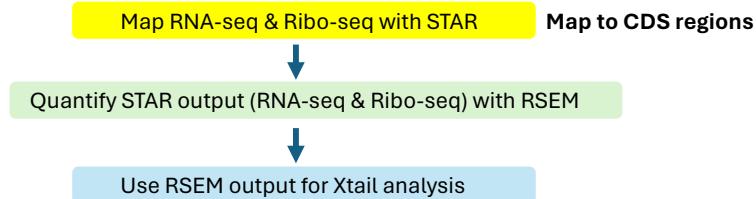
Use reads map to CDS (coding sequence) not the whole mRNA



56

Analyzing Differential Translation

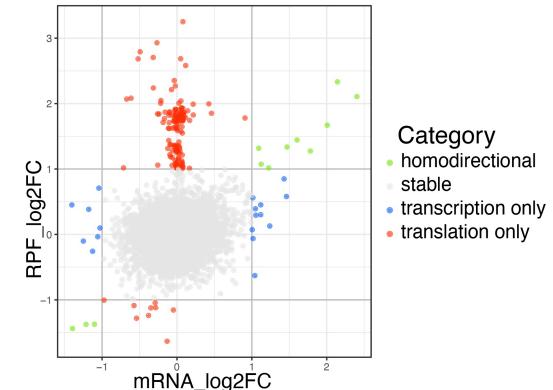
Xtail as an example.



https://github.com/kaufm202/Riboseq_ASMB_2024/blob/main/Translation%20Efficiency.md (example code)
<https://rdrr.io/github/xryanglab/xtail/f/vignettes/xtail.Rmd> (xtail vignettes)

57

Xtail output log2 fold change plot



58

1. Xtail (Differential translation)

2. RiboTaper (ORF identification)

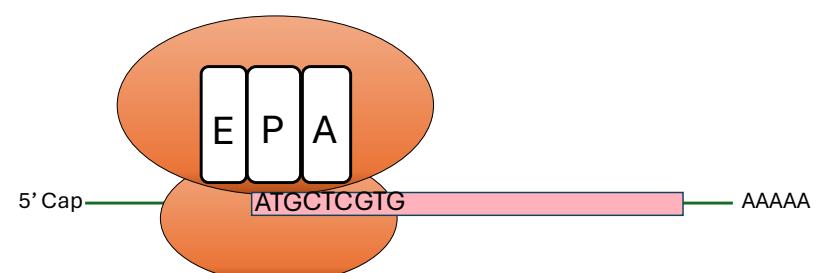
Need:

- (1) P-site offset
- (2) RNA-seq and Ribo-seq BAM files from STAR alignment

3. RiboPlotR (visualization)

4. GWIPS and TAIR websites (visualization)

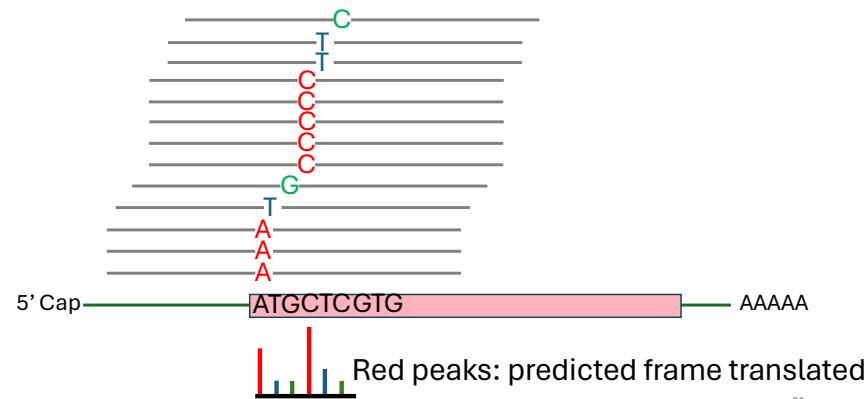
Intro 1: How could we detect 3-nt periodicity?



59

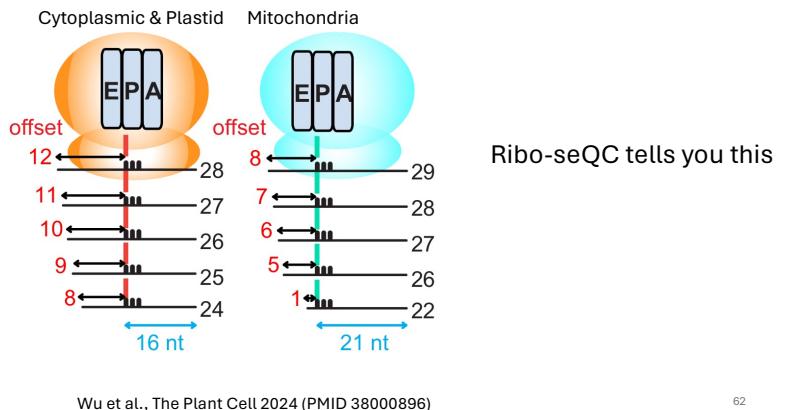
60

Intro 1: How could we detect 3-nt periodicity?



61

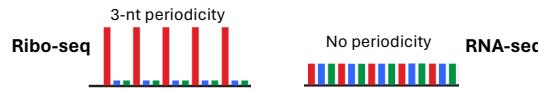
Intro 2: What is P-site offset (cut-off)?



62

ORF discovery with RiboTaper

RiboTaper (Calviello et al., Nat Methods 2016, PMID: 26657557)



1. Run Ribo-seQC to obtain P-site offset (i.e., cutoff) for each read length
2. Make RiboTaper annotation files
3. Provide both RNA-seq and Ribo-seq bam files (from STAR output)
4. Run RiboTaper

RiboTaper link: https://ohlerlab.mdc-berlin.de/software/RiboTaper_126/
 Example code: https://github.com/kaufm202/Riboseq_ASMB_2024/tree/main

63

RiboTaper output files

```

May 9 05:35 all_calculations_ccdsgenes_annotation_new
May 9 00:05 Centered_RNA
May 7 21:35 data_tracks
May 9 01:24 find_ORF_results.pdf
May 9 09:06 RIBO_max.fasta
May 9 07:52 orfs_found
May 9 09:37 orfs_found_nonccds
May 9 10:20 ORFs_genes_found
May 9 10:20 ORFs_max_file
May 9 10:11 ORFs_NONCCDS ←
May 9 10:20 protein_db_max.fasta
May 8 23:43 RIBO_max.fasta
May 9 01:39 RIBO_max_all_processed
May 9 05:39 quality_check_plots.pdf
May 9 05:11 results_ccds
May 9 05:31 results_exonscds
May 9 05:31 results_nccds
May 9 05:35 results_nccds_annotation
May 8 22:41 RIBO_best.bam
May 9 00:28 RIBO_best_counts_ccds
May 9 02:08 RIBO_best_counts_exonscds
May 9 02:32 RIBO_best_counts_nonccds
May 8 22:39 RIBO_unique.bam
May 9 00:09 RIBO_unique_counts_ccds
May 9 01:37 RIBO_unique_counts_exonscds
May 9 02:30 RIBO_unique_counts_nonccds
May 8 23:30 RNA_best.bam
May 9 00:49 RNA_best_counts_ccds
May 9 02:28 RNA_best_counts_exonscds
May 9 02:51 RNA_best_counts_nonccds
May 9 02:35 RNA_best_nccds.bam
May 9 00:26 RNA_unique_counts_ccds
May 9 01:56 RNA_unique_counts_exonscds
May 9 03:30 RNA_unique_counts_nonccds
May 9 01:38 tmp_cde
May 9 10:13 tmp_nccds
May 9 10:20 translated_ORFs_filtered_sorted.bed
May 9 10:16 translated_ORFs_sorted.bed

```

ORFs_max_filt contains the information for discovered ORFs
 (It is a tab delim file which can be easily analyzed in R)

64

RiboTaper output files

1. Xtail (Differential translation)
2. RiboTaper (ORF identification)
- 3. RiboPlotR (visualization)**
4. GWIPS and TAIR websites (visualization)

RiboPlotR

Wu and Hsu

Plant Methods 2021

PMID: 34876166

Need:

- (1) P-site information from RiboTaper (or Ribo-seQC)

For more info: <https://github.com/hsinyenwu/RiboPlotR>

65

```

May 9 05:35 all_calculations_cdsgenes_annot_new
May 9 00:05 Centered_RNA
May 7 21:35 data_tracks
May 9 10:20 final_ORF_results.pdf
May 9 09:36 ORFs_CDS
May 9 10:20 orfsfound
May 9 09:37 p_sites_all_nccds
May 9 10:20 ORFs_genes_found
May 9 10:20 ORFs_max
May 9 10:20 ORFs_max_file
May 9 09:36 exonsccds
May 9 10:20 protein_db_max.fasta
May 8 23:43 p_sites_all
May 7 20:59 p_sites_all_processed
May 9 09:36 quality_nccds_plots.pdf
May 9 09:36 results_nccds
May 9 05:31 results_exonsccds
May 9 05:31 results_nccds
May 9 05:31 results_nccds_annot
May 9 05:31 results_nccds_bam
May 9 00:28 RIBO_best_counts_ccds
May 9 02:02 RIBO_best_counts_exonsccds
May 9 03:32 RIBO_best_counts_nccds
May 8 22:32 RIBO_unique_bam
May 9 01:36 RIBO_unique_counts_ccds
May 9 01:37 RIBO_unique_counts_exonsccds
May 9 03:11 RIBO_unique_counts_nccds
May 8 23:30 RNA_best_bam
May 9 01:22 RNA_best_counts_ccds
May 9 03:51 RNA_best_counts_nccds
May 8 23:05 RNA_unique_bam
May 9 01:22 RNA_unique_counts_ccds
May 9 01:38 RNA_unique_counts_exonsccds
May 9 03:30 RNA_unique_counts_nccds
May 9 09:36 tmp_ccds
May 9 10:13 tmp_nccds
May 9 10:20 translated_ORFs_filtered_sorted.bed
May 9 10:16 translated_ORFs_sorted.bed

```

P_sites_all a bed file for Ribo-seq P-sites

Process **P_sites_all** file for RiboPlotR
(see <https://github.com/hsinyenwu/RiboPlotR>)

Counts Chr Position Strand

1	1	1000000	+
3	1	10000007	+
3	1	10000010	+
3	1	10000016	+
1	1	10000018	+
4	1	10000019	+

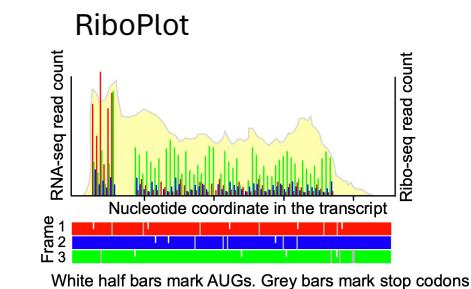
(Position: the 1st nucleotide of the P-site)

66

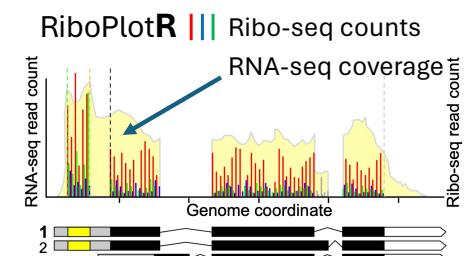
Ribo-seq data visualization

Genome Browser	
Manually enter gene id	Do not require coding experience Y
Visualize all isoforms together	Y
Gene structure	Y
Select different isoforms	
Periodicity with 3 colors	
Loop for multiple genes	
Generate pdf for publication	

67



<https://pythonhosted.org/riboplot/riboplot.html>



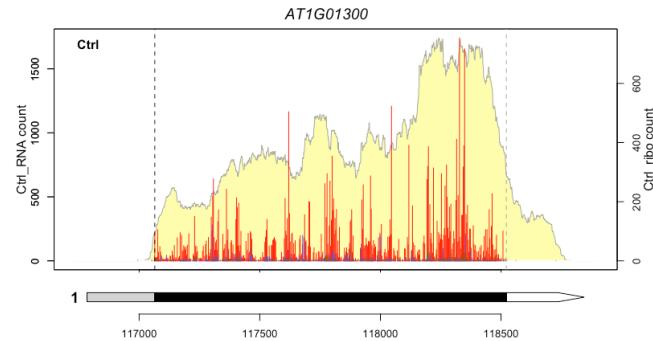
<https://github.com/hsinyenwu/RiboPlotR>

68

Ribo-seq visualization with RiboPlotR examples

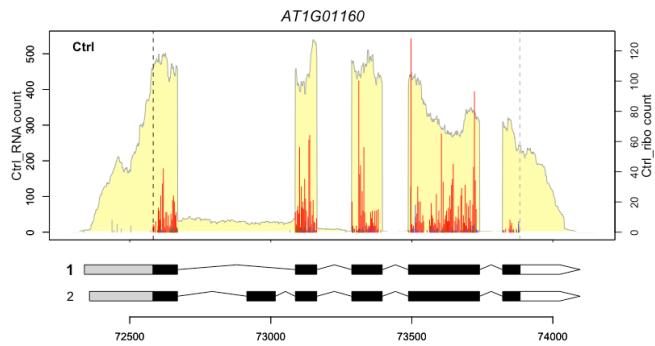
69

Ribo-seq visualization: example



70

Ribo-seq visualization: example

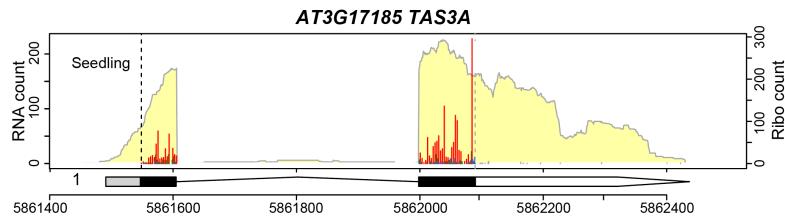


71

Ribo-seq visualization with RiboPlotR sORFs or uORFs

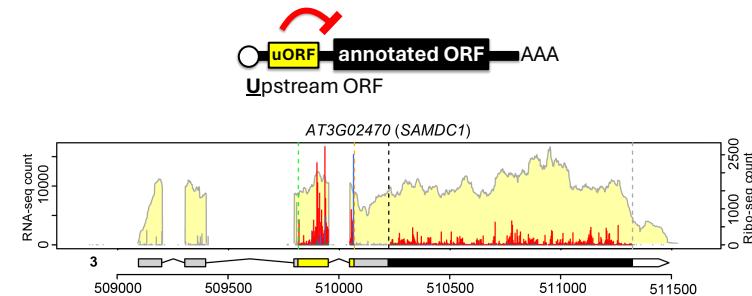
72

Ribo-seq visualization: translation events on non-coding RNAs



73

Ribo-seq visualization with RiboPlotR: Translated uORFs



74

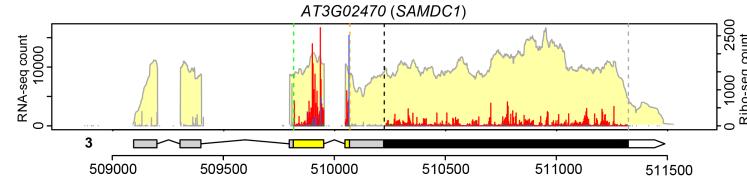
The problem for visualizing uORFs

Seqname	Source	Feature	Start	Stop	Score	Strand	Frame	Attribute
1	Araport11	gene	3631	5899	.	+	.	gene_id "AT1G01010"; gene_biotype "protein_coding";
1	Araport11	mRNA	3631	5899	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	exon	3631	3913	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	exon	3996	4276	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	exon	4486	4605	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	exon	4706	5095	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	exon	5174	5326	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	exon	5439	5899	.	+	.	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	CDS	3760	3913	.	+	0	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	CDS	3996	4276	.	+	2	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	CDS	4486	4605	.	+	0	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	CDS	4706	5095	.	+	0	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	CDS	5174	5326	.	+	0	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";
1	Araport11	CDS	5439	5630	.	+	0	gene_id "AT1G01010"; transcript_id "AT1G01010.1"; gene.biotype "protein_coding";

GTF/GFF3 files only allows one CDS (encodes one peptide) for each transcript.

75

What to do if you want to visualize uORFs?



- RiboPlotR will take in one separate GTF file for “each uORF”. This means **one uORF needs one gtf file**.

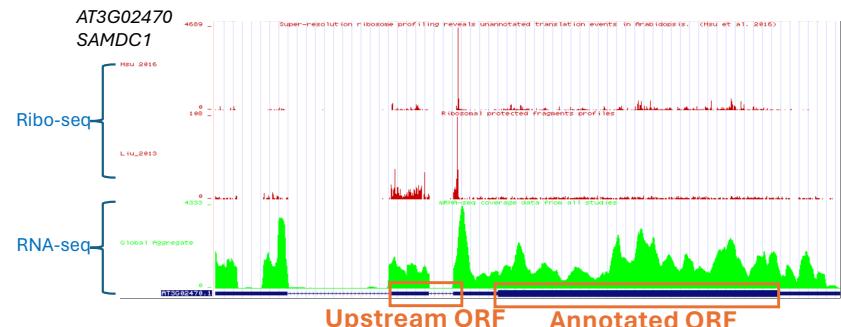
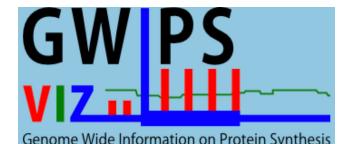
See https://github.com/hsinyenwu/RiboPlotR_addition for how to generate uORF gtf. See <https://github.com/hsinyenwu/RiboPlotR> for generating plots with uORFs.

76

1. Xtail (Differential translation)
2. RiboTaper (ORF identification)
3. RiboPlotR (visualization)
4. **GWIPS and TAIR websites (visualization)**

GWIPS website

- <https://gwips.ucc.ie/cgi-bin/hgGateway>
- Arabidopsis (3 datasets), Maize (1 dataset)



78

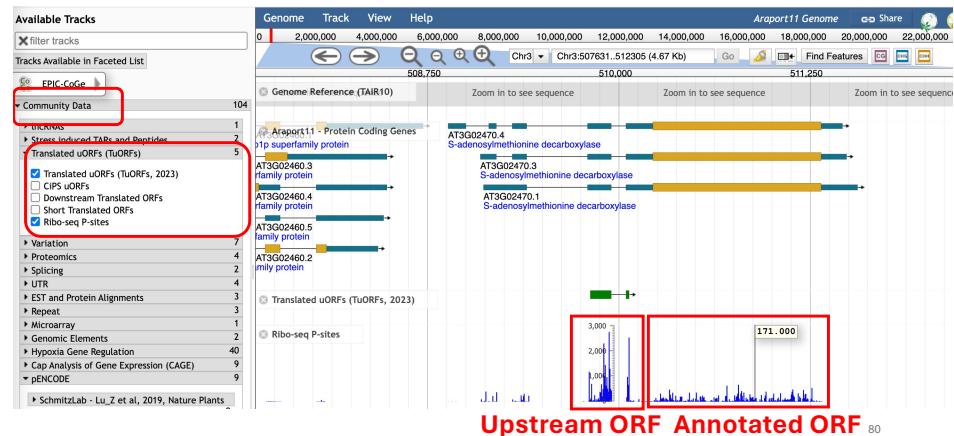
Our latest Ribo-seq data on TAIR

The Arabidopsis Information Resource (TAIR),
<https://www.arabidopsis.org>

A screenshot of the TAIR website showing the Ribo-seq data for locus **AT4G34590**. The main panel displays a map of the genome with a yellow bar representing the protein-coding genes. A red circle highlights a specific region on the map. The left sidebar includes links for Summary, Transcripts, Maps and Mapping Data (which is selected), Sequences, Protein Data, Expression, Gene Ontology, and Homology.

77

Our latest Ribo-seq data on TAIR



79

Upstream ORF Annotated ORF 80

Our latest Ribo-seq data on TAIR

<https://www.arabidopsis.org>

SAMDC1: AT3G02470

Any questions?

1st half

Experiment

- 1. Introduction to Ribo-seq
- 2. Getting ribosome footprints
- 3. Making the sequencing libraries

(Q&A + 5 min break)

2nd half

Data analysis

- 4. Preprocessing & quality control
- 5. Translation efficiency
- 6. ORF discovery
- 7. Ribo-seq data visualization

Survey link: <https://shorturl.at/KbtYQ>