
The Development of a Ribosome Profiling Protocol to study
Translation in the Yeast *Yarrowia lipolytica*



UCC

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland



LUKE POWER

119225992

SUPERVISORS:

PROFESSOR PAVEL BARANOV

DOCTOR JOHN MORRISSEY

MRES MOLECULAR BIOLOGY & BIOCHEMISTRY

DATE: 31/12/20

SCHOOL OF BIOCHEMISTRY

SCHOOL OF MICROBIOLOGY

Contents

Contents	1
Table of Figures	4
Main Figures	4
Supplementary Figures	4
Abbreviations	4
About CHASSY	Error! Bookmark not defined.
Abstract	5
Introduction	5
Results	8
Development of a ribosome profiling protocol in <i>Y. lipolytica</i> W29	8
Utilizing polysome gradients to show digestion	11
Using Trips-Viz for Data Analysis	13
Triplet Periodicity	14
Translation Expression Values	17
Illustrative profiles of footprint densities at individual genes	18
An example of uORF mediated regulation of translation	20
Reproducibility	21
Discussion	22
Method	23
Footprint preparation	23
Culturing	23
Cell harvesting	23
Cell lysis	24
Lysis Buffer Preparation	24
Mixer Mill Cryogenic Lysis	24
	1

Lysate Clarification	25
Qubit Readings	25
RNase 1 Treatment	25
Polysome Buffer	25
Preparation of Density Gradients	26
Ultra-centrifugation	27
UV Fractionation and Monosome Collection	27
RNA Isolation	28
Selecting Wells	28
Trizol	28
Library Generation	29
Size selection from gel	29
Gel extraction	30
PNK dephosphorylation and Linker Ligation	30
Enzymatic pre-adenylation of linker using Mth RNA ligase	30
Dephosphorylation	31
Linker ligation	32
Gel Extraction	32
Reverse transcription	33
RT Polyacrylamide Gel	34
Gel extraction	34
Circularisation	35
PCR amplification	35
PCR Cycle	37
PCR Gel	37
Gel extraction	37
	2

Bioinformatic analysis and processing	38
Initial processing of data	38
Utilizing Trips-Viz	38
Stock Solutions and Buffers	39
YPD Media	39
Lysis Buffer	39
Polysome Buffer	39
Polysome Gradient Buffers	40
RNA Elution Buffer	40
15% PAGE Gel (with Urea)	40
7.5% PAGE Gel (with Urea)	41
8% polyacrylamide gel for PCR	41
Supplementary Figures	42
Cell culture and harvesting	42
Footprint fragment isolation and purification	43
Footprint fragment end repair and ligation linking	44
Reverse Transcription of ligated RPFs	45
cDNA library generation	46
References	47

Table of Figures

Main Figures

1. A diagram detailing the general steps involved in Ribosome profiling.	Page 9
2. Polysome Profiles of <i>Y. lipolytica</i>	Pages 11/12
3. Metagene Profiles and Triplet Periodicity	Pages 15/16
4. Translation Expression Distribution	Page 18
5. Single Transcript plots	Page 20
6. uORF mediated regulation of Translation	Page 21
7. Pearson Coefficient	Page 22

Supplementary Figures

1. Growth curve of <i>Y. lipolytica</i>	Page 44
2. Size selection gel of <i>Y. lipolytica</i>	Page 45
3. Linker ligation gel of <i>Y. lipolytica</i>	Page 46
4. Reverse transcription gel of <i>Y. lipolytica</i>	Page 47
5. PCR gel of <i>Y. lipolytica</i>	Page 48

Abbreviations

Abbreviation	Full Title
Ribo-Seq	Ribosome Profiling
cDNA	Complementary DNA
RPF	Ribosome Protected Fragments
PCR	Polymerase Chain Reaction
mRNA	Messenger RNA
PNK	Polynucleotide Kinase
YPD	Yeast extract, Peptone, Dextrose
CHX	Cycloheximide
CDS	Coding Sequence
ORF	Open Reading Frame
uORF	Upstream Open Reading Frame

Abstract

Ribosome Profiling is an emerging technique that provides a genome wide-view of translation *in-vivo*, allowing an in-depth analysis of translation within a cell. It offers significant advantages over other techniques such as RNA-sequencing for interpreting transcriptional data about cells. Ribosome profiling can accelerate our understanding of complex biological processes happening within the cell and in turn, can be utilized to explore new ways for industrial exploitation of various cells. Here ribosome profiling, with the aid of a specialized ribo-seq data browser have been used to uncover new information about translation of the *Yarrowia lipolytica* transcriptome, an unusual yeast with interesting properties, with the aims of promoting its use as a candidate for biotechnology.

Introduction

Y. lipolytica is a yeast species of growing interest in the biopharmaceutical industry due to its advantageous properties over conventional yeast like *Saccharomyces cerevisiae*. *Y. lipolytica* is an oleaginous yeast species with the ability to utilize unusual carbon sources such as hydrocarbons including alkanes and paraffins (Abdel-Mawgoud et al., 2018, Xu et al., 2016, Liu et al., 2015). In addition, it possesses the ability to accumulate high abundances of intracellular lipids which could potentially be geared towards biochemical products which use these as precursors (Egermeier et al., 2017, Magnan et al., 2016). It has seen some success in industry in the production of citric acid, a common chemical used as a preserving agent for cosmetics (Miller and Alper, 2019, Larroude et al., 2018a, Larroude et al., 2018b), amongst other compounds of high value such as omega fatty acids or conjugated linoleic acids (Qiao et al., 2017, Zhang et al., 2013, Xue et al., 2013). At present, there is a lack of insight with regards to gene expression of *Y. lipolytica* that is publicly available. Studies have been done using RNA-seq and mass spectrometry (Zhang et al., 2016, Wang et al., 2018) to enhance our knowledge of translation in *Y. lipolytica*, however they do not offer the same wealth of information on gene expression as ribosome profiling does. While information regarding transcription of genes in *Y. lipolytica* is widely available, mRNA abundance does not always accurately predict protein synthesis levels, which is a weakness of that technology (Brar and Weissman, 2015, Ingolia, 2016). To overcome this weakness, ribosome profiling was used to accurately and reliably provide an insight into translation in *Y. lipolytica*, with the aim of establishing a pipeline to explore post-transcriptional control and to help bridge the gap of knowledge with regards to the transcriptome of *Y. lipolytica*.

Ribosome profiling is an emerging technology that has become has risen in popularity in the field of translational biology, due to the insight it can provide into translational control of gene expression and regulation (Ingolia et al., 2009). It works off the concept of flash-freezing the cells with liquid nitrogen to stall translation, the addition of lysis buffers containing translation inhibiting drugs such as lactimidomycin or cycloheximide (CHX) (the addition of the drugs in the lysis buffer is the prevent any run-off of ribosomes) (Schneider-Poetsch et al., 2010, Duncan and Mata, 2017), some form of detergent (such as triton-X100) to disrupt cell membranes and the use of a physical homogeniser which allows for the cells contents to leave the cell, with the subsequent use of ribonucleases to cleave and destroy regions of mRNA that reside outside of the ribosome, leaving only ~28-30nt fragments protected by ribosome (Ribosome Protected Fragments/RPFs or footprints)(McGlinicy and Ingolia, 2017, Brar and Weissman, 2015). These RPFs undergo a series of isolation steps and reactions to generate a cDNA library, which is then sequenced and mapped to provide the location of all translating ribosomes across the transcriptome that were active prior to flash freezing. Added with the known sequence of the transcriptome of the organism of interest, ribosome profiling can be used to show one; where ribosomes are actively translating, and two; how external stress factors such as temperature, pH, oxidative stress, and drug treatment can impact the synthesis of proteins. A change in temperature can cause for an upregulation of proteins such as the family of heat shock proteins, which act as chaperones, assisting in the stabilization of proteins through assisting folding or unfolding of protein structures. A change in nutrient sources available can show how the cell reacts to adapt to this by altering its regulation of different enzymes. Treating cells with particular drugs that inhibit growth aspects or that regulate the cells defence mechanisms can help visualise the internal response the cell takes against the drug of question (Ingolia et al., 2012, Michel and Baranov, 2013). At CHASSY, the scientific end goals are to enhance production of compounds of interest, and to make cells more robust for various industrial conditions. Using these insights generated from ribo-seq could enhance acceleration of drug discovery and/or biotechnological advancements with the aim of unlocking and upgrading *Y. lipolytica* W29 potential as a cell factory for the biotechnology sector.

Here we have generated the first set of ribosome profiling data for *Y. lipolytica* W29. The ribosome profiling data was mapped to the reference transcriptome of *Y. lipolytica* W29 and is available through the transcriptome browser Trips-Viz (TRanscriptome wide Information on Protein Synthesis Visualized) (available here: <http://trips.ucc.ie/>). Trips-Viz is a computational environment which allows for an interactive graphical exploration of data generated from ribosome profiling, RNA-seq and mass spectrometry (Kiniry et al., 2019). Trips-Viz was developed by Stephen Kiniry and his colleagues in

University College Cork. It is a powerful tool that is capable of visualizing ribosome profiling and RNA-seq data at the level of a single transcript profile up to looking at the entirety of the transcriptome. Trips-Viz is able to look at data past individual single gene transcripts and allows for the analysis of metadata such as triplet periodicity, read length distributions of RPF's, and different breakdowns of reads. Trips-Viz ability to visualize mapped reads to the transcriptome gives useful information such as relieving the difficulty of mapping short reads over spliced junctions (a regulated process that allows for a gene to encode for multiple proteins by the use of different exons in transcribed mRNA). Trips-Viz allows for the visualization of multiple datasets on a singular graph, enabling for the comparison of multiple species variants of a single gene drug treatment and its effects on gene expression.

Results

Development of a ribosome profiling protocol in *Y. lipolytica* W29

For this study, we decided to use the *Y. lipolytica* strain W29, which has been used to develop many variants of *Y. lipolytica* including the strain E150/CLIB122, a publicly available sequenced and annotated genome whose genomic sequence currently acts as the genomic reference sequence (Liu and Alper, 2014). We chose the strain W29 as it a fully sequenced genome with all chromosomes assembled. It also contains the highest number of protein coding genes which is essential for transcriptome analysis (W29 contains a total of 7949 protein coding genes, while the strain E150/CLIB122 contains 6449 protein coding genes.) (Magnan et al., 2016, Zhang et al., 2019, Ledesma-Amaro and Nicaud, 2016). Figure 1 shows a diagram adopted from (McGlinchy and Ingolia, 2017) which shows the outline of the procedure followed in this experiment. Ribosome profiling was carried out in W29 cells harvested during log-phase growth in rich YPD media at 25°C. A total of four samples were made, two biological replicates BR 1 and BR 2, with a technical replicate made for each (BR1-1, BR1-2, BR2-1 and BR2-2). The addition of two technical replicates was done to ensure that there would be enough material present for sequencing as they would be pooled prior to reverse transcription. Cells were harvested through the use of a 1L Durapore glass filter assembly with a nitrocellulose membrane filter. Yeast is scraped from the filter into a 50ml Falcon tube containing liquid nitrogen to flash freeze the cells. 2mLs polysome lysis buffer which included CHX and triton-X100 detergent was added to the falcon tube drop-wise. The frozen cell/lysis buffer mixture were lysed cryogenically using steel grinding balls and jars in a mixer mill which are shaken at extremely fast speeds while keeping the jars and samples well below freezing temperatures. This lysate was then clarified using centrifugation and RNA was quantified using a Qubit fluorometer. The lysate was divided into 30µg aliquots to make a pool of samples used for ribosome digestion. Samples were treated with RNase 1 to digest mRNA outside the ribosome, leaving a ribosome protected fragment (RPF).

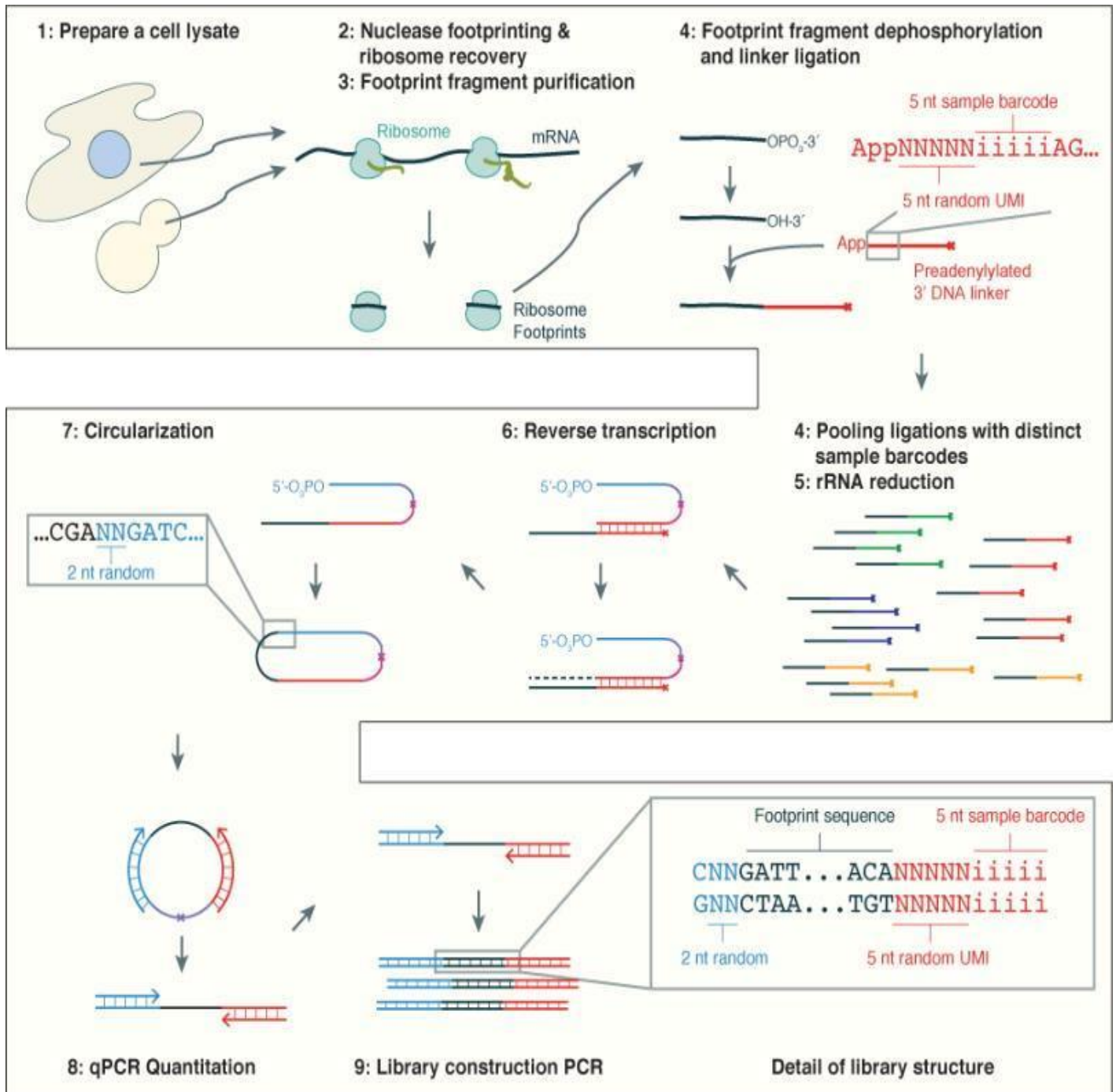
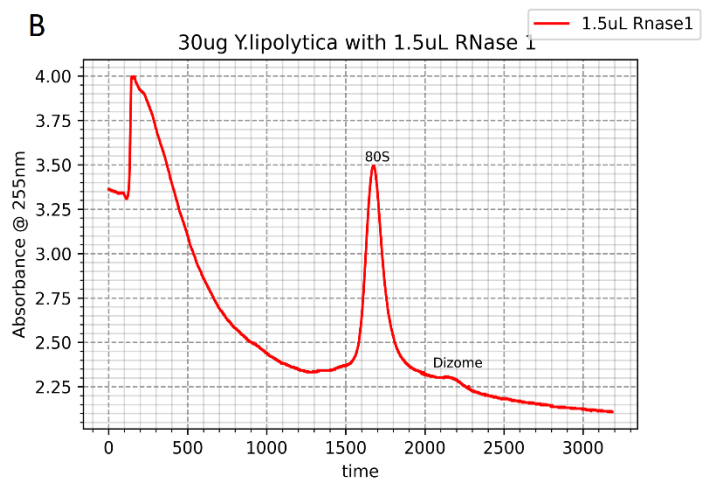
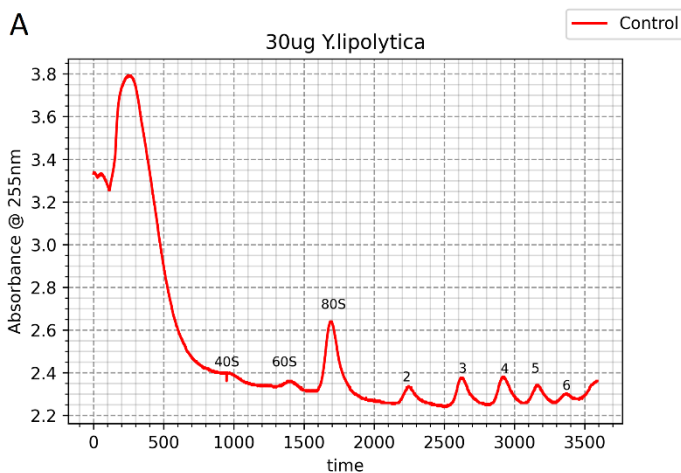


Figure 1. A diagram detailing the general steps involved in Ribosome profiling. Adapted from (McGlinchy and Ingolia, 2017).

RNAse 1 was selected to be the ribonuclease as it is used routinely for *S. cerevisiae* in ribo-seq experiments. The ribosomes present in yeast are generally quite resilient to harsh digestion and they do not suffer any loss to structural integrity. This allows for a more robust digestion and as a result lower amount of contaminants being present (Gerashchenko and Gladyshev, 2017). Samples are subjected to RNAse 1 digestion for 45 minutes at room temperature, 200 rpm before being purified using a sucrose gradient to size-select monosomes. After recovery from the sucrose gradients, the now isolated monosomes are denatured to release the RPFs and run through a 15% PAGE Urea size selection gel, alongside 26 and 34 nucleotide bands which act as markers. Bands present between these markers contain the ribosomal footprints and are excised from the gel and eluted overnight (see supplementary figure S2). These recovered RPFs are treated with T4 Polynucleotide Kinase (T4 PNK) to repair the damage caused by RNAse1 cleavage. As RNAse 1 cleaves at the 3' side of nucleotide bases through 2', 3'-cyclic monophosphate intermediates, T4 PNK works through the removal of 3' phosphoryl groups and prepares the RPF's for subsequent linker ligation. Footprints are treated with pre-adenylated linkers and T4 Rnl2(tr) K227Q to add unique molecular identifiers (UMIs) to allow for multiplexing before sequencing. Samples were ran through another 15% PAGE-Urea gel and the ligated products were excised from gel and eluted overnight (see supplementary figure S3). Upon elution of the samples, the technical replicates were pooled, leaving two biological replicates (BR1 and BR2). These two samples undergo reverse transcription with the addition of a reverse transcription primer, generating a cDNA template of the footprints. The cDNA was size-selected and eluted from a 7.5% PAGE gel overnight. Upon recovery of the templates, Circligase II was employed to circularise the single-stranded cDNA template, prior to PCR amplification. PCR was used to generate a library which was sent for high-throughput sequencing. An important factor to consider when doing PCR is the number of cycles to use for amplification, as the higher the cycle number the more cDNA amplified, but there is also a higher risk of having PCR duplicates which are undesirable as they inflate homozygosity falsely. A trial PCR gel was ran using a variable numbers of cycles in an effort to determine the best number of cycles of PCR amplification to do (see supplementary figure S5). Seven cycles was selected sufficient to generate a cDNA library large enough to send for sequencing. Post-sequencing and analysis, the location of 77 million footprints were mapped to the coding regions of the transcriptome.

Utilizing polysome gradients to show digestion

The principle of monosome-polysome separation is utilising a sucrose gradient made up of a linear sucrose gradient (from 10% to 50% sucrose in this case) to act as a separation buffer for the cell lysate to be separated according to size). The cell lysate is added to the sucrose gradient (by placing it on-top of the sucrose gradient and subsequently ultra-centrifuged at an extremely high speed (35000rpm for 3 hours at 4°C). This allows for the separation of the components of the lysate based on their size. After RNase 1 digestion, the contents of the cell are cleaved into small fragments that are low in mass. This material (such as tRNAs and non-translated mRNA) is collected near the top of the gradient where the sucrose concentration is lighter, while larger components (such as the monosomes) are centrifuged down through the gradient. As the gradient passed through the absorbance detector, the solution was aliquoted onto a 96 well plate which was later analysed with a plate reader to determine the wells(s) which contained the monosome fraction. These monosome fractions were purified using TRIzol extraction (refer to methods). During the gradient collection process, a csv file was generated of the flow through by measuring the absorbance of light at 225nm every 100ms (millisecond). The use of sucrose gradients allows for the visualisation of ribonuclease digestion by comparing a graph in which samples that did not undergo digestion (negative control) against a graph in which samples had RNase1 digestion.



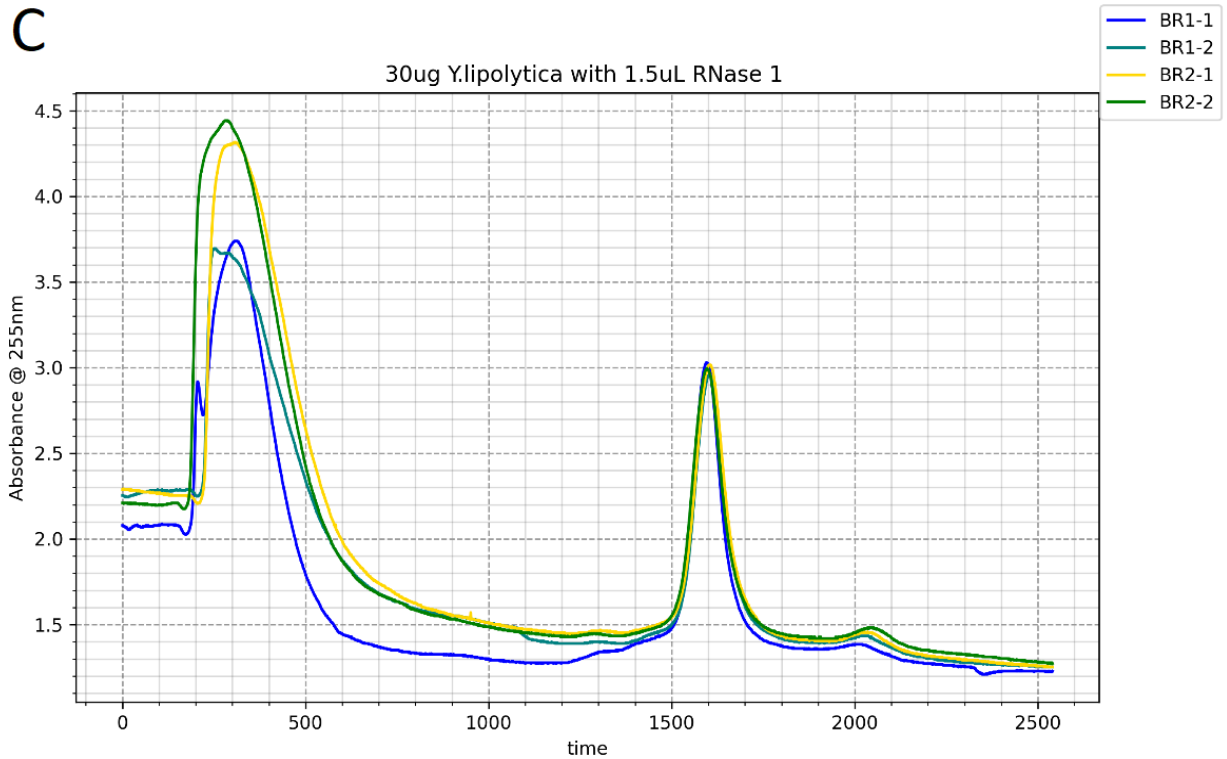


Figure 2A shows an individual polysome profile graph of *Y. lipolytica* W29 lysate that did not undergo RNase 1 digestion (negative control). This was done through following the method outlined in this paper but omitting the addition of the ribonuclease. The polysome gradient in Figure 2A shows the locations of the peaks corresponding to different ribosome complexes which are indicated. Upon RNase I digestion, the mRNA between individual ribosomes within these polysome complexes are digested. The polysomes are therefore destroyed and migrate as 80S monosomes as evident in the polysome profile shown in Figure 2B. Figure 3A is a combined polysome profile graph of the 4 sets of *Y. lipolytica* W29 lysates that were used to sequence the transcriptome. The four samples can be differentiated by the colour key given in the top right corner of the graph. Figure 3A confirms a robust digestion as each samples 80S peak are uniform in terms of both height and time taken to elute.

Using Trips-Viz for Data Analysis

After the ribo-seq library was sequenced, the data received contained reads from both biological replicates as one file. The file was split into two individual replicates using Cutadapt which recognises barcodes for each biological replicate and allowed the separation of reads. Processing this data involves the use of demultiplexing the linker sequences so that the FASTQ file can be split into two separate files, for each biological replicate. The separation allowed for individual quality control to be done on each replicate using FastQC, a tool that can provide basic quality metrics such as GC content and sequencing quality (available here: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). To generate the files required for Trips-Viz, Cutadapt (Martin, 2011) was used to remove adapters present on the reads, and bowtie (Langmead et al., 2009) was used to align these reads to the reference genome. Both of these tools are freely available and are not specific to ribosome profiling. Bowtie was used over its successor bowtie 2 as ribo-seq generates shorter reads. Bowtie 2 is more suited to longer gapped alignments while bowtie is better for reads around 25 nucleotides in length (which is closer to the desired length of footprints). Following this, Samtools was used to convert the alignment SAM file (Sequence Alignment Map) file into a sorted SAM file and subsequently an indexed BAM (Binary Alignment Map) file. HT-SEQ counts the number of RPF's aligned to each gene using a protein-coding gene annotation from the published W29 GTF (Gene Transfer Format) file (Anders et al., 2015, Magnan et al., 2016).

One issue that arose during analysis of transcriptome information was a lack of 5' leader and 3' trailer regions, as no transcriptomic methods were used to generate this data. Therefore, the sequences available for transcripts included only CDS (coding sequence) information from start to stop codon for each mRNA. However, we computationally solved this problem using a custom python programming script. We developed code which extended coding regions 300 nucleotides upstream and downstream to represent 5' leaders and 3' trailers respectively. This created a pseudo-transcriptome which could reveal useful information of translation in 5' leaders such as uORFs or 3' trailers such as stop codon readthrough. By adding this extension, it allowed for the detection of uORFs that would have otherwise been missed. Figure 4 shows a single transcript plot of *GCN4*, a well-known example of why this pseudo-transcript file was needed. From the figure it can be seen that there are 3 uORFs that would have been missed had the pseudo-transcript fasta file been omitted.

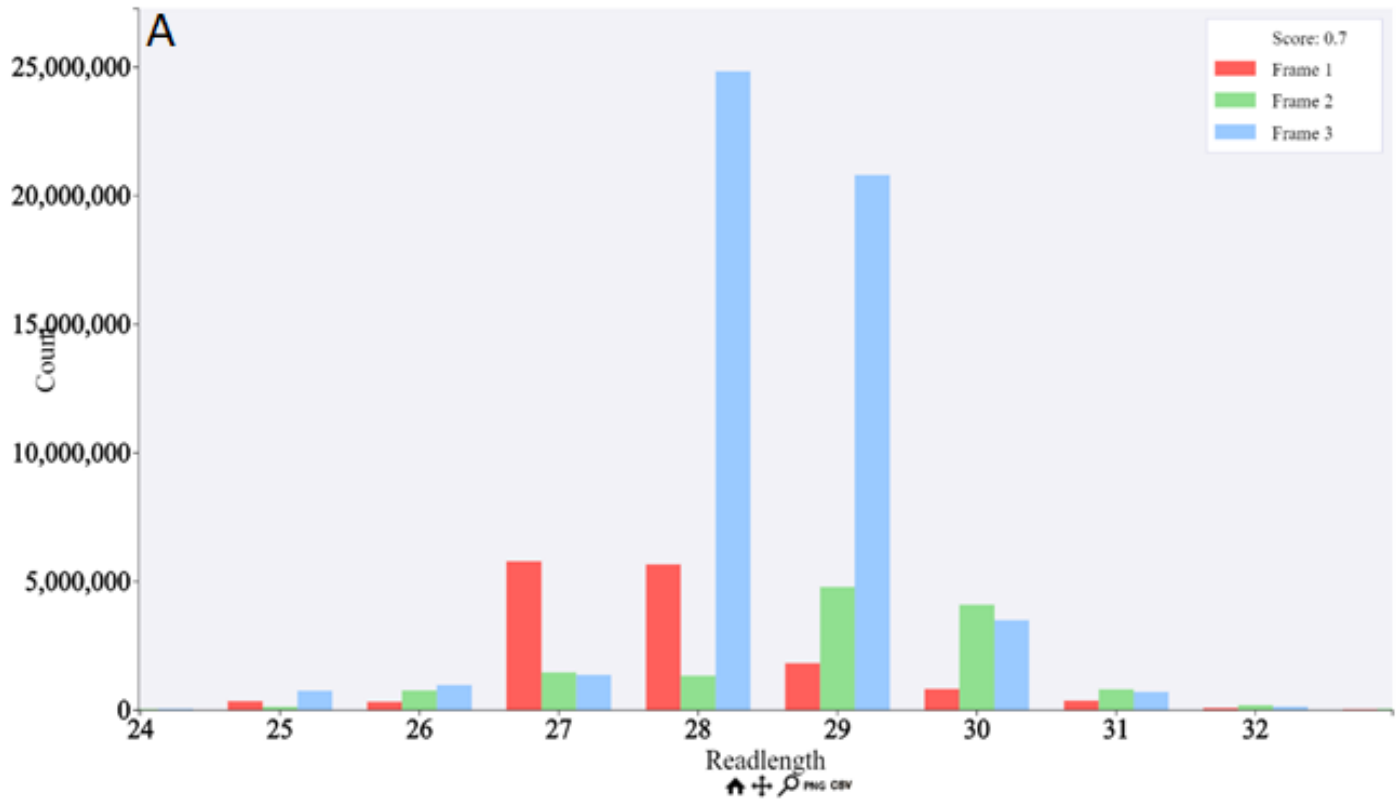
Another issue that arose during analysis, was the lack of names available to identify genes such as *GCN4*, as they appeared only as locus tags rather than named genes. By using Blast to find homologs between *S. cerevisiae* (reference genome S288c taken from SGD) and *Y. lipolytica*, we were able to add

gene names to ~44% of the proteome. When filtering, we used an E-value of 0.01, a bit score of 50 and removed any homolog with less than 30% identity.

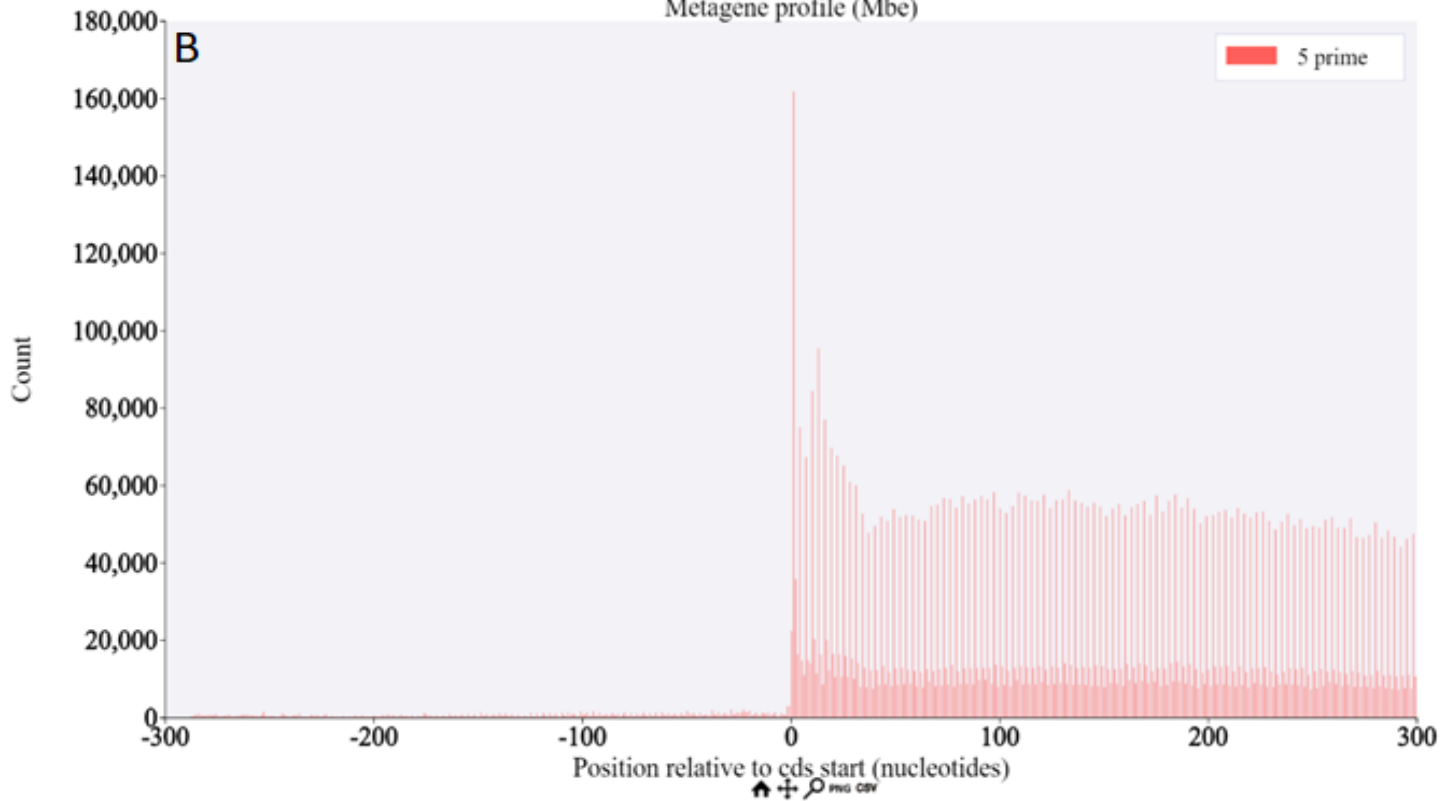
Triplet Periodicity

Trips-Viz allows for the visualisation of triplet periodicity, a periodic pattern that has a phase of three, in both metagene profiles and a separate compiled read breakdown chart. Because the ribosome decodes mRNA one codon at a time and moves by a three-nucleotide step along the mRNA chain as it decodes, the precise location of the ribosome is dependent on its reading frame. As a result of this, the probability of ribonuclease cleavage is not the same for different sub codon positions. This leads to the periodic pattern in footprint distributions of a repeat every three nucleotides. As such, triplet periodicity can be used as a measure of the quality of the ribosome profiling data. When the signal is strong and the coverage is high, it can be possible to detect what are known as dual coding regions, regions of the translated sequence that are present in more than one frame. This can be in either the same mRNA or in an alternatively spliced transcript (Michel et al., 2012). Figure 3A is a compiled triplet periodicity plot using the data sets from both biological replicates. From the key present in the top right corner of the graph, it can be seen that the frame 3 profile (colour coded blue) is the highest expressed frame for the read lengths of 28 nucleotides and 29 nucleotides. The score for this breakdown is 0.7 which indicates a strong bias towards the third frame in particular. Generally, a score of 0.6 or above is considered a strong score for the data set whereas lower score values of 0.3 and 0.4 indicate weak periodicity. The graph itself would not show a singular peak higher than the other frames, but rather a more closely aligned series of peaks. Figures 3B and 3C are metagene profiles located around different regions of interest. From these two graphs the triplet periodicity of the ribo-seq data can be visualised. From it, a clear triplet pattern can be observed in which there is one large peak followed by two small peaks. This trend is caused due to clear and sufficient RNase 1 digestion (in which there is a low amount of variance between digestions).

Triplet periodicity (MfK)



Metagene profile (Mbe)



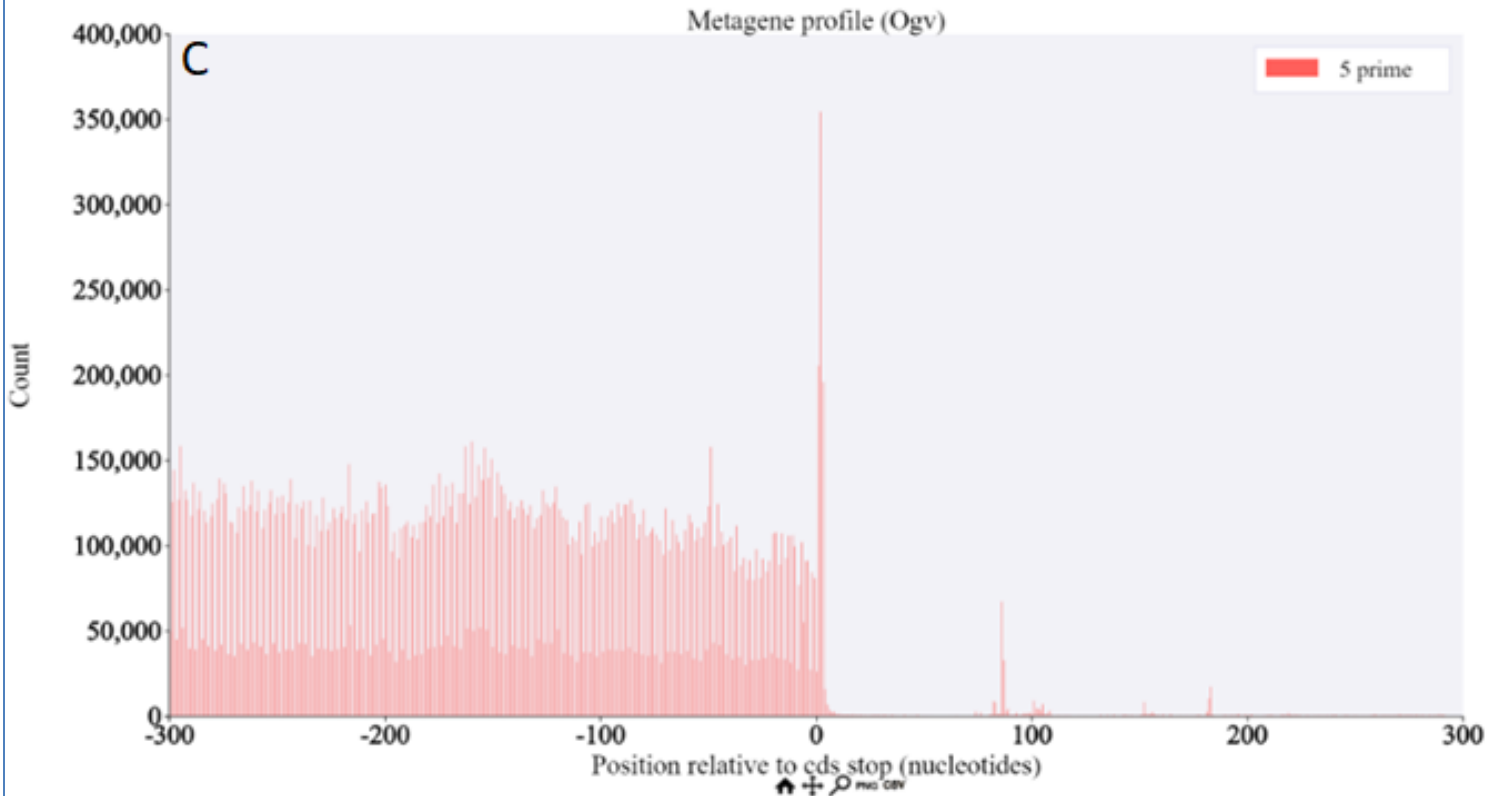


Figure 3A is a compiled triplet periodicity/reads breakdown plot, detailing a combined read count for the two biological replicates used. In figure 3A it can be seen that there is a prominent peak in the third frame for both read lengths 28 and 29. Figures 3B and 3C are metagene profiles based around specific positions of interest (Figure 3B is the position relative to the CDS start codon where $x=0$, while figure 3C is the position relative to the CDS stop codon where $x=0$) From both figures, a clear triplet-like pattern of a large peak being trailed by two smaller peaks is present. This pattern is more prevalent in figure 3B which is the CDS start position.

Translation Expression Values

One feature that Trips-Viz provides is a gene count table that indicates the number of reads aligned to the sequences of all protein-coding transcripts. This table can be downloaded as a csv file, allowing for further analysis such as distribution of gene expression or differential expression analysis. Here, we were interested in the distribution of gene expression (i.e. ribosomes per gene). Figure 4 is a translation expression distribution histogram, with three examples highlighted, *JEN1*, *GCN4* and *EF1A*. *JEN1* is a monocarboxylate symporter of the plasma membrane involved in the uptake of carbon sources like lactate and pyruvate. The expression of *JEN1* in this case was low, likely due to cells being grown in YPD (Yeast Peptone Dextrose) media which contains 2% glucose as a carbon source. The second highlighted gene is *GCN4*, a gene responsible for the activation of a large amount of amino acid biosynthetic genes. Lastly, *EF1A* is highlighted on the far end of the histogram, and is among the most highly expressed genes. If *Y. lipolytica* was grown in a different set of conditions, the locations of the genes highlighted would differ, due to the shift in regulation of various genes.

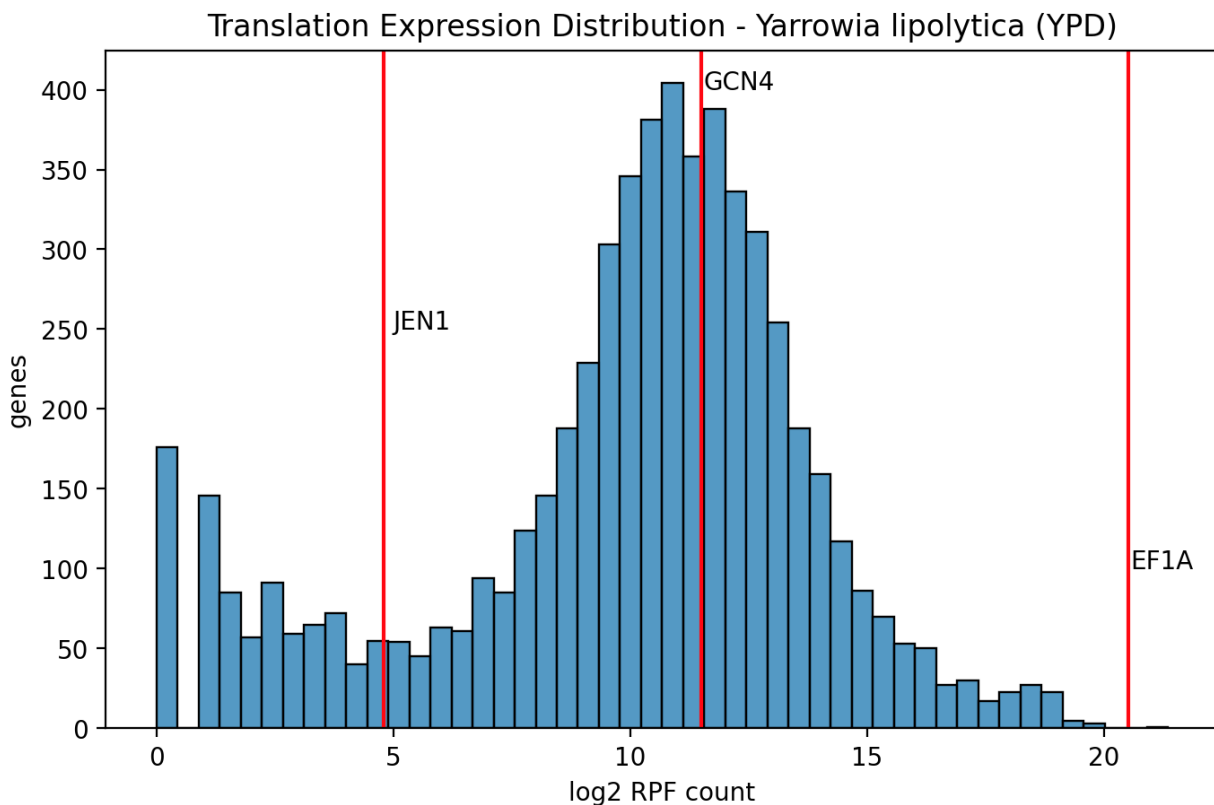
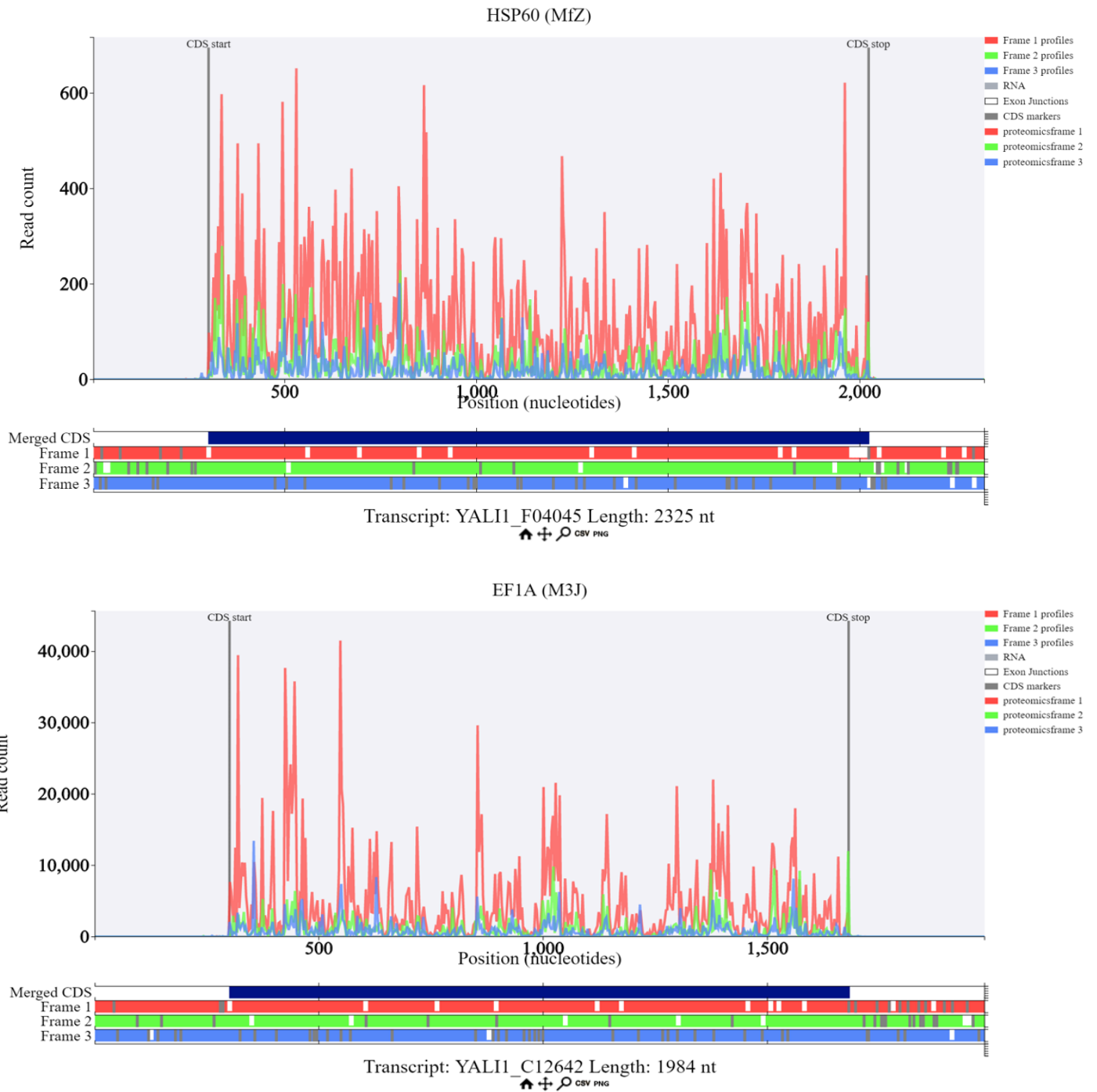


Figure 4 is a histogram detailing the distribution of expression of genes in *Y. lipolytica* W29. The location of three genes are highlighted as examples.

Illustrative profiles of footprint densities at individual genes

TRIPS-Viz provide visualisations of footprints distributions across individual transcripts, offering a wealth of information relating to individual genes. The purpose of these plots is to allow the user to see ribosome profiling and RNA-seq reads that align to an individual transcript. The reads are colour coded depending on what reading frame is supported. If the triplet periodicity of the reads from the graph is strong, peaks of one colour should be higher than the others, giving a clear picture of which frame is being translated. The graph is split into two parts, a main window containing a plot and a second section below known as the ORF Architecture. The ORF architecture contains the same three colour coded frames, while also showing AUG codons (short white lines appearing in bars) and stop codons (longer grey lines), yielding visualisation and manual identification of ORFs. Figure 5A is an example of an individual transcript plot, looking at the transcript for Heat Shock Protein 60 (*HSP60*), a mitochondrial chaperonin which assists in the folding of amino acid chains into their 3D structure (Grundtman et al., 2011). It can be seen that the majority of reads in the plot in Figure 5A support translation of annotated reading frame (read). It also can be seen that nearly all reads are mapped within the annotated CDS region (indicated with long vertical lines). Figure 5B shows a single transcript plot of Elongation Factor 1 alpha (*EF1A*). *EF1A* plays a large role in the enzymatic delivery of aminoacyl tRNAs to the ribosome, by encoding for the isoform of the alpha subunit of the EF1 complex (Talapatra et al., 2002). From figure 5B, it can be seen that there is a strong bias for the gene to be initiated from a frame 1 profile. As *EF1A* is involved in elongation and translation it is expected for a high number of reads from ribosome profiling to be attached to *EF1A*. In the case of this experiment, just under a million reads were aligned to *EF1A* alone. The graph shows that there is no initiation occurring outside of the main CDS region (in either the 5' leader or 3' trailer). As *EF1A* is a transcription factor it is not unusual to see a high number of reads attached to the gene. When looking at the ORF architecture, there are a multitude of start codons present at the start of the CDS region in frame one while stop codons appear intermittently throughout the frame one profile. From the figure it can be seen that there was no initiation of translation occurring in either the 5' leader or 3' trailer.



Figures 5A and 5B both show individual transcript plots for two genes (figure 5A being the transcript plot for *HSP60*, while figure 5B being the transcript plot for *EF1A*). Both transcripts show a prominent frame one profile for expression which can be seen in from the main graph as well as the ORF architecture below.

An example of uORF mediated regulation of translation

General Control Nonderepressible 4 (*GCN4*) is a transcription factor involved in the activation of many amino acid biosynthetic genes and is often called a master regulator of gene expression. It regulates close to 10% of the gene expression of the *S. cerevisiae* genome. *GCN4*'s expression is tightly regulated at both transcriptional and translational levels (Hinnebusch and Natarajan, 2002) and responds to amino acid starvation. Figure 6 illustrates the single transcript plot of *GCN4*. Unlike the previous figure (figure 5A and 5B) *GCN4* contains three uORFs that are present in the 5' untranslated region which regulate the translation of this gene. There is a fourth uORF that is not visible in figure 6. This is due to RNA-seq not being done in parallel with ribo-seq and the pseudo-transcriptome fasta file used to generate each plot only taking 300 nucleotides before the CDS start and 300 nucleotides after CDS stop codon. Extending this region of nucleotides could resolve this issue. It can be seen in figure 6 that most of the translation occurs in the second uORF which is in the 3rd reading frame (blue). Under normal conditions it is expected to see the translation of *GCN4* being regulated from outside the CDS region, while under stress conditions the translation of *GCN4* would be altered with peaks of initiation appearing inside the main CDS region. However, these results suggest *GCN4* contains some amount of basal translation but likely to increase massively under amino acid starvation.

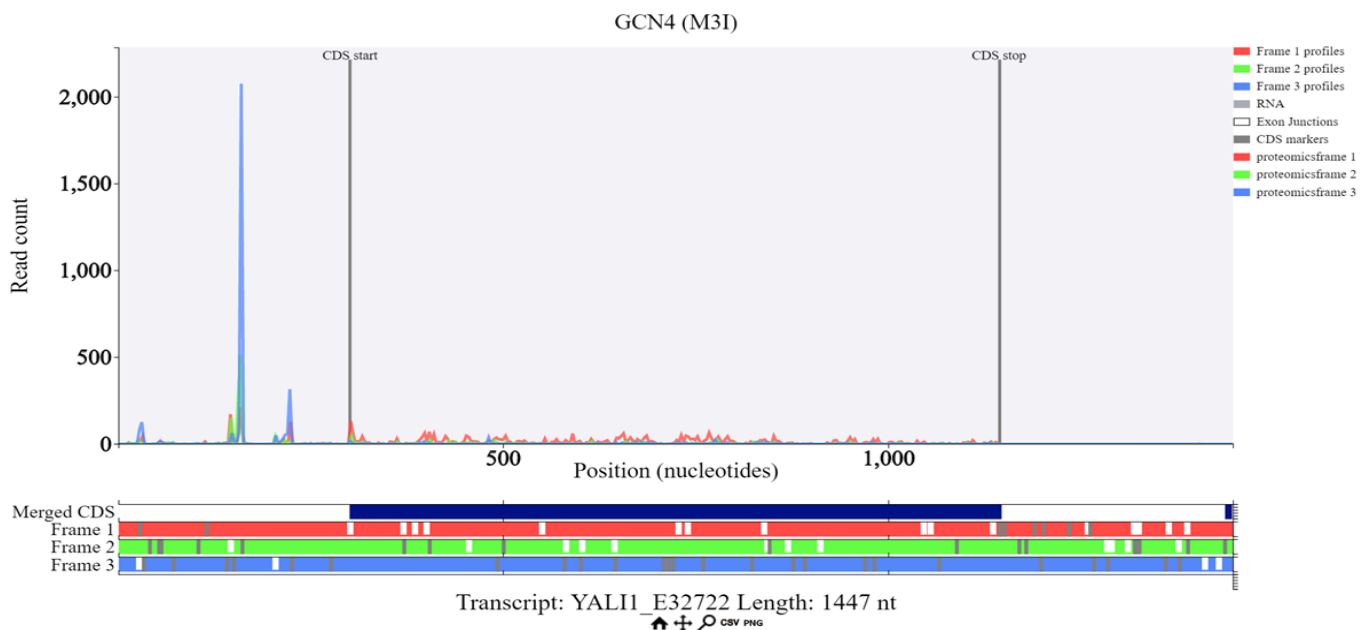


Figure 6 shows a single transcript plot of *GCN4*. Only three of the four uORFs that regulate the translation *GCN4* are present. It can be seen that there is a predominant 3rd frame profile for expression of *GCN4*.

Reproducibility

To measure the levels of reproducibility, a correlation is measured between the number of reads aligned to each gene of the two biological replicates. The Pearson correlation coefficient is 1.0, indicating a perfect correlation and can be seen in the scatter plot shown in figure 7.

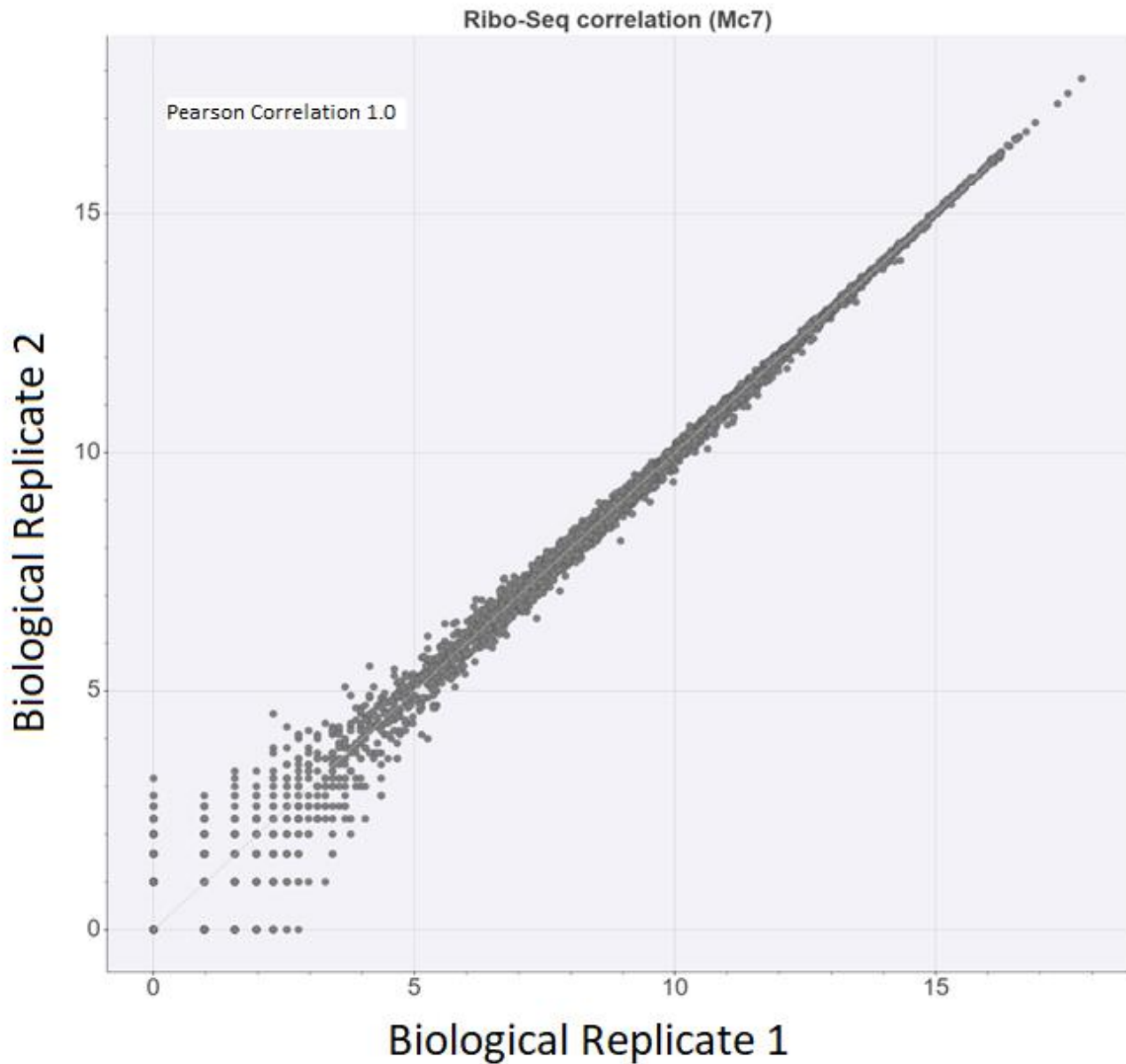


Figure 7: The scatter plot showing correlation between the number of RPFs aligned to each gene between two replicates.

Discussion

Here we generated and analysed the first ribosome profiling data for *Y. lipolytica* W29, which was generated using a custom method derived from the McGlincy and Ingolia protocol for *Saccharomyces cerevisiae* in 2017 (McGlincy and Ingolia, 2017). We determined the position of 77 million footprints, giving a genome-wide view of translation of in the cell. Figures 3A, B and C show two metrics which confirmed that the ribo-seq in *Y. lipolytica* worked, with Figure 3A showing that the quality of ribo-seq data generated was good with a score of 0.7. Figures 3B and 3C further showcase a lack of reads in the 3' trailers and 5' leaders confirming the ribosome profiling worked in *Y. lipolytica*. Figure 7 provides shows that the method is reproducible with a perfect correlation between the two biological replicates.

The main aim of this study was to establish a working protocol and computational analysis pipeline for future work to be done on *Y. lipolytica*. We did not examine *Y. lipolytica* under any stress conditions as this was a pilot experiment to establish ribosome profiling in *Y. lipolytica*. Future work could focus on utilising ribosome profiling to investigate external stresses such as high temperatures or the addition of alkanes as substrates for *Y. lipolytica* to further investigate some of its more unusual properties to help advance the knowledge available for *Y. lipolytica*. By doing ribosome profiling in cultures grown in rich YPD media, and harvested during log phase, we ensured no external stress conditions would have taken place. We showed reproducible results using our customized method which with the addition of polysome gradients, allows us to determine if nuclease digestion worked or not from an early stage. This addition shows that it is possible for this method to adopted and adjusted for different species such as *Kluyveromyces marxianus*. Figures 2B and 2C confirm that the ribosomes present in *Y. lipolytica* do not suffer much loss in structural integrity from RNase1 as the 80S peaks present in both figures are strong clear peaks that are both sharp and uniform across multiple samples.

The addition of the *Y. lipolytica* transcriptome to Trips-Viz will allow full ribosome profiling computational analysis for this species. Trips-viz will be a valuable resource for researchers to upload ribosome profiling data for quality control, differential expression and investigation of translational regulation within the cell. Both global and single gene analysis will be critical for future ribosome studies with *Y. lipolytica* to understand how this species adapts to changes in environmental conditions. These may include industrial conditions such as high temperatures in industrial bioreactors and *Y. lipolytica*'s utilization of unusual carbon sources. This protocol and computational analysis software will allow researchers use ribosome profiling as a platform to study *Y. lipolytica* and develop more robust cell factory strains suitable for industrial use.

Methods

Footprint preparation

Culturing

YPD (yeast extract peptone dextrose) media was made up and used to culture the W29 cells.

Liquid cultures were made up as followed:

Ingredient	1000ml/1L	750ml
1% Yeast Extract	10g	7.5g
2% Peptone	20g	15g
2% Glucose	20g	15g

Agar plates were made up as followed:

Ingredient	15ml/1 plate	150ml/10 plates
1% Yeast Extract	0.15g	1.5g
2% Peptone	0.3g	3.0g
2% Glucose	0.3g	3.0g
2% Agar	0.3g	3.0g

Medium was then made up using MilliQ water to the required amount and autoclaved after preparation.

Table 1: YPD agar plates and YPD liquid culture recipes.

Cell harvesting

1. Pour agar plates under Bunsen burner flame to ensure an aseptic environment.
2. Once plates have cooled, streak plates using a pipette tip, ensuring that no streaks overlap.
3. Place the plates in a dry environment at 25°C.
4. Leave the plates for four days to grow.
5. Autoclave two volumetric flasks and a graduated cylinder.
6. Transfer a single colony into a 50ml Falcon tube filled with 5ml of YPD liquid culture.
7. Leave in an incubator at 25°C overnight at an RPM of 200.
8. Measure cells OD600 using a 1 in 10 dilution of sample from overnight culture into YPD media.
9. Using $C1V1 = C2V2$ calculate out the required amount of overnight sample to add to a 150ml flask of YPD media to give an OD600 reading of 0.06.
10. Incubate cells to an OD600 of 0.6 minimum.
11. Prepare a 50mL falcon tube/rack and fill it with liquid nitrogen, place the rack on a cardboard or Styrofoam box.
12. Pierce the cap with a screwdriver to make a single hole.

13. Rapidly filter and harvest the yeast with the 1L Durapore glass filter assembly with GE whatman membrane filters.
14. Use cell scrapers to quickly collect the yeast on the membrane and immediately dip into a 50mL falcon filled with liquid nitrogen.
15. Place the tube in -80oC for storage. The liquid nitrogen will boil off through the pierced cap.

Cell lysis

Lysis Buffer Preparation

Component	Volumes	Concentrations
1M Tris-HCL 8.0	200uL	20mM
2M KCl	750uL	150mM
1M MgCl ₂	50uL	5mM
CHX (100mg/ml)	10uL	100ug/mL
1M DTT	10uL	1mM
20% Triton X-100	500uL	1%
Water	8.39mL	

Table 2: Lysis Buffer recipes.

Note, CHX and DTT can be added fresh on day of use.

1. Take the sample falcon tube from the -80oC and fill it with liquid nitrogen.
2. Using a p1000, slowly drop 2mL lysis buffer into the falcon, so that the lysis buffer forms small droplets rather than a large block.

The samples are ready for lysis or can be stored in -80oC until ready.

Mixer Mill Cryogenic Lysis

1. Chill the grinding jars and balls in a liquid nitrogen filled Styrofoam box.
2. Drain out any liquid nitrogen present in the falcon.
3. Add frozen yeast to grinding jar, add the ball and seal tightly.
4. Re-chill the jar in liquid nitrogen.
5. Loosen the grinding jar ¼ turn, place in mixer mill and grind for 3mins at 15Hz.
6. Tighten the jar and return to liquid nitrogen.
7. Repeat for 5 more cycles.
8. Chill the jars, along with two spatulas, in liquid nitrogen.

- Partly fill the conical tube with liquid nitrogen and place it upright in a liquid nitrogen bath. Open the grinding jar and recover the lysate powder into the conical flask, re-chill spatulas as needed.
- Cap the tube of yeast lysate using tinfoil with a pierced hole and place it in the -80oC upright, until the liquid nitrogen evaporates. Samples can be stored in -80oC indefinitely.

Lysate Clarification

- Thaw the yeast lysate gently (on ice or in cold room) and immediately centrifuge the tube at 3000g, 4°C for 5mins.
- Recover supernatant into 2mL RNase-free tube.
- Further clarify the supernatant by spinning at 20,000g, 4oC for 10mins and recover the supernatant again in 1.5mL epi (typically ~1.5mL).
- Store Ribo-Seq samples indefinitely at -80oC or proceed directly to RNase I treatment.

Qubit Readings

- Mix 199uL buffer and 1uL dye in 1.5mL epi.
- Add 1uL sample to the tube.
- Take readings (select correct assay: RNA BR)

For **ribosome profiling**, aliquot 30ug of total RNA into 1.5mL RNase-free tubes.

RNase 1 Treatment

Polysome Buffer

Component	Volumes	Concentrations
1M Tris-HCL 8.0	200uL	20mM
2M KCl	750uL	150mM
1M MgCl ₂	50uL	5mM
CHX (100mg/ml)	10uL	100ug/mL
1M DTT	10uL	1mM
Water	8.89mL	

Table 3: Polysome Buffer recipe.

- Thaw RNase1 and lysate on ice.
- Add polysome buffer (lysis buffer (5mM MgCl₂) without Triton-X100) to make final volume 200uL.
- Transfer 1.5uL RNase1 to each lysate, place on T-shaker for 45 minutes, 25oC, 220rpm.
- For a control, add SUPERase*In (optional)
- Add 10uL SUPERase*In to stop nuclease digestion.

Preparation of Density Gradients

Prepare the two sucrose buffers as followed.

Component	4 Gradients	6 Gradients
1M Tris pH 7.5	1mL	1.5mL
2M KCl	3.5mL	5.35mL
1M MgCl ₂	250uL	375uL
CHX (100mg/ml)	50uL	75uL
1M DTT	50uL	75uL
20U/uL Superasin	50uL	75uL
MilliQ Water	45.15mL	67.73mL
10% Sucrose Gradient Buffer	3.0g	4.5g
	27.8mL of Buffer	41.7mL of Buffer
50% Sucrose Gradient Buffer	14.9g	22.35g
	20.8mL of Buffer	31.2mL of Buffer

Table 4: Density Gradients recipe.

1. Make up the polysome buffer stock first in a 50ml falcon tube (if doing four gradients).
2. Weigh out the two sucrose measurements into two separate 50ml falcon tubes.
3. Add the buffer to these two falcon tubes and vortex.
4. Set on a bench top mixer, covering them from light.
5. Using a marker, draw a line on each gradient tube at the halfway point.
6. Place each tube into the gradient rack.
7. Once solution is fully dissolved, using two syringes, first add the 10% sucrose solution to the bottom of the gradient tube until the meniscus is level with the halfway mark.
8. Using the second syringe, add the 50% sucrose solution to the bottom of the gradient tube until the meniscus of the 50% solution reaches the halfway mark.

Note: the 10% solution is lighter and should sit on top of the 50% solution

9. Add the gradient tube caps to each tube and remove the overflow of buffer solution from the top.
10. Turn on the Gradient Master Machine and ensure that the plate is levelled.
11. Slowly place the gradient rack onto the plate (the rack is magnetic so care must be taken to not disturb the contents of the gradient tube).
12. Using the rotor SW41ti, mix the gradients in their tubes (takes approx. 6 minutes).
13. Once the gradients are done, move the gradient rack and tubes into the cold room until sample is ready to be added.

Ultra-centrifugation

1. Book a slot to use the centrifuge (sign book at correct date).
2. Select 4oC and then select pre-cool heat.
3. Thaw lysates on ice, thaw the gradients in cold room.
4. Remove 200uL from each gradient tube, keep in a 2mL Eppendorf.
5. Add lysate to gradient tubes in equal amounts.
6. Remove rotor and rack from cold room.
7. Remove caps from the metal tubes, insert gradient tubes into the rack of larger metal tubes, for balancing purposes, if you use 2 gradients, and fill them at 1 and 4. Reseal caps.
8. Insert all of the tubes to the correct numbered position on the rotor, make sure they are fitted properly.
9. Lift the rotor from the red base and carefully insert into the rotor junction inside the chamber.
10. Slide the chamber door to close.
11. Set the RPM to 35,000 and time to 3hrs.
12. Press vacuum to depressurize.
13. Press enter and then start to begin.
14. Return in approx. 10mins to make sure vacuum/rotor is working.
15. Fill in logbook with name, signature, date, problems.
16. When finished, wait until rpm is zero, and turn off vacuum. Turn off using switch on left hand side.

UV Fractionation and Monosome Collection

1. Assemble the apparatus. We use 1.5mL/min flow speed. Also get a UV 96 well plate and timer. Sensitivity = 1.0-0.5, wait for light on output reader to go green (warm up).
2. Using the large syringe, uptake some 60% CsCl solution, take from the reusable bottle, not the stock. Remove air from syringe.
3. You will want to run through water to make sure everything is clean and operational.
4. Fill the polypropylene tube with milli-Q water, stick the needle through the already present hole with needle and run through for a few mins.
5. Using rapid reverse, remove some CsCl back into the syringe. Leave a noticeable amount remaining in the tube and discard. This tube can be reused so wash it.
6. When tube is removed, start the forward pump to remove any air bubbles until CsCl comes through, then stop. You are now ready for your samples
7. Ensure that the polysome machine is connected to laptop and open the TracerDAQ program.

8. Set the y axis to variable and start the TracerDAQ recording.
9. Wait until the first large peak is visible and the graph begins to level off before inserting the outflow tube into the 1st well of a 96 well plate, hold for 12 seconds and move to the next and so on.
10. Stop collection of outflows once graph has fully levelled off (around 8 minutes), cover 96 well plate and move directly to next step or store in -20 overnight.
11. Disassemble all parts and wash with distilled water and leave them on a layer of tissue to dry on bench. For the inflow tube (thick), wash with the distilled water tap. For the thinner outflow tube, use a p1000 to wash through 1mL Ethanol, then 1mL milli-Q water, then 2 x 1mL empty pipette to push through any remaining water or moisture.

RNA Isolation

Selecting Wells

1. Thaw 96 well plate in a box of ice (if left overnight).
2. Once all wells are clear and no bubbles are present, wipe the bottom of plate with a cloth to ensure no condensation is present.
3. Place the plate in the plate reader and set the absorbance to 240nm.
4. Read plate and save table of results.
5. From the table, select the wells that have the highest abs value, these will be the wells of which the monosome was collected within.

Trizol

1. Add 750uL of TRizol LS per 0.25mL sample volume. This is recommended for lysates, but for clarified lysates and monosome fractions this volume of Trizol is not necessary.
2. Pipette up and down to mix.
3. Incubate at room temperature for 5 minutes to permit complete dissociation of the nucleoprotein complex.
4. Add 200uL of chloroform per 750uL TRizol added.
5. Incubate for 2-3 minutes.
6. Centrifuge 15mins at 12,000g 4oC.
7. Carefully transfer aqueous phase to new tube at 45-degree angle.

Library Generation

Size selection from gel

Prepare a 15% PAGE gel with Bis acrylamide and Urea as followed.

Component	7.5mL	15mL	50mL
40% aacrylamide/bis acrylamide (19:1)	2.815mL	5.63mL	18.76mL
Urea	3.6g	7.2g	24g
10X TBE	750uL	1.5mL	5mL
Water	950uL	1.9mL	6.33mL

Dissolve at 37°C then filter using an air vacuum.

Component	7.5mL	15mL	50mL
10% APS (Ammonium Sulphate)	18.75uL	37.5uL	125uL
TEMED	3.75uL	7.5uL	25uL
Final Volume	7.5mL	15ml	50mL

Table 4: 15% Page gel recipe.

An excess of 10% APS can be added to help solidify the gel faster.

1. Prepare gel solution as shown, once prepared, add immediately to gel plates and insert comb
2. Leave to polymerise for 2 hours, then set up the tank and add 1x TBE buffer for 30mins.
3. Using a syringe, clean the wells with buffer from the tank.
4. Once polymerised, pre-run at 15mA (300V) for 20-30mins (1x TBE).
5. Clean again with syringe.
6. Prepare samples:
 1. mRNA sample: 4uL + 2uL 3x loading dye.
 2. Footprint sample: 12uL + 6uL 3x loading dye.
 3. Marker: 1uL marker + 7uL water + 4uL 3X loading dye
7. Allow for gaps between each sample well if possible, as to ensure that overloading won't occur.
8. Run gel for 70mins at 15mA (300V).
9. After gel is run, add 20ml of 1xTBE buffer from the tank to a square plate.
10. Add 2uL of 1000X SYBR Gold to the plate and place on a bench top shaker for 2 minutes.
11. Add gel to the plate and leave on bench top shaker for a further 3 minutes.
12. View gel in geldock reader under Ethidium Bromide settings.
13. Print out Gel sample.

Gel extraction

Component	Volume	Concentration
3M NaOAc pH 5.5	1mL	300mM
0.5M EDTA	20uL	1mM
10% SDS	250uL	0.25% v/v
MilliQ Water	8.999973mL	

Table 5: RNA elution buffer recipe.

1. Cut out required gel representing footprints.
2. Place each gel slice in a clean non-stick microfuge tube.
3. Add 500uL RNA gel elution buffer.
4. Shake overnight in shaker to extract RNA from slices – 1000rpm
5. Briefly centrifuge and collect the liquid, transfer to new epi.
6. Add 1.5uL glycol-blue
7. Add 500uL of isopropanol
8. Precipitate in -20oC overnight or -80 for 1 hour.
9. Pellet by centrifugation for 30mins at 20,000g, 4oC.
10. Remove supernatant and wash pellet in 1mL cold 80% Ethanol.
11. Carefully pipette all liquid from the tube and place it sideways on a rack. Leave for 10mins to air dry.
12. Re-suspend in 4uL Milli-Q water.

PNK dephosphorylation and Linker Ligation

Enzymatic pre-adenylation of linker using Mth RNA ligase

Linker oligonucleotides require enzymatic pre-adenylation prior to ligation to RNA fragments with T4 Rnl2(tr) K227Q.

1. Combine :
 - 1.2 µL linker oligonucleotide at 100 µM
 - 2 µL 10X 5' DNA adenylation action buffer
 - 2 µL 1 mM ATP
 - 12.8 µL water
 - 2 µL Mth RNA Ligase
2. Incubate for 1 h at 65°C, then heat-inactivate the enzyme by incubation at 85 °C for 5 min.
3. Add 30 µL water to the sample and then purify using the Zymo Oligo Clean & Concentrator kit according to the manufacturer's instructions, except elute in 6 µL nuclease-free water.

- Store at -20°C and avoid repeated freeze–thaw.

Index	Primer	Barcode	Oligo sequence
1	NI-810	ATCGT	5'-/5Phos/NNNNNATCGTAGATCGGAAGAGCACACGTCTGAA/3ddC/
2	NI-811	AGCTA	5'-/5Phos/NNNNNAGCTAAGATCGGAAGAGCACACGTCTGAA/3ddC/
3	NI-812	CGTAA	5'-/5Phos/NNNNNCGTAAAGATCGGAAGAGCACACGTCTGAA/3ddC/
4	NI-813	CTAGA	5'-/5Phos/NNNNNCTAGAAGATCGGAAGAGCACACGTCTGAA/3ddC/
5	NI-814	GATCA	5'-/5Phos/NNNNNGATCAAGATCGGAAGAGCACACGTCTGAA/3ddC/
6	NI-815	GCATA	5'-/5Phos/NNNNNGCATAAGATCGGAAGAGCACACGTCTGAA/3ddC/
7	NI-816	TAGAC	5'-/5Phos/NNNNNTAGACAGATCGGAAGAGCACACGTCTGAA/3ddC/
8	NI-817	TCTAG	5'-/5Phos/NNNNNTCTAGAGATCGGAAGAGCACACGTCTGAA/3ddC/

Table 6: Linker sequences.

Note: each sample uses 0.5uL of linker.

Dephosphorylation

- Set up each phosphorylation reaction as follows:

Component	Volume (uL)	Final
RNA sample	3.5 uL	NA
T4 PNK buffer (10X)	0.5 uL	1X
T4 PNK (10U/uL)	0.5 uL	5U
SUPERase*In	0.5 uL	10U

Table 7: PNK reaction recipe.

- Incubate for 1 h at 37°C .
- Prepare the linker ligation by adding components as follows directly to the dephosphorylation reaction:

Linker ligation

Component	Volume (uL)	Final
50% w/v PEG-8000	3.5 uL	17.5%
10X T4 RNA ligase buffer	0.5 uL	1X
Preadenylated linker (20uM)	0.5 uL	1uM
T4 Rnl2(tr) K227Q (200U/uL)	0.5 uL	100U

Table 8: Linker ligation reaction recipe.

1. Incubate the ligation reaction for 3 h at 22°C. Leave on bench (not shaking).
2. Run a 15% PAGE gel and size select for ligated products, run the A linker as a marker.
 - a. mRNA sample: 10uL + 5uL 3x loading dye.
 - b. Marker: 1uL marker + 7uL water + 4uL 3X loading dye.
3. After gel is run, add 20ml of 1xTBE buffer from the tank to a square plate.
4. Add 2uL of 1000X SYBR Gold to the plate and place on a bench top shaker for 2 minutes.
5. Add gel to the plate and leave on bench top shaker for a further 3 minutes.
6. View gel in geldock reader under Ethidium Bromide settings.

Gel Extraction

1. Cut out required gel representing footprints.
2. Place each gel slice in a clean non-stick microfuge tube.
3. Add 500uL RNA gel elution buffer.
4. Shake overnight in shaker to extract RNA from slices ~ 1000rpm.
5. Briefly centrifuge and collect the liquid, transfer to new Eppendorf.
6. Add 1.5uL glycol blue.
7. Add 500uL of isopropanol.
8. 8. Precipitate in -20°C overnight or -80 for 1 hour.
9. 9. Pellet by centrifugation for 30mins at 20,000g, 4°C.
10. Remove supernatant and wash pellet in 1mL cold 80% Ethanol.
11. Carefully pipette all liquid from the tube and place it sideways on a rack. Leave for 10mins to air dry.
12. Resuspend in 10uL Milli-Q water.

Reverse transcription

1. As markers, use RT primer and a linker.
2. Add 2uL reverse transcription primer (NI-802) at 1.25uM to all samples and controls.
3. Denature for 5 min at 65°C in a PCR machine and then place on ice.
4. Cool the PCR machine to 50°C.
5. Set up the following reaction:

Component	Volume (uL)	Final
RNA sample and primer	12 uL	NA
5X Protoscript II buffer	4 uL	1X
dNTPs	1 uL	10mM each
0.1M DTT	1 uL	5mM
SUPERase.In (20U/uL)	1 uL	20U
Protoscript II (200U/uL)	1 uL	20U

Table 9: Reverse transcription reaction recipe.

1. Incubate for 30min at 50°C (Protoscript) or 55°C (superscript III).
2. Hydrolyse the RNA template by adding 2.2uL 1M NaOH to each tube and incubate at 70°C for 20min.
3. Add 28uL water, bringing volume to 50uL.
4. Purify sample using oligo clean and concentrator kit, except elute in 6uL water.
5. Proceed to RT gel.

RT Polyacrylamide Gel

Component	7.5mL	15mL	50mL
40% acrylamide/bis-acrylamide (19:1)	1.41mL	2.82mL	9.4mL
urea	3.6g	7.2g	24g
10X TBE	0.75mL	1.5mL	5mL
water	2.36mL	4.72mL	15.73mL

Dissolve at 37°C then filter using an air vacuum.

Component	7.5mL	15mL	50mL
10% APS (Ammonium Sulphate)	18.75uL	37.5uL	125uL
TEMED	3.75uL	7.5uL	25uL
Final Volume	7.5mL	15ml	50mL

Table 10: Reverse transcription PAGE gel recipe

1. Once polymerised, prerun at 15mA (300V) for 20mins in 1X TBE.
2. Denature samples 90sec at 80oC.
3. As marker, use 2uL RT reverse primer (1341)
4. Each sample: 8uL cDNA + 4uL loading dye.
5. Run gel at 15mA (300V) for 45 minutes to separate non-extended primer from the RT product well.
6. The light blue dye runs at approx. 30nts.

Gel extraction

1. Cut out required gel representing footprints.
2. Place each gel slice in a clean non-stick microfuge tube.
3. Add 750uL RNA gel elution buffer.
4. Shake overnight in shaker to extract RNA from slices ~ 1000rpm.
5. Briefly centrifuge and collect the liquid, transfer to new epi.
6. Add 1.5uL glycol blue.
7. Add 750uL of isopropanol.
8. Precipitate in -20°C overnight or -80 for 1 hour.
9. Pellet by centrifugation for 30mins at 20,000g, 4°C.

10. Remove supernatant and wash pellet in 1mL cold 80% Ethanol.
11. Carefully pipette all liquid from the tube and place it sideways on a rack. Leave for 10mins to air dry.
12. Resuspend in 12uL Milli-Q water.

Circularisation

Component	Volume (uL)	Final
cDNA	12 uL	NA
CirLigase II 10X buffer	2 uL	1X
5M Betaine	4 uL	1M
50mM MnCl ₂	1 uL	2.5mM
CirLigase II (100U/uL)	1 uL	100U

Table 11: Circularisation reaction recipe

1. Resuspend RT precipitates in 12uL water in a PCR tube.
2. Prepare circularization reaction as shown in table.
3. Incubate for 1 hour at 60°C in thermal cycler.
4. Heat inactivate the enzyme at 80°C for 10 minutes.
5. The circularization products may be stored at -20°C.

PCR amplification

1. The goal here is to aim for maximum of 6 PCR cycles to reduce PCR bias.
2. For each PCR, it is important to have a particular ratio of template to volume. You ideally want no more than 5%. This strategy ensures that the primer concentration remains greater than 10 times the concentration of extended PCR product, and thus that in later cycles primer annealing and extension predominates over re-annealing of the two template strands.
3. For each sample,
 1. Prepare a 10uL reaction with 0.5uL template
 2. Prepare a 50uL reaction with 2.5uL template
4. Attempt cycles: 6, 8, 10, 12, 14
5. For different cycles, press pause at the end of the elongation cycle and remove a PCR tube.
6. Once cycle for each sample is determined, set up a 50uL reaction and resuspend.
7. Gel extraction, resuspend in 11uL.

Component	Amount per run (μ l) for 50uL	Final
5X Phusion HF buffer	10 uL	1X
dNTPs 10 mM ea.	1 uL	0.2 mM each
10 μ M Forward Primer	2.5 uL	500 nM
10 μ M Reverse Primer	2.5 uL	500 nM
Circularized cDNA template	see above	1–256 pM
Nuclease free water	q.s. to 49.5 μ l	NA
Phusion polymerase (2 U/ μ l)	0.5 uL	1 U

Component	1X reaction (20uL)	1X reaction (10uL)
Phusion buffer (5X)	4 uL	2 uL
dNTPs (10mM)	0.4 uL	0.2 uL
Illumina forward primer (10uM)(1350)	1 uL	0.5 uL
Illumina reverse index primer (10uM)(1461-1469)	1 uL	0.5 uL
circularized PCR template	0.5 uL	0.25 uL
mQ water	12.9 uL	6.45 uL
phusion DNA polymerase (2U uL ⁻¹)	0.2 uL	0.1 uL

Table 12: PCR Reaction recipe.

The forward primer is NI-798. Reverse primers are NI-799, 822–826

PCR Cycle

Initial denaturation: 30 seconds at 98°C

Cycles 7-16:

Denaturation: 10 seconds at 98°C

Annealing: 10 seconds at 65°C

Extension: 5 seconds at 72°C

Final extension: 5 minutes at 72°C

Use different PCR primer for each sample to be pooled:

PCR Gel

8% polyacrylamide gel

10X TBE solution	500µl	750 µl
19:1 acrylamide:bisacrylamide solution (40%)	1 ml	1.5 ml
mQ H2O	3.5 ml	5.25 ml
TEMED	10 µl	15 µl
Ammonium Persulfate (10% w/v) (APS)	35 µl	52.5 µl

Table 13: PCR gel recipe.

1. Stain gel for 3 mins with 1X SYBER gold in 1X TBE.
2. Cut product bands and continue with an overnight gel extraction by adding 750uL gel extraction buffer.
3. Rotate at ~700rpm overnight.

Gel extraction

1. Spin down tubes briefly and transfer the buffer to fresh 1.5mL epi.
2. Add 1.5uL glycoblue and 750uL isopropanol.
3. Spin for 30mins at 20,000 x g for 30mins and discard supernatant.
4. Add 750uL 80% EtOH and spin again for 15 mins and discard supernatant.
5. Let the pellet air dry for 5 minutes.
6. Resuspend pellet in 15uL Tris pH 8.0.
7. Store at -20°C.

Bioinformatic analysis and processing

Initial processing of data

The following software was used for processing the fastq files. All software was installed using the conda environment package manager. As samples were multiplexed using different UMI's, generating separate files could be done easily by demultiplexing the reads and separating them based on their barcode sequence added during the linker ligation step of the protocol. Cutadapt (Martin, 2011) (<https://cutadapt.readthedocs.io/en/stable/>) was used to remove the common adapter found in both biological replicates followed by the use of Bowtie (<http://bowtie-bio.sourceforge.net/manual.shtml>), a short read aligner to align reads to a reference genome (Langmead et al., 2009). The reference genome that was used for alignment was downloaded from Genbank which can be found on Pubmed (available here: https://www.ncbi.nlm.nih.gov/genome/194?genome_assembly_id=277073). Following alignment using Bowtie, the files were individually formatted into SAM and BAM file formats using Samtools (Li et al., 2009), a set of utility tools designed for the interaction and post processing of short DNA sequence reads (<http://www.htslib.org/>), for further analysis using HTseq (<https://htseq.readthedocs.io>) (Anders et al., 2015).

Utilizing Trips-Viz

Three files were needed to populate Trip-Viz, a GTF file containing all the genetic information regarding *Y. lipolytica* W29, a genomic fasta containing the entire sequence for each chromosome on a singular line, and the aforementioned pseudo-transcript fasta, a fasta file containing the CDS regions for every protein coding gene, plus an extension of 300 nucleotides upstream and downstream. Each file was generated through python using files taken from Genbank (available here: https://www.ncbi.nlm.nih.gov/genome/194?genome_assembly_id=277073). As the strain of *Y. lipolytica* used for ribosome profiling was not well annotated, the sequenced data was blasted against the *S. cerevisiae* reference database to provide annotation. This was done through the BLAST command line for linux. The results from this were placed in a data matrix generated in python using a package called Pandas (<https://pandas.pydata.org/>). The following filters were added to remove weak links for annotation, one; the percentage identity must be above 30%, two; the E-value was capped at 0.01, meaning there is 1% chance of a hit happening at random and three; ensuring that bit score was no lower than 50. Bit score relates to the likelihood of a sequence database in which the current match could be found by chance.

Stock Solutions and Buffers

YPD Media

Ingredient	1000ml/1L	750ml
1% Yeast Extract	10g	7.5g
2% Peptone	20g	15g
2% Glucose	20g	15g

Ingredient	15ml/1 plate	150ml/10 plates
1% Yeast Extract	0.15g	1.5g
2% Peptone	0.3g	3.0g
2% Glucose	0.3g	3.0g
2% Agar	0.3g	3.0g

Lysis Buffer

Component	Volumes	Concentrations
1M Tris-HCL 8.0	200uL	20mM
2M KCl	750uL	150mM
1M MgCl ₂	50uL	5mM
CHX (100mg/ml)	10uL	100ug/mL
1M DTT	10uL	1mM
20% Triton X-100	500uL	1%
Water	8.39mL	

Polysome Buffer

Component	Volumes	Concentrations
1M Tris-HCL 8.0	200uL	20mM
2M KCl	750uL	150mM
1M MgCl ₂	50uL	5mM
CHX (100mg/ml)	10uL	100ug/mL
1M DTT	10uL	1mM
Water	8.89mL	

Polysome Gradient Buffers

Component	4 Gradients	6 Gradients
1M Tris pH 7.5	1mL	1.5mL
2M KCl	3.5mL	5.35mL
1M MgCl ₂	250uL	375uL
CHX (100mg/ml)	50uL	75uL
1M DTT	50uL	75uL
20U/uL Superasin	50uL	75uL
MilliQ Water	45.15mL	67.73mL
10% Sucrose Gradient Buffer	3.0g	4.5g
	27.8mL of Buffer	41.7mL of Buffer
50% Sucrose Gradient Buffer	14.9g	22.35g
	20.8mL of Buffer	31.2mL of Buffer

RNA Elution Buffer

Component	Volume	Concentration
3M NaOAc pH 5.5	1mL	300mM
0.5M EDTA	20uL	1mM
10% SDS	250uL	0.25% v/v
MilliQ Water	8.999973mL	

15% PAGE Gel (with Urea)

Component	7.5mL	15mL	50mL
40% acrylamide/bis acrylamide (19:1)	2.815mL	5.63mL	18.76mL
Urea	3.6g	7.2g	24g
10X TBE	750uL	1.5mL	5mL
Water	950uL	1.9mL	6.33mL

Dissolve at 37°C then filter using an air vacuum.

This stock prepared can be stored away from light for future use.

Component	7.5mL	15mL	50mL
10% APS (Ammonium Sulphate)	18.75uL	37.5uL	125uL
TEMED	3.75uL	7.5uL	25uL
Final Volume	7.5mL	15ml	50mL

7.5% PAGE Gel (with Urea)

Component	7.5mL	15mL	50mL
40% acrylamide/bis-acrylamide (19:1)	1.41mL	2.82mL	9.4mL
urea	3.6g	7.2g	24g
10X TBE	0.75mL	1.5mL	5mL
water	2.36mL	4.72mL	15.73mL

Dissolve at 37°C then filter using an air vacuum.

This stock prepared can be stored away from light for future use.

Component	7.5mL	15mL	50mL
10% APS (Ammonium Sulphate)	18.75uL	37.5uL	125uL
TEMED	3.75uL	7.5uL	25uL
Final Volume	7.5mL	15ml	50mL

8% polyacrylamide gel for PCR

Component	5mL	7.5mL
10X TBE solution	500µL	750 µL
19:1 acrylamide:bisacrylamide solution (40%)	1 ml	1.5 ml
mQ H2O	3.5 ml	5.25 ml
TEMED	10 µL	15 µL
Ammonium Persulfate (10% w/v) (APS)	35 µL	52.5 µL

Supplementary Figures

Cell culture and harvesting

Y. lipolytica is a mesophilic yeast that grows comfortably at 24 degrees. As ribosome profiling had not been performed before in W29, the main focus was to establish a framework and so, the yeast was placed in rich liquid cultures of YPD media and grown in favourable conditions as to not alter translation. Cells were harvested during the start of log-phase growth to ensure that the cells would be translationally active. Readings were taken every 30 minutes on a spectrophotometer. Readings underwent a 1 in 10 dilution to give an accurate read. This was done in triplicate and the graph consists of the average value for each time point.

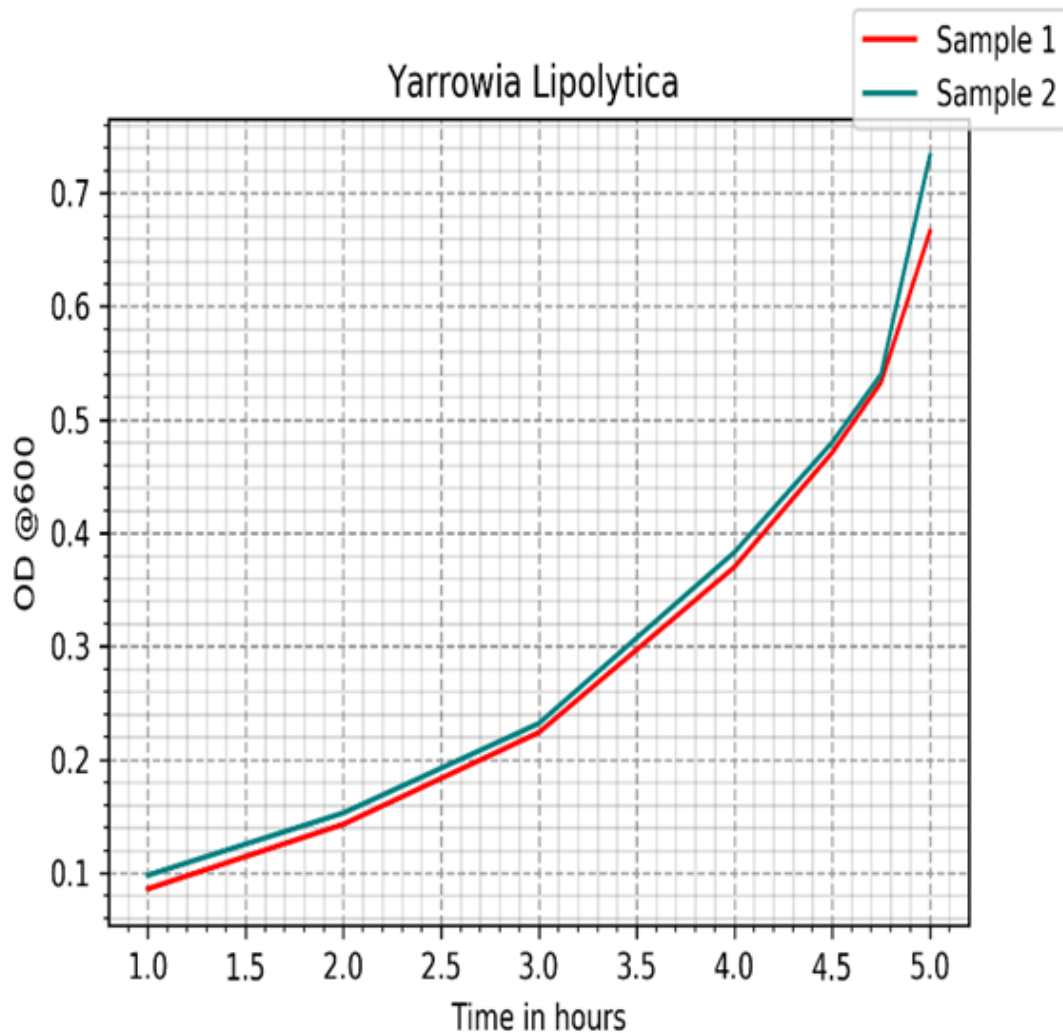


Figure S1: Growth curve of *Y. lipolytica* cultured in YPD liquid media.

Footprint fragment isolation and purification

After phenol/chloroform extraction, further isolation and purification was required to ensure that only the RPFs remained in solution. A size selection gel was used to collect the footprint fragments. As ribosomes are approximately 28 nucleotides in length, the size selection gel should contain a double band marker containing bands that are above and below the 28-nucleotide length but also be close enough to ensure that no unwanted material is captured within the marker. The gel samples were bracketed with two markers containing a 32-nucleotide long band and a 26-nucleotide long band. The samples were loaded onto a 15% PAGE-Urea gel and ran for 70 minutes. The gel was viewed under a gel doc (Bio Rad) and the bands within the two markers were excised from the gel and eluted.

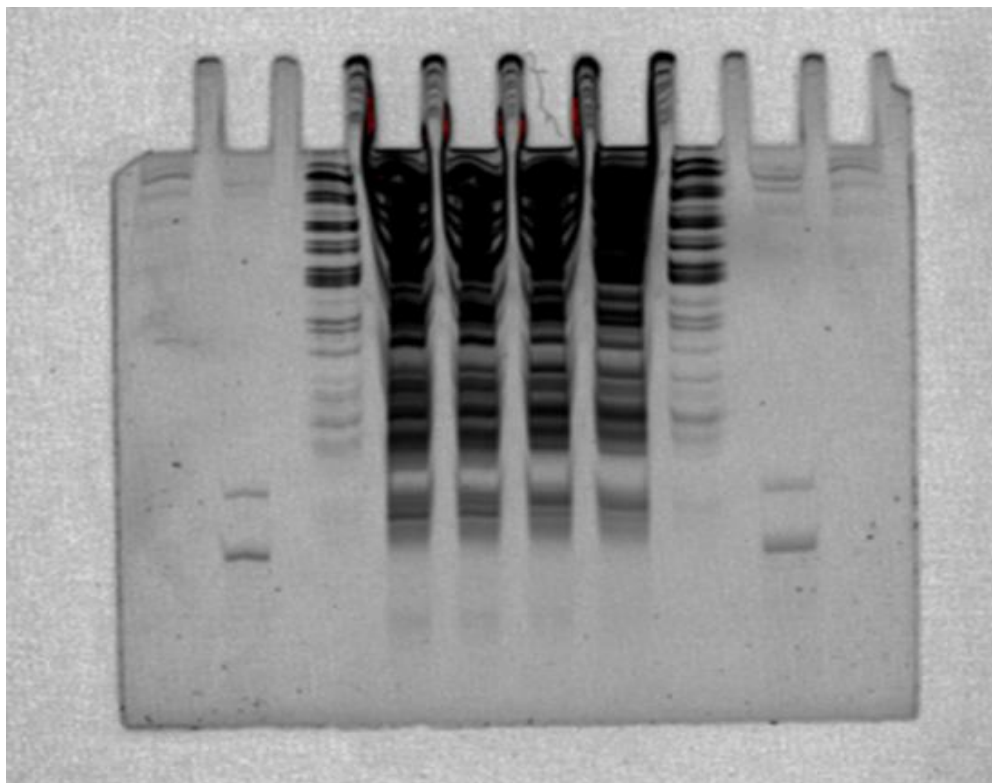


Figure S2. Size selection gel

A total of four samples were used in the size selection gel, two separate biological samples (BR1 and BR2) with two technical replicates of each. The four samples were bracketed by two markers on either side to help visualise bands of interest, while also reducing possible error of slanting when excising the gel slices from the gel. Well contents from left to right: 26/32 Marker, BR1-1, BR1-2, BR2-1, BR2-2, 26/32 Marker.

Footprint fragment end repair and ligation linking

Eluted and recovered RPFs are treated with T4 Polynucleotide Kinase (T4 PNK) to repair the damage caused by RNase1 cleavage. As RNase 1 cleaves at the 3' side of nucleotide bases through 2', 3'-cyclic monophosphate intermediates, T4 PNK works through the removal of 3'phosphoryl groups and prepares the RPF's for subsequent linker ligation. Samples are treated with pre-adenylated linkers and T4 Rnl2(tr) K227Q to add UMIs to allow for multiplexing before sequencing. Samples were bracketed with two negative controls containing the 26/32 nucleotide band marker from the size selection gel that underwent linker ligation without PNK treatment. These controls would not be ligated and therefore not create a heavier product. The ligated products were viewed under a gel dock and isolated for excision.

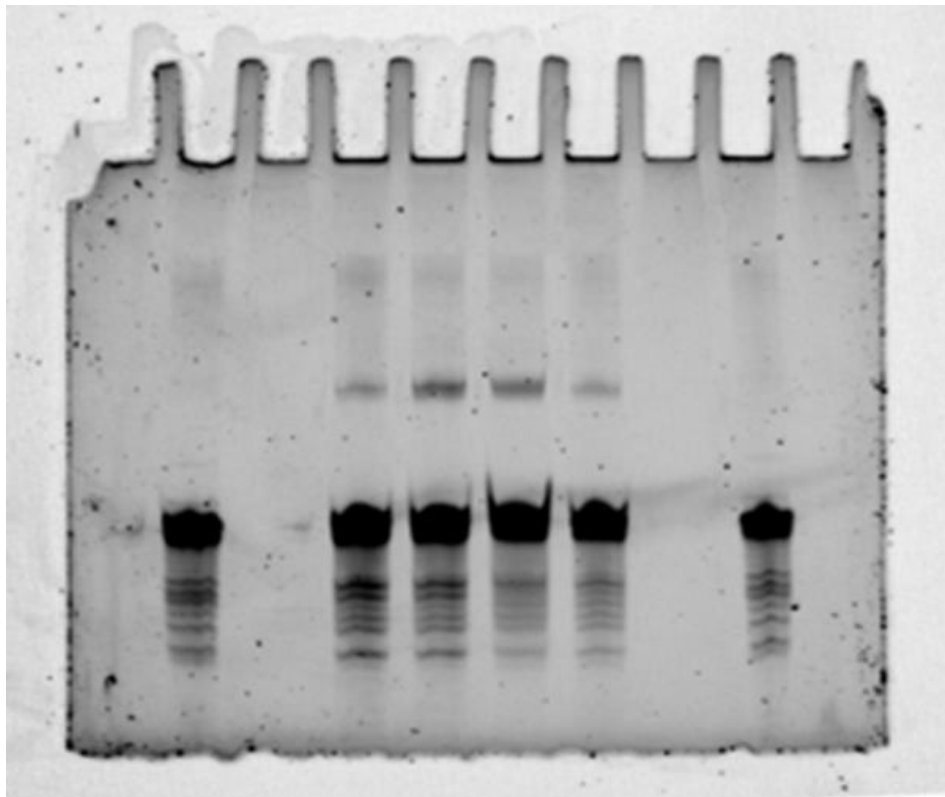


Figure S3: Linker Ligation Gel

From the gel, it can be seen that the ligated produce is much heavier than the un-ligated product. The bands of interest are present in the centre of the gel while the un-ligated material appears closer to the bottom of the gel. The four samples are bracketed by negative controls. Well contents from left to right: Negative control, BR1-1, BR1-2, BR2-1, BR2-2, Negative control.

Reverse Transcription of ligated RPFs

Reverse transcription is the process of generating complementary DNA (cDNA) from RNA. It is a fundamental tool for the study of transcription and translation as it converts genetic information from an unstable condition (RNA) to a much more stable condition in the form of cDNA. Samples were pooled prior to reverse transcription to ensure that there would be enough material moving forward to reduce the total number of PCR cycles required. Samples were treated with Protoscript II and left to incubate for 30 minutes before being hydrolysed with NaOH. Samples were placed on a 7.5% PAGE Urea gel and ran at 300V for 45 minutes. After the gel was ran, samples were viewed under a gel dock and excised for overnight elution.

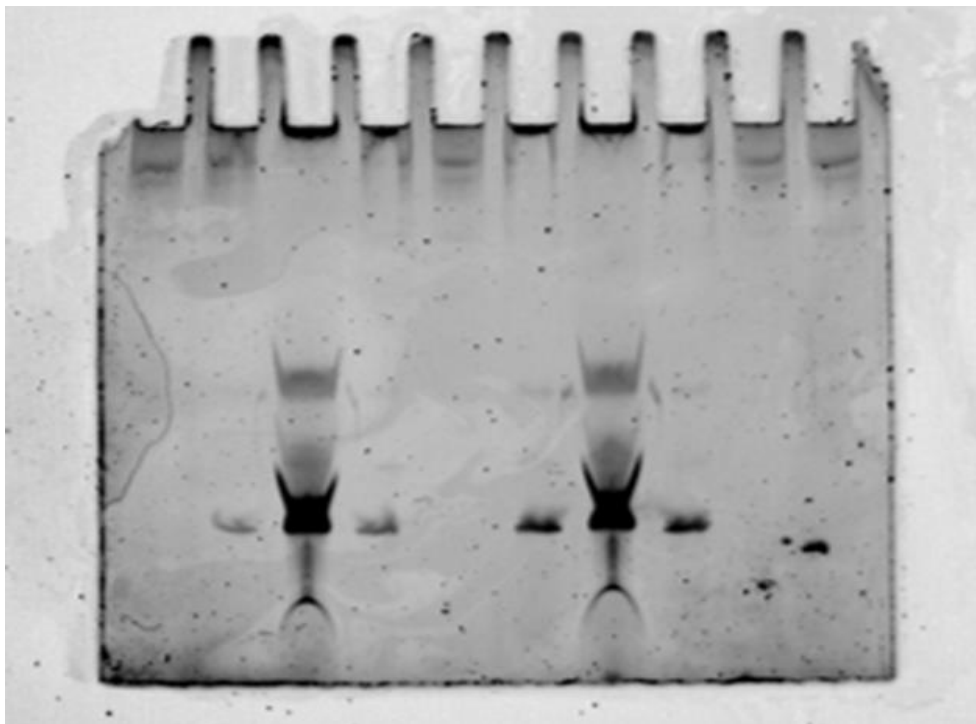


Figure S4: Reverse transcription gel

Samples were pooled together prior to reverse transcription. Bands of interest were the heavier bands that appeared halfway up the gel. The lighter bands that appear towards the lower end of the gel are the reverse transcription primers as these are added in excess.

cDNA library generation

The final step in generating a cDNA library involves the use of PCR to amplify the material that will be sent off to get sequenced. There were precautions when running PCR mainly lowering the amount of cycles as much as possible and ensuring that the ratio of template to volume was no greater than 5%. Putting an emphasis on having a low amount of PCR cycles was to reduce the likelihood of PCR duplicates, a form of contamination caused by a duplication of a read from one DNA fragment. This leads to a false increase to homozygosity. Prior to the actual PCR gel that was run a trial was set up to check how many cycles were required. The trial consisted of four samples, each undergoing a PCR for a different number of cycles.

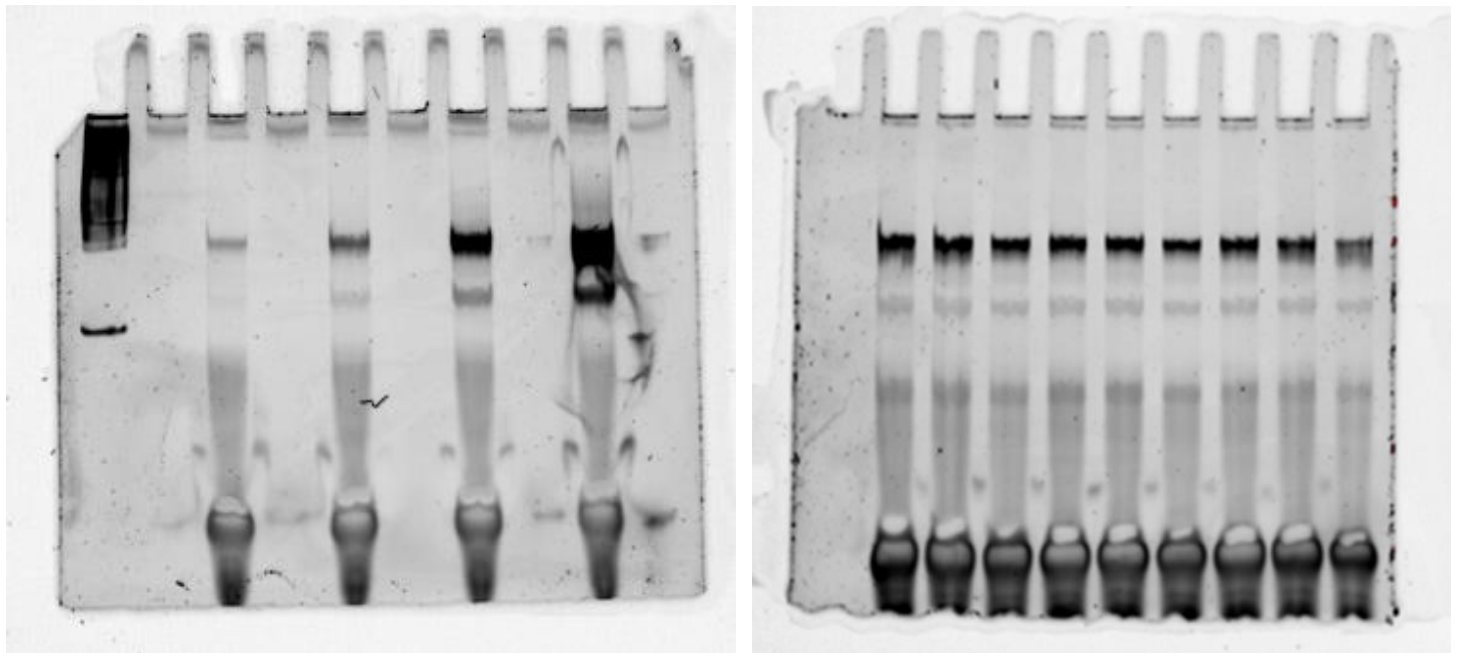


Figure S5A: Trial PCR Gel

From left to right, 6 cycles, 8 cycles, 10 cycles, and 12 cycles. The trial PCR gel was done to ensure that no PCR duplicates would be present in the ribo-seq dataset after library generation. 6 cycles were deemed too little the amount of material needed, while 10 and 12 cycles appeared to be too much material. Figure S5B: PCR Gel. The two biological replicates were pooled together, and nine templates underwent PCR. From the trial gel, it was deemed that 7 cycles were enough to generate the material required for sequencing.

Acknowledgements

I would like to thank John Morrissey and Paval Baranov for their help and supervision throughout all aspects of the research project. I would also like to thank all members of the LPTI lab for their help and teachings throughout. I would like to thank Darren Fenton for his mentorship on the ribosome profile protocol and assistance. Finally, I would like to thank Pat Allen for his help and guidance with all equipment used throughout the protocol. This research was supported by the CHASSY project which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 720824.

References

- Abdel-Mawgoud, A. M., Markham, K. A., Palmer, C. M., Liu, N., Stephanopoulos, G. and Alper, H. S. (2018) 'Metabolic engineering in the host *Yarrowia lipolytica*', *Metab Eng*, 50, pp. 192-208.
- Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq—a Python framework to work with high-throughput sequencing data', *Bioinformatics*, 31(2), pp. 166-169.
- Brar, G. A. and Weissman, J. S. (2015) 'Ribosome profiling reveals the what, when, where and how of protein synthesis', *Nat Rev Mol Cell Biol*, 16(11), pp. 651-64.
- Duncan, C. D. S. and Mata, J. (2017) 'Effects of cycloheximide on the interpretation of ribosome profiling experiments in *Schizosaccharomyces pombe*', *Sci Rep*, 7(1), pp. 10331.
- Egermeier, M., Russmayer, H., Sauer, M. and Marx, H. (2017) 'Metabolic Flexibility of *Yarrowia lipolytica* Growing on Glycerol', *Frontiers in microbiology*, 8, pp. 49-49.
- Gerashchenko, M. V. and Gladyshev, V. N. (2017) 'Ribonuclease selection for ribosome profiling', *Nucleic Acids Res*, 45(2), pp. e6.
- Grundtman, C., Kreutmayer, S. B., Almanzar, G., Wick, M. C. and Wick, G. (2011) 'Heat shock protein 60 and immune inflammatory responses in atherosclerosis', *Arteriosclerosis, thrombosis, and vascular biology*, 31(5), pp. 960-968.
- Hinnebusch, A. G. and Natarajan, K. (2002) 'Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress', *Eukaryotic cell*, 1(1), pp. 22-32.
- Ingolia, N. T. (2016) 'Ribosome Footprint Profiling of Translation throughout the Genome', *Cell*, 165(1), pp. 22-33.
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. and Weissman, J. S. (2012) 'The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments', *Nat Protoc*, 7(8), pp. 1534-50.
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. and Weissman, J. S. (2009) 'Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling', *Science*, 324(5924), pp. 218-23.
- Kiniry, S. J., O'Connor, P. B. F., Michel, A. M. and Baranov, P. V. (2019) 'Trips-Viz: a transcriptome browser for exploring Ribo-Seq data', *Nucleic Acids Res*, 47(D1), pp. D847-d852.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009) 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biology*, 10(3), pp. R25.
- Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J. M. and Ledesma-Amaro, R. (2018a) 'A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β -carotene', *Biotechnol Bioeng*, 115(2), pp. 464-472.
- Larroude, M., Rossignol, T., Nicaud, J. M. and Ledesma-Amaro, R. (2018b) 'Synthetic biology tools for engineering *Yarrowia lipolytica*', *Biotechnol Adv*, 36(8), pp. 2150-2164.
- Ledesma-Amaro, R. and Nicaud, J.-M. (2016) '*Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids', *Progress in Lipid Research*, 61, pp. 40-50.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078-9.
- Liu, H.-H., Ji, X.-J. and Huang, H. (2015) 'Biotechnological applications of *Yarrowia lipolytica*: Past, present and future', *Biotechnology Advances*, 33(8), pp. 1522-1546.
- Liu, L. and Alper, H. S. (2014) 'Draft Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* PO1f, a Commonly Used Metabolic Engineering Host', *Genome announcements*, 2(4), pp. e00652-14.
- Magnan, C., Yu, J., Chang, I., Jahn, E., Kanomata, Y., Wu, J., Zeller, M., Oakes, M., Baldi, P. and Sandmeyer, S. (2016) 'Sequence Assembly of *Yarrowia lipolytica* Strain W29/CLIB89 Shows Transposable Element Diversity', *PLoS One*, 11(9), pp. e0162363.

- Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *2011*, 17(1), pp. 3.
- McGlincy, N. J. and Ingolia, N. T. (2017) 'Transcriptome-wide measurement of translation by ribosome profiling', *Methods*, 126, pp. 112-129.
- Michel, A. M. and Baranov, P. V. (2013) 'Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale', *Wiley Interdiscip Rev RNA*, 4(5), pp. 473-90.
- Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F. and Baranov, P. V. (2012) 'Observation of dually decoded regions of the human genome using ribosome profiling data', *Genome Res*, 22(11), pp. 2219-29.
- Miller, K. K. and Alper, H. S. (2019) 'Yarrowia lipolytica: more than an oleaginous workhorse', *Appl Microbiol Biotechnol*, 103(23-24), pp. 9251-9262.
- Qiao, K., Wasylenko, T. M., Zhou, K., Xu, P. and Stephanopoulos, G. (2017) 'Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism', *Nat Biotechnol*, 35(2), pp. 173-177.
- Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., Green, R., Shen, B. and Liu, J. O. (2010) 'Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin', *Nat Chem Biol*, 6(3), pp. 209-217.
- Talapatra, S., Wagner, J. D. O. and Thompson, C. B. (2002) 'Elongation factor-1 alpha is a selective regulator of growth factor withdrawal and ER stress-induced apoptosis', *Cell Death & Differentiation*, 9(8), pp. 856-861.
- Wang, G., Li, D., Miao, Z., Zhang, S., Liang, W. and Liu, L. (2018) 'Comparative transcriptome analysis reveals multiple functions for Mhy1p in lipid biosynthesis in the oleaginous yeast Yarrowia lipolytica', *Biochim Biophys Acta Mol Cell Biol Lipids*, 1863(1), pp. 81-90.
- Xu, P., Qiao, K., Ahn, W. S. and Stephanopoulos, G. (2016) 'Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals', *Proc Natl Acad Sci U S A*, 113(39), pp. 10848-53.
- Xue, Z., Sharpe, P. L., Hong, S. P., Yadav, N. S., Xie, D., Short, D. R., Damude, H. G., Rupert, R. A., Seip, J. E., Wang, J., Pollak, D. W., Bostick, M. W., Bosak, M. D., Macool, D. J., Hollerbach, D. H., Zhang, H., Arcilla, D. M., Bledsoe, S. A., Croker, K., McCord, E. F., Tyreus, B. D., Jackson, E. N. and Zhu, Q. (2013) 'Production of omega-3 eicosapentaenoic acid by metabolic engineering of Yarrowia lipolytica', *Nat Biotechnol*, 31(8), pp. 734-40.
- Zhang, B., Chen, H., Li, M., Gu, Z., Song, Y., Ratledge, C., Chen, Y. Q., Zhang, H. and Chen, W. (2013) 'Genetic engineering of Yarrowia lipolytica for enhanced production of trans-10, cis-12 conjugated linoleic acid', *Microb Cell Fact*, 12, pp. 70.
- Zhang, H., Wu, C., Wu, Q., Dai, J. and Song, Y. (2016) 'Metabolic Flux Analysis of Lipid Biosynthesis in the Yeast Yarrowia lipolytica Using ¹³C-Labeled Glucose and Gas Chromatography-Mass Spectrometry', *PLOS ONE*, 11(7), pp. e0159187.
- Zhang, S., Jagtap, S. S., Deewan, A. and Rao, C. V. (2019) 'pH selectively regulates citric acid and lipid production in Yarrowia lipolytica W29 during nitrogen-limited growth on glucose', *Journal of Biotechnology*, 290, pp. 10-15.