Extensive post-transcriptional buffering of gene expression in the response to oxidative stress in baker's yeast

William R. Blevins[1,#], Teresa Tavella[1,#], Simone G. Moro[1], Bernat Blasco-Moreno[2], Adrià Closa-Mosquera[2], Juana Díez[2], Lucas B. Carey[2,3], M. Mar Albà[1,2,4,*]

[1] Evolutionary Genomics Groups, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute (IMIM)-Universitat Pompeu Fabra (UPF), Barcelona, Spain

[2] Health and Experimental Sciences Department, Universitat Pompeu Fabra(UPF), Barcelona, Spain.

[3] Center for Quantitative Biology and Peking-Tsinghua Joint Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.

[4] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain.

[#]Shared first co-authorship

[*]To whom correspondence should be addressed.

Keywords: ribosome profiling, RNA-Seq, translation, oxidative stress, yeast

Abstract

Cells responds to diverse stimuli by changing the levels of specific effector proteins. These changes are usually examined using high throughput RNA sequencing data (RNA-Seq); transcriptional regulation is generally assumed to directly influence protein abundances. However, the correlation between RNA-Seq and proteomics data is in general quite limited owing to differences in protein stability and translational regulation. Here we perform RNA-Seq, ribosome profiling and proteomics analyses in baker's yeast cells grow in rich media and oxidative stress conditions to examine gene expression regulation at various levels. With the exception of a small set of genes involved in the maintenance of the redox state, which are regulated at the transcriptional level, modulation of protein expression is largely driven by changes in the relative ribosome density across conditions. The majority of shifts in mRNA abundance are compensated by changes in the opposite direction in the number of translating ribosomes and are predicted to result in no net change in protein level. We also identify a subset of mRNAs which is likely to undergo specific translational repression during stress and which includes cell cycle control genes. The study suggests that post-transcriptional buffering of gene expression may be more common than previously anticipated.

Introduction

In recent years high throughput RNA sequencing (RNA-Seq) has become the method of choice for measuring shifts in gene expression between cells grown in different conditions [1]. However, diverse studies have shown that mRNA levels only partially explain protein levels in the cell [2–5]. In yeast, the correlation between mRNA and protein abundance is typically in the range 0.6-0.7 [2]. In addition, the ratio between protein and mRNA levels may vary across different conditions [3]. For instance, substantial differences in this ratio have been observed during osmotic stress in yeast [6] or after the treatment of human cells with epidermal growth factor [7].

In contrast to RNA-Seq, which measures the total amount of mRNA in the cell, ribosome profiling (Ribo-Seq) only captures those mRNAs that are being actively translated [8]. Each Ribo-Seq read corresponds to one translating ribosome, providing a quantitative view of the amount of protein produced by the cell at any given time. Although this remains an indirect estimate of protein abundance, it has several advantages over proteomics, such as the fact that with Ribo-Seq virtually all translated sequences can be captured, and that one can apply the same pipelines and statistical methods as for RNA-Seq to identify differentially expressed genes.

The response to oxidative stress in the yeast *Saccharomyces cerevisiae* involves a general decrease in mRNA translation initiation as well as the selective transcriptional activation of a set of proteins involved in the maintenance of the redox state of the cell [9-11]. A previous study reported changes in the ratio between the normalized number of Ribo-Seq and RNA-Seq reads, or translational efficiency (TE), of hundreds of genes upon oxidative

59 stress [9], suggesting extensive translational regulation. However, changes in TE alone do
60 not necessarily imply changes in the abundance of the translated proteins. Here, by
61 performing a separate analysis of Ribo-Seq and RNA-Seq data, we show that the majority
62 of genes that show statistically significant differences at the RNA-Seq level do not show
63 similar differences at the Ribo-Seq level, suggesting that, in most cases, changes in
64 mRNA abundance are compensated by changes in ribosome density and are not
65 propagated to the protein level. Our approach also uncovers a subset of differentially
66 expressed genes in which regulation appears to be mainly exerted at the translational level.
67

68 **Results**

69

70 Ribosome profiling experiments in normal and stress conditions

71

72 We extracted ribosome-protected RNA fragments, as well as complete polyadenylated
73 RNAs, from *Saccharomyces cerevisiae* grown in rich media (normal) and in H_2O_2-
74 induced oxidative stress conditions (stress)(Figure 1). We then sequenced the ribosome-
75 protected RNA fragments (Ribo-Seq) as well as complete mRNAs (RNA-Seq) using a
76 strand-specific protocol. The Ribo-Seq data provided a snapshot of the translatome, each
77 read corresponding to one translating ribosome, whereas the number of RNA-Seq reads
78 mapping to a gene was used to quantify the relative abundance of the transcript.

79

80 After quality control of the sequencing reads we obtained 31-36 million Ribo-Seq reads
81 and 12-15 million RNA-Seq reads per sample (Supplementary Table S1). We mapped the
82 reads to the genome and generated a table of read counts per gene for each of the samples.

83 After filtering out non-expressed genes (see Methods), the table contained data for 5,419
84 *S. cerevisiae* annotated genes (ORFs).
85
86 We normalized the RNA-Seq and Ribo-Seq table of counts by calculating normalized
87 counts per million (CPM) in logarithmic scale, or $\log_2$CPM (Supplementary Figure S1).
88 The correlation coefficient between the average Ribo-Seq and RNA-Seq $\log_2$CPM
89 expression values was 0.84 in normal conditions and 0.87 in stress conditions (Figure 2
90 A and B, respectively). As the differences in $\log_2$CPM between RNA-Seq or Ribo-Seq
91 replicates were negligible (Figure 2 C and D, Supplementary Table S2), these values
92 reflect the amount of disagreement between total mRNA and translated protein
93 abundances.
94
95 Ribo-Seq shows a higher correlation with proteomics than RNA-Seq
96
97 The next step was to compare the quantification of gene expression by RNA-Seq and
98 Ribo-Seq to that obtained using proteomics. We extracted the protein fraction from yeast
99 grown in normal and stress conditions and estimated the abundance of different yeast
100 proteins, i.e. the proteome, using mass spectrometry information (Figure 1). We could
101 reliably quantify the protein products of 2,200 genes (see Methods), representing about
102 40% of the genes quantified by RNA-Seq or Ribo-Seq. Normalized protein abundances
103 between pairs of proteomics replicates showed correlation coefficients in the range 0.83-
104 0.93 (Supplementary Table S3), lower than for RNA-Seq or Ribo-Seq replicates (>0.99).
105
106 In normal conditions the correlation coefficient between the transcriptome (RNA-Seq)
107 and the proteome relative abundance units was 0.46. This increased to 0.71 when

108 comparing the translatome (Ribo-Seq) and the proteome units (Figure 3). This indicates
109 that Ribo-Seq-based quantification of gene expression provides a more accurate picture
110 of protein abundance than RNA-Seq data. The average correlation coefficient between
111 the three pairs of proteome replicates was 0.91, setting up a maximum value for any
112 correlation. Differences between RNA-Seq and proteomics quantification estimates may
113 arise because of differences in the half life of the proteins with respect to their cognate
114 mRNAs as well as variations in the translation rate or ribosome density across the
115 transcripts. As the value of 0.71 (Ribo-Seq *versus* proteomics) is intermediate between
116 0.46 (RNA-Seq *versus* proteomics) and 0.91 (proteomics replicates), the two above
117 mentioned factors appear to be relevant to explain the strong uncoupling between mRNA
118 and protein abundance in this system.
119
120 In stress conditions the correlation coefficient between the transcriptome and proteome
121 was 0.62, somewhat higher than in normal conditions. The correlation coefficient
122 between the translatome and the proteome was 0.67, again higher than the same value
123 between the transcriptome and the proteome but lower than the correlation between the
124 proteome stress replicates (0.86). Taken together, these results are consistent with the
125 hypothesis that differences in ribosome density play a role in modulating protein
126 expression.
127
128 Analysis of three nucleotide periodicity
129
130 In actively translated regions mapped Ribo-Seq reads exhibit a characteristic three
131 nucleotide periodicity that results from the codon-to-codon ratcheting movement of the
132 ribosome along the coding sequence [8]. We used the program RibORF [10] to assess the

133 nucleotide periodicity and homogeneity of the Ribo-Seq reads in the annotated coding
134 sequences. According to this analysis, in the vast majority of genes (98%, 5198 out of
135 5304 analyzed genes) the annotated ORF appeared to be translated in both normal and
136 stress conditions, validating our approach of considering all the reads that mapped to the
137 annotated ORFs for the quantification of protein translation.
138
139 In a small fraction of genes, however, we found evidence of alternative translated ORFs
140 (Supplementary Table S4). One example was TOS8, which encodes a homeodomain-
141 containing transcription factor. In this gene active translation of the canonical 831 amino
142 acid long protein by RibORF was only detected in stress conditions; in contrast, a protein
143 of only 81 amino acids was the main translated polypeptide in rich media. The shorter
144 alternative ORF was on a different reading frame to the main protein product and showed
145 no homology to any previously characterized protein. These cases illustrate how detailed
146 examination of the distribution of the Ribo-Seq reads may help uncover proteins that have
147 remained hidden within longer ORFs.
148
149 Ribo-Seq estimates of changes in gene expression are more conservative
150
151 We next calculated the gene expression level fold change (FC) between the two
152 conditions, using RNA-Seq and Ribo-Seq data separately. The $\log_2$FC distribution based
153 on the Ribo-Seq data had a lower variance than the $\log_2$FC distribution using RNA-Seq
154 data (Figure 4A). This indicated a higher range of variation in the mRNA levels, as
155 estimated by RNA-Seq, than in the ribosome-protected fragments. This was consistent
156 with the existence of post-transcriptional buffering of gene expression, as also reported
157 for inter-specific gene expression comparisons of *S.cerevisiae* and *S.paradoxus* [11].

158

159 We considered the possibility that the about 2.5 times higher number of Ribo-Seq reads
160 than RNA-Seq read in the original datasets biased the comparison of log_2FC distributions.
161 In order to test it we subsampled the mapped reads so as to have a similar number of reads
162 in all the RNA-Seq and Ribo-Seq samples (Supplementary Tables S5 and S6). The results
163 were very similar to those observed without subsampling (Supplementary Figure S2),
164 indicating that these observations have a biological origin.

165

166 We also used an alternative method, multidimensional scaling (MDS) [12], to quantify the
167 distance between Ribo-Seq and RNA-Seq gene expression measurements (Figure 4B).
168 We found that the distance between Ribo-Seq normal and stress conditions was shorter
169 that the distance between RNA-Seq normal and stress conditions, which was consistent
170 with the previous observation that log_2FC variance was lower for Ribo-Seq than for RNA-
171 Seq.

172

173 Extensive post-transcriptional buffering of gene expression

174

175 We next performed differential gene expression analysis, separately for Ribo-Seq and
176 RNA-Seq data, using multivariable linear regression with the Limma package [13]. Limma
177 provides a list of differentially expressed genes with the corresponding adjusted p-values.
178 We selected genes with an adjusted p-value < 0.05 and a log_2FC larger than one standard
179 deviation; the latter corresponded to a minimum FC of 1.49 for RNA-Seq data and 1.36
180 for Ribo-Seq data. We used the standard deviation instead of a fixed value to
181 accommodate for the differences in the width of the log_2FC distributions. The number of
182 genes that were differentially expressed was 1,530 for RNA-Seq and 536 for Ribo-Seq.

183

184 The correlation between RNA-Seq and Ribo-Seq gene $\log_2$FC values was quite low (0.18),
185 indicating an important disconnect between the two kinds of data (Figure 4C). Only 127
186 genes showed a significant change in the same direction i.e. homodirectional changes.
187 Genes that were up-regulated during stress according to both RNA-Seq and Ribo-Seq
188 included protein functions known to be activated at the transcriptional level in response
189 to stress, such as hexoquinases or heat shock proteins [14]. The number of genes annotated
190 with the Gene Ontology (GO) term 'oxidation reduction process' was similar for RNA-
191 Seq or Ribo-Seq up-regulated genes (17 and 15, respectively), supporting that these genes
192 are essentially regulated at the level of transcription and can be effectively detected with
193 both kinds of sequencing data.

194

195 The vast majority of genes were only significant at the transcriptome or the translatome
196 levels (1,413 and 409 genes, respectively; Figure 4C). The first group was formed by
197 genes that showed significant changes in relative transcript abundance but not in the
198 relative number of ribosome-protected fragments, supporting extensive post-
199 transcriptional buffering of gene expression. The data indicated that about a quarter of the
200 genes in the genome may be undergoing compensatory changes: when mRNA levels
201 increase ribosome density per transcript decreases and the other way round. The levels of
202 the proteins encoded by these genes are not expected to change despite significant
203 changes in the corresponding mRNA abundance.

204

205 The second group, translatome-only differentially expressed genes, represented cases in
206 which mRNA levels did not change but the density of ribosomes per transcript showed a
207 significant increase or decrease in stress relative to normal. This would be consistent with

208 the expression of these genes being primarily modulated at the level of translation. We
209 identified many more genes under differential translational repression than activation
210 (360 *versus* 49, Figure 4C), suggesting that the former mechanism may be more prevalent
211 that the first one in response to stress.
212
213 Finally, we found a subset of cases showing opposite changes in RNA-Seq and Ribo-Seq
214 data. The main group was formed by 70 genes showing increased mRNA levels but
215 decreased translation in stress *versus* normal. One simple explanation would be that, for
216 these genes, there is an mRNA fraction that is stored in a translational inactive highly
217 stable form, whereas the rest is translated at the usual level. More complex scenarios
218 could involve a combination of transcriptional and translational regulatory events.
219
220 Dissecting differential regulation by functional class
221
222 To better understand the biological relevance of our observations, we investigated if
223 certain functional classes were significantly enriched among the sets of differentially
224 expressed genes. We used DAVID [15] to identify significantly over-represented functional
225 clusters (Figure 4D). Only one class, 'oxidation-reduction process', was enriched among
226 genes up-regulated during stress both using RNA-Seq and Ribo-Seq data. This is
227 consistent with transcriptional activation of this set of genes upon stress, increasing the
228 signal for both total mRNA and the translated fraction.
229
230 Three other classes – 'translation', 'ATPase' and 'proteasome' – showed increased
231 mRNA levels during stress, but this was not reflected in an increase in the translated
232 fraction. These classes may be particularly prone to undergo compensated mRNA

233 changes. Among genes that were differentially expressed only when we used Ribo-Seq
234 data 'cell wall', 'mitochondrial intermembrane space' and 'catalytic activity' were
235 enriched among up-regulated genes, whereas 'cell cycle' was enriched among down-
236 regulated genes (Figure 4D).
237
238 Translational efficiency and protein level changes
239
240 To obtain further insights into the regulatory mechanisms of gene expression during
241 oxidative stress in yeast we also compared the translational efficiency (TE; Ribo-Seq
242 normalized counts divided by RNA-Seq normalized counts) of the different genes in the
243 two conditions using the program Ribodiff [16]. We detected 470 genes that showed
244 significantly increased TE during stress (adjusted p-value < 0.05; see Methods); about 82%
245 of them were cases in which the relative mRNA levels had decreased during stress but
246 this change had been compensated by an increase in ribosome density so that no
247 significant changes in the amount of translated protein would be expected (transcriptome
248 downregulated, Table 1). In only about 3% of cases increased TE was associated with
249 translational activation and increased protein production (translatome upregulated, Table
250 1).
251
252 In the case of genes with significantly lower TE in stress than in normal conditions the
253 percentage of compensatory cases was also the predominant scenario, accounting for 50%
254 of the genes in the class (356 out of 714, Table 1). The second most numerous group were
255 genes likely to be actively repressed at the level of translation, accounting for 29% of the
256 genes with significantly decreased TE (29%). The latter genes showed no change in
257 mRNA levels but the relative number of associated ribosomes was lower in stress than in

258 normal conditions, which would be expected to lead to a decrease in the protein levels.
259 This group included 12 genes from the cell cycle functional category (Supplementary
260 Table S7).
261

262 ## Discussion

263

264 The adaptation of organisms to variations in different environmental conditions is
265 associated with the activation or repression of gene expression. These changes are usually
266 studied at the level of complete mRNA molecules using microarrays or next generation
267 sequencing. However, changes in mRNA concentration do not necessarily reflect changes
268 in their encoded protein products [7,11] .
269

270 Here we have explored the usefulness of ribosome profiling data to close the gap between
271 mRNA and protein abundance estimates. Each ribosome profiling read corresponds to
272 one translating ribosome and thus the number of reads that map to a gene reflects the
273 amount of protein that is being made [8,17]. Numerous recent studies have used ribosome
274 profiling to gain insights into novel translation regulatory mechanisms [18,19] or to discover
275 new translated RNA sequences [20-23]. However, there is a lack of studies addressing how
276 ribosome profiling can be used to improve the estimates of protein abundance changes
277 over RNA-Seq-based estimates. Our study shows that Ribo-Seq provides better estimates
278 of protein abundance than RNA-Seq and that the results of differential gene expression
279 analyses are drastically altered if we use Ribo-Seq of RNA-Seq as the source sequencing
280 data.
281

282 The abundance of the different proteins in the cells is usually estimated using mass
283 spectrometry proteomics data [24,25]. This provides a direct measurement of protein
284 abundance that can account for the variations in the stability of different proteins;
285 however, proteomics methods are much less sensitive than current RNA sequencing
286 approaches and not all proteins can be detected in routine analyses [26]. In addition, the
287 results obtained with high-throughput sequencing are more reproducible across biological
288 replicates than those obtained with mass spec proteomics; this confers the former studies
289 increased power to perform differential gene expression analyses.
290
291 Previous studies in yeast indicated that Ribo-Seq showed a higher correlation with
292 proteomics data than RNA-Seq, but these conclusions were drawn after comparing data
293 obtained from different laboratories [8]. Here we generated RNA-Seq, Ribo-Seq and
294 proteomics data for yeast grown in identical conditions, leading to less biased
295 comparisons. These results support the hypothesis that the Ribo-Seq read counts provide
296 a better approximation to protein levels than RNA-Seq read counts.
297
298 We observed that many of the genes that were detected as significantly up- or down-
299 regulated in stress by RNA-Seq did not show any significant changes using the Ribo-Seq
300 data, indicating frequent post-transcriptional buffering of gene expression. Intriguingly,
301 studies comparing the expression of orthologous genes from closely related species have
302 also reported that gene expression is in general more variable when measured by RNA-
303 Seq than Ribo-Seq [11,27]. We found that, during oxidative stress, genes encoding ribosomal
304 proteins and members of the proteasome and ATPase complexes tended to show
305 increased mRNA levels but, at the same time, the rate of translation decreased. We also
306 have to consider that some mRNAs could be transiently stored in P-bodies or stress

307 granules [28–30], becoming inaccessible to the translation machinery. Translation of these
308 transcript could be rapidly reactivated when the stress disappears.
309
310 Transcripts encoding proteins involved in the cell cycle appeared to be modulated
311 differently. In this case there was no apparent change in the number of mRNA molecules
312 but ribosome density decreased, presumably reflecting lower translation rates. Repression
313 of this class of proteins may be related to a slow down of cell division under stress; the
314 cells grown under oxidative stress showed approximately half doubling times when
315 compared to those grown in rich media.
316
317 The results of this study illustrate the importance of performing ribosome profiling
318 experiments to differentiate between changes in mRNA that are likely to result in changes
319 in the protein levels to those that are not. Although obtaining Ribo-Seq data is more
320 labour-intensive than RNA-Seq, the protocols are being simplified and its use is rapidly
321 growing [31–33]. The methodological framework we have developed can be applied to other
322 datasets and help advance our understanding of gene regulation in other conditions.
323

324 **Methods**

325

326 Biological material

327

328 We grew *S. cerevisiae* (S288C) in 500 ml of rich media [34]. In order to induce oxidative
329 stress, 30 minutes before harvesting we added diluted H_2O_2 to the media for a final
330 concentration of 1.5 mM. The cells were harvested in log growth phase (OD600 of ~0.25)
331 via vacuum filtration and frozen with liquid nitrogen.

332

333 Ribosome profiling

334

335 In order to capture ribosome protected mRNAs, cyclohexamide was added one minute
336 before the cells were harvested. Cyclohexamide is commonly used as a protein synthesis
337 inhibitor in order to prevent ribosome run-off and the subsequent loss of ribosome-
338 transcript complexes. One third of each culture was used for ribosome profiling (Ribo-
339 Seq); the rest was reserved for RNA-Seq.

340

341 Cells were lysed using the freezer/mill method (SPEX SamplePrep); after preliminary
342 preparations, lysates were treated with RNaseI (Ambion), and subsequently with
343 SUPERaseIn (Ambion). Monosomal fractions were collected; SDS was added to stop any
344 possible RNAse activity, then samples were flash-frozen with N_2(l). Digested extracts
345 were loaded in 7%-47% sucrose gradients. RNA was isolated from monosomal fractions
346 using the hot acid phenol method. Ribosome-Protected Fragments (RPFs) were selected
347 by isolating RNA fragments of 28-32 nucleotides (nt) using gel electrophoresis. The
348 preparation of sequencing libraries for Ribo-Seq and RNA-Seq was based on a previously
349 described protocol [35]. Pair-end sequencing reads of size 35 nucleotides (2x35bp) were
350 produced for Ribo-Seq and RNA-Seq on MiSeq and NextSeq platforms, respectively.
351 The data has been deposited at NCBI Bioproject PRJNA435567
352 (https://www.ncbi.nlm.nih.gov/bioproject/435567).

353

354 Processing of the sequencing data

355

356 The RNA-Seq data was filtered using Trimmomatic with default parameters (version
357 0.36)[36]. In the Ribo-Seq data we discarded the second read pair as it was redundant and
358 of poorer quality than the first read, and then used Cutadapt [37] to eliminate the adapters
359 and to trim five and four nucleotides at 5' and 3' edges, respectively. Ribosomal RNA
360 was depleted from the Ribo-Seq data *in silico* by removing all reads which mapped to
361 annotated rRNAs. Ribo-Seq reads shorter than 25 nucleotides were not used.
362
363 After quality check and read trimming, the reads were aligned against the *S. cerevisiae*
364 genome (S288C R64-2-1) using Bowtie 2 [38]. For annotation we used a previously
365 generated *S. cerevisiae* transcriptome containing 6,184 annotated coding sequences plus
366 1,009 non-annotated assembled transcripts (see Supplementary data). SAMtools [39] was
367 used to filter out unmapped reads.
368
369 We counted the number of reads that mapped to each gene with HTSeq-count [40]. We used
370 the mode 'intersection strict' to generate a table of counts from the data; the procedure
371 removed about 5% of the reads in the case of RNA-Seq, and 8% in the case of Ribo-Seq.
372 Only genes in which the average read count of the two replicates was larger than 10 in all
373 conditions (normal and stress, for RNA-Seq and for Ribo-Seq) were kept. The filtered
374 table of counts contained data for 5,419 genes; nearly all of them corresponded to
375 annotated genes (5,312 genes).
376
377 For subsampling the number of mapped reads we used SAMtools [39]. We used the function
378 'samtools view' with option '-s 0.X', where X is the percentage of reads that we wish to
379 keep.
380

381 Analysis of three nucleotide periodicity in the mapped Ribo-Seq reads

382

383 We used RibORF [10] to analyze the mapped Ribo-Seq. We analyzed all possible ORFs
384 with a minium length of 9 amino acids and at least 10 mapped reads. WE analyzed 5,304
385 annotated ORFs. RibORF counts the number of reads that fall in each frame and
386 calculates the distribution of reads along the length of the ORF. We used the original
387 proposed cutoff (score > 0.7) to predict translated ORFs.

388

389 Quantification of protein abundance by mass spectrometry

390

391 For our proteomics experiment, we analysed 3 replicates per condition by LCMSMS
392 using a 90-min gradient in the Orbitrap Fusion Lumos. These samples were not treated
393 with cyclohexamide. As a quality control measure, BSA controls were digested in parallel
394 and ran between each sample to avoid carry-over and assess the instrument performance.
395 The peptides were searched against SwissProt Yeast database, using the Mascot v2.5.1
396 search algorithm. The search was performed with the following parameters: peptide mass
397 tolerance MS1 7 ppm and peptide mass tolerance MS2 0.5 Da; three maximum missed
398 cleavages; trypsin digestion after K or R except KP or KR; dynamic modifications
399 oxidation (M) and acetyl (N-term), static modification carbamidomethyl (C). Protein
400 areas were obtained from the average area of the three most intense unique peptides per
401 protein group. Considering the data from all 6 samples, we detected proteins from 3,336
402 genes. We limited our quantitative analysis to a subset of 2,200 proteins which had
403 proteomics hits for at least 3 unique peptides; this filter eliminates noise arising from
404 technical challenges of quantifying lowly abundant proteins with LCMSMS.

405

406 Differential gene expression analysis

407

408 The table of counts was normalized to $\log_2$ Counts per Million ($\log_2$CPM) using the
409 function 'cpm' in the R package edgeR [41]. Before performing differential gene expression
410 analysis, we normalized the data using Trimmed Mean of M-values (TMM) from the
411 same package. Finally, we applied the Limma voom method [13] to identify differentially
412 expressed genes, separately for RNA-Seq and Ribo-Seq data (adjusted p-value < 0.05 and
413 $|\log_2 FC| > 1$ SD($\log_2$FC)) .

414

415 We applied the same pipeline to the proteomics data using normalized area values as a
416 quantitative measure of protein abundance. To ensure robustness of the differential
417 expression analysis we used genes which had at least 3 unique peptides and could be
418 quantified in all 6 replicates (1,580 genes); the procedure did not identify any significantly
419 up or down regulated genes, using an adjusted p-value < 0.05. Low sensitivity of this
420 procedure is expected considering the relatively poor correlation of the mass spec
421 replicates (r between 0.83 and 0.93).

422

423 Analysis of functional clusters

424

425 We identified significantly enriched functional clusters in differentially expressed genes
426 using DAVID [15]. The analysis was done separately for over- and under-expressed genes
427 and for RNA-Seq and Ribo-Seq derived data. Only clusters with enrichment score ≥ 1.5
428 and adjusted p-val < 0.05 were retained. In each cluster we chose a representative Gene
429 Ontology (GO) term [42], with the highest number of genes inside the cluster. Figure 4

430 integrates the results obtained with the Ribo-Seq and the RNA-Seq data, the $\log_{10}$ fold
431 enrichment of the significant GO terms is plotted.
432
433 Analysis of translational efficiency
434
435 We searched for genes with significantly increased or decreased translational efficiency
436 (TE)[8] using the RiboDiff program [16]. We selected genes significant at an adjusted p-value
437 < 0.05 and showing $\log_2(TE_{stress}/TE_{normal})$ higher than 0.67 or lower than -0.67 (plus or
438 minus one standard deviation of the distribution).
439

440 **Acknowledgements**

441 We acknowledge the Proteomics Unit of Center for Regulatory Genomics and Universitat
442 Pompeu Fabra for the isolation of proteins from yeast cultures. We are also grateful to
443 Robert Castelo for advice during this project. The work was funded by grants BFU2015-
444 65235-P, BFU2015-68351-P and BFU2016-80039-R, from Ministerio de Economía e
445 Innovación (Spanish Government) - FEDER (EU), and from grant PT17/0009/0014 from
446 Instituto de Salud Carlos III – FEDER. We also received funding from the "Maria de
447 Maeztu" Programme for Units of Excellence in R&D (MDM-2014-0370) and from
448 Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya (AGAUR),
449 grant number 2014SGR1121, 2014SGR0974, 2017SGR01020 and, predoctoral
450 fellowship (FI) to W.B. We also acknowledge support from the EU Erasmus Programme
451 to T.T.
452

453 **Author contributions**

454 WRB, JD, LBC and MMA designed the experiments. WRB performed the growth
455 experiments in LBC's lab. WB performed the initial sequencing data quality filtering,
456 read mapping, identification of translated ORFs, correlations between proteomics and
457 sequencing data. TT performed the differential gene expression and translational
458 efficiency analyses as well as GO terms enrichment. SGM performed the subsampling
459 analyses and correlations between different sets of sequencing data. BBM carried out the
460 ribosome profiling protocol in JD's lab. ACM performed the multidimensional scaling
461 analysis. TT, WRB and MMA wrote the manuscript.
462

463 **Competing interests**

464 The authors declare no competing interests.
465

466 **Data availability**

467 Supplementary data files have been uploaded to Figshare and can be accessed at
468 http://dx.doi.org/10.6084/m9.figshare.5809812. This includes the transcriptome genomic
469 coordinates, the gene table of counts, lisst of differentially expressed genes and
470 gene/protein abundance estimates derived from RNA-Seq, Ribo-Seq and proteomics. The
471 original sequencing data is at https://www.ncbi.nlm.nih.gov/bioproject/435567 (NCBI
472 Bioproject PRJNA435567).
473

474 **Supplementary information**

475 The supplementary file contains the supplementary tables and figures mentioned in the
476 text.
477

478 **Figure legends**

479

480 **Figure 1. Experimental design.** Baker's yeast (*S. cerevisiae*) was grown in rich media
481 and oxidative stress conditions in parallel. The cultures were used to extract total RNA,
482 ribosome-protected RNA fragments and proteins.

483

484 **Figure 2. Representative gene expression correlations between RNA sequencing**
485 **samples. A.** RNA-Seq normal replicate 1 *versus* Ribo-Seq normal replicate 1. **B.** RNA-
486 Seq stress replicate 1 *versus* Ribo-Seq stress replicate 1. **C.** RNA-Seq normal replicate 1
487 *versus* RNA-Seq normal replicate 2. **D.** Ribo-Seq normal replicate 1 *versus* Ribo-Seq
488 normal replicate 2. Expression units are CPM in logarithm scale; R: Spearman correlation
489 value. N: normal growth conditions (two replicates N1 and N2); S: stress conditions (two
490 replicates S1 and S2).

491

492 **Figure 3. Proteomics shows a stronger correlation with Ribo-Seq than with RNA-**
493 **Seq data. A.** RNA-Seq *versus* proteomics, normal growth conditions. **B.** RNA-Seq *versus*
494 proteomics, oxidative stress. **C.** Ribo-Seq *versus* proteomics, normal growth conditions.
495 **D.** Ribo-Seq *versus* proteomics, oxidative stress. CPM: counts per million for RNA-Seq
496 and RNA-Seq data (represented in logarithmic scale, average between replicates). $\log_2$
497 normalized area: relative abundance for proteomics data (average between replicates). R:
498 Spearman correlation value. Plot and correlations comprise 2200 genes for which $\geq$3
499 unique peptides were detected by LCMSMS.

500

501 **Figure 4. Integrated analysis of RNA sequencing and ribosome profiling data. A.**
502 Distribution of gene expression fold change (FC) values. FC was calculated as the ratio
503 between the number of reads in oxidative stress and normal conditions. We took the
504 average number of reads per gene among the replicates. The standard deviation of $\log_2$FC
505 was 0.44 for Ribo-Seq (RP) and 0.57 for RNA-Seq (RNA). **B.** Multidimensional scaling
506 (MDS) plot using the gene expression values of each sample. MDS was based on the
507 $\log_2$CPM values for each gene. Data was for 5,419 S. cerevisiae genes. RP: Ribo-Seq data;
508 RNA: RNA-Seq data; N: normal growth conditions; S: stress conditions. Two sequencing
509 replicates were generated per condition. **C.** Correlation between log fold change (FC)
510 gene expression values. The X axis corresponds to the RNA-Seq data, or transcriptome,

511 the Y axis to the Ribo-Seq data, or translatome. Coloured dots correspond to differentially
512 expressed genes. In the legend homodirectional means up-regulated, or down-regulated,
513 both at the transcriptome and translatome levels; opposite_change is up-regulated at one
514 level and down-regulated at the other one; translatome means significant differences in
515 Ribo-Seq only; transcriptome means significant differences in RNA-Seq only. **D.**
516 Significant gene functional classes among differentially expressed genes. Shown is a 2-
517 D plot of the enrichment score values, in logarithmic scale, provided by the software
518 DAVID for differentially expressed genes using RNA-Seq (transcriptome) or Ribo-Seq
519 (translatome) data. Significant enrichment scores are associated with a p-val < 0.05.
520 Functional classes associated with positive values are significantly enriched among up-
521 regulated genes, and functional classes with negative values are significantly enriched
522 among down-regulated genes. Non-significant enrichment scores are given a value of 0
523 in the plot.
524

525 ## References

526

527 1. Rapaport, F. *et al.* Comprehensive evaluation of differential gene expression
528 analysis methods for RNA-seq data. *Genome Biol.* **14,** R95 (2013).
529 2. de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global
530 signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5,** 1512–26
531 (2009).
532 3. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression
533 control. *Nature* **473,** 337–342 (2011).
534 4. Payne, S. H. The utility of protein and mRNA correlation. *Trends Biochem. Sci.*
535 **40,** 1–3 (2015).
536 5. Ponnala, L., Wang, Y., Sun, Q. & van Wijk, K. J. Correlation of mRNA and
537 protein abundance in the developing maize leaf. *Plant J.* **78,** 424–440 (2014).
538 6. Lee, M. V. *et al.* A dynamic model of proteome changes reveals new roles for
539 transcript alteration in yeast. *Mol. Syst. Biol.* **7,** 514–514 (2014).
540 7. Tebaldi, T. *et al.* Widespread uncoupling between transcriptome and translatome
541 variations after a stimulus in mammalian cells. *BMC Genomics* **13,** 220 (2012).
542 8. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S.
543 Genome-wide analysis in vivo of translation with nucleotide resolution using

544 ribosome profiling. *Science* **324,** 218–23 (2009).
545 9. Gerashchenko, M. V., Lobanov, A. V. & Gladyshev, V. N. Genome-wide
546 ribosome profiling reveals complex translational regulation in response to
547 oxidative stress. *Proc. Natl. Acad. Sci.* **109,** 17394–17399 (2012).
548 10. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and
549 pseudogenes are translated and some are likely to express functional proteins.
550 *Elife* **4,** e08890 (2015).
551 11. Mcmanus, C. J., May, G. E., Spealman, P. & Shteyman, A. Ribosome profiling
552 reveals post-transcriptional buffering of divergent gene expression in yeast. 422–
553 430 (2014).
554 12. Borg, I. & Groenen, P. J. F. *Modern multidimensional scaling*. (Springer, 1997).
555 13. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock
556 linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15,** R29
557 (2014).
558 14. Morano, K. A., Grant, C. M. & Moye-Rowley, W. S. The response to heat shock
559 and oxidative stress in Saccharomyces cerevisiae. *Genetics* **190,** 1157–95 (2012).
560 15. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative
561 analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*
562 **4,** 44–57 (2009).
563 16. Zhong, Y. *et al.* RiboDiff: detecting changes of mRNA translation efficiency
564 from ribosome footprints. *Bioinformatics* **33,** 139–141 (2017).
565 17. Ingolia, N. T. Ribosome Footprint Profiling of Translation throughout the
566 Genome. *Cell* **165,** 22–33 (2016).
567 18. Jungfleisch, J. *et al.* A novel translational control mechanism involving RNA
568 structures within coding sequences. *Genome Res.* **27,** 95–106 (2017).
569 19. Yordanova, M. M. *et al.* AMD1 mRNA employs ribosome stalling as a
570 mechanism for molecular memory formation. *Nature* **553,** 356–360 (2018).
571 20. Aspden, J. L. *et al.* Extensive translation of small ORFs revealed by Poly-Ribo-
572 Seq. *Elife* e03528 (2014).
573 21. Ruiz-Orera, J., Messeguer, X., Subirana, J. A. & Alba, M. M. Long non-coding
574 RNAs as a source of new peptides. *Elife* **3,** e03523 (2014).
575 22. Raj, A. *et al.* Thousands of novel translated open reading frames in humans
576 inferred by ribosome footprint profiling. *Elife* **5,** (2016).
577 23. Ruiz-Orera, J., Verdaguer-Grau, P., Villanueva-Cañas, J. L., Messeguer, X. &

578 Albà, M. M. Translation of neutrally evolving peptides provides a basis for de
579 novo gene evolution. *Nat. Ecol. Evol.* **2**, 890–896 (2018).
580 24. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute
581 quantification of proteins and phosphoproteins from cell lysates by tandem MS.
582 *Proc. Natl. Acad. Sci.* **100,** 6940–6945 (2003).
583 25. Edfors, F. *et al.* Gene-specific correlation of RNA and protein levels in human
584 cells and tissues. *Mol. Syst. Biol.* **12,** 883 (2016).
585 26. Slavoff, S. A. *et al.* Peptidomic discovery of short open reading frame-encoded
586 peptides in human cells. *Nat. Chem. Biol.* **9,** 59–64 (2013).
587 27. Stadler, M. & Fire, A. Conserved translatome remodeling in nematode species
588 executing a shared developmental transition. *Plos Genet.* **9,** e1003739 (2013).
589 28. Zid, B. M. & O'Shea, E. K. Promoter sequences direct cytoplasmic localization
590 and translation of mRNAs during starvation in yeast. *Nature* **514,** 117–121
591 (2014).
592 29. Khong, A. *et al.* The Stress Granule Transcriptome Reveals Principles of mRNA
593 Accumulation in Stress Granules. *Mol. Cell* **68,** 808–820.e5 (2017).
594 30. Luo, Y., Na, Z. & Slavoff, S. A. P-Bodies: Composition, Properties, and
595 Functions. *Biochemistry* **57,** 2424–2431 (2018).
596 31. Reid, D. W., Shenolikar, S. & Nicchitta, C. V. Simple and inexpensive ribosome
597 profiling analysis of mRNA translation. *Methods* **91,** 69–74 (2015).
598 32. Xie, S.-Q. *et al.* RPFdb: a database for genome wide information of translated
599 mRNA generated from ribosome profiling. *Nucleic Acids Res.* **44,** D254–D258
600 (2016).
601 33. Liu, W., Xiang, L., Zheng, T., Jin, J. & Zhang, G. TranslatomeDB: a
602 comprehensive database and cloud-based analysis platform for translatome
603 sequencing data. *Nucleic Acids Res.* **46,** D206–D212 (2018).
604 34. Tsankov, A. M., Thompson, D. A., Socha, A., Regev, A. & Rando, O. J. The
605 Role of Nucleosome Positioning in the Evolution of Gene Regulation. *PLoS Biol.*
606 **8,** e1000414 (2010).
607 35. Ingolia, N. T., Brar, G. a, Rouskin, S., McGeachy, A. M. & Weissman, J. S. The
608 ribosome profiling strategy for monitoring translation in vivo by deep sequencing
609 of ribosome-protected mRNA fragments. *Nat. Protoc.* **7,** 1534–50 (2012).
610 36. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
611 Illumina sequence data. *Bioinformatics* **30,** 2114–20 (2014).

612 37. Martin, M. Cutadapt removes adapter sequences from high-throughput
613 sequencing reads. *EMBnet.journal* **17.1,** (2011).
614 38. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-
615 efficient alignment of short DNA sequences to the human genome. *Genome Biol.*
616 **10,** R25 (2009).
617 39. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
618 **25,** 2078–2079 (2009).
619 40. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with
620 high-throughput sequencing data. *Bioinformatics* **31,** 166–9 (2015).
621 41. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor
622 package for differential expression analysis of digital gene expression data.
623 *Bioinformatics* **26,** 139–140 (2010).
624 42. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene
625 Ontology Consortium. *Nat. Genet.* **25,** 25–9 (2000).
626

627 # Tables

	Translatome upregulated	Translatome downregualted	Transcriptome upregulated	Transcriptome downregulated	Other
Increased TE under stress	14	0	0	385	71
Decreased TE under stress	0	208	356	0	150

628

629 **Table 1. Genes with significantly increased or decrease translational efficiency during**

630 **oxidative stress.** TE: gene translational efficiency. Ribodiff p-value < 0.05 and

631 $|\log_2(TE_{stress}/TE_{normal})| > 0.67$. Translatome/Transcriptome defintions as in Figure 5.

632

633

634 **Figures**

635

636 **Figure 1**

637

638

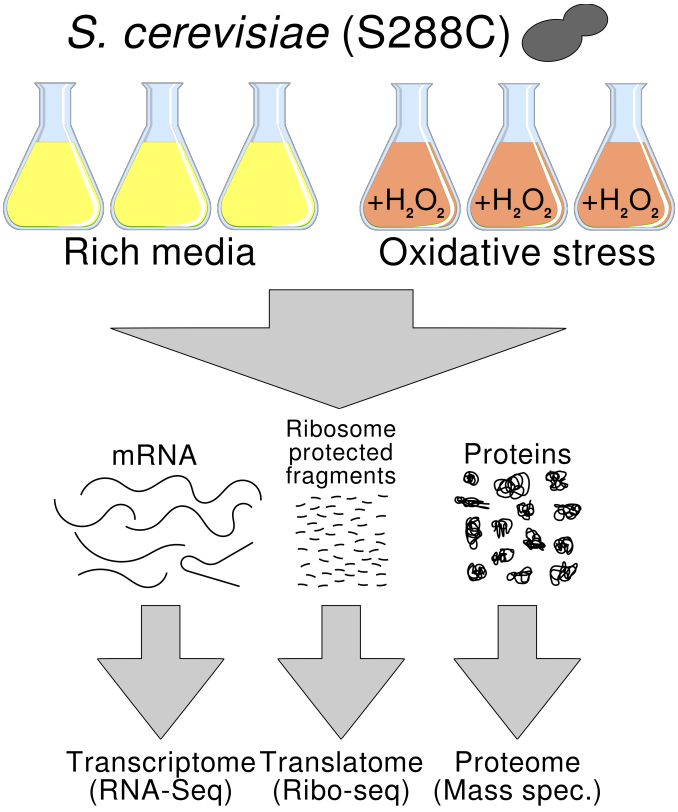


Figure 2

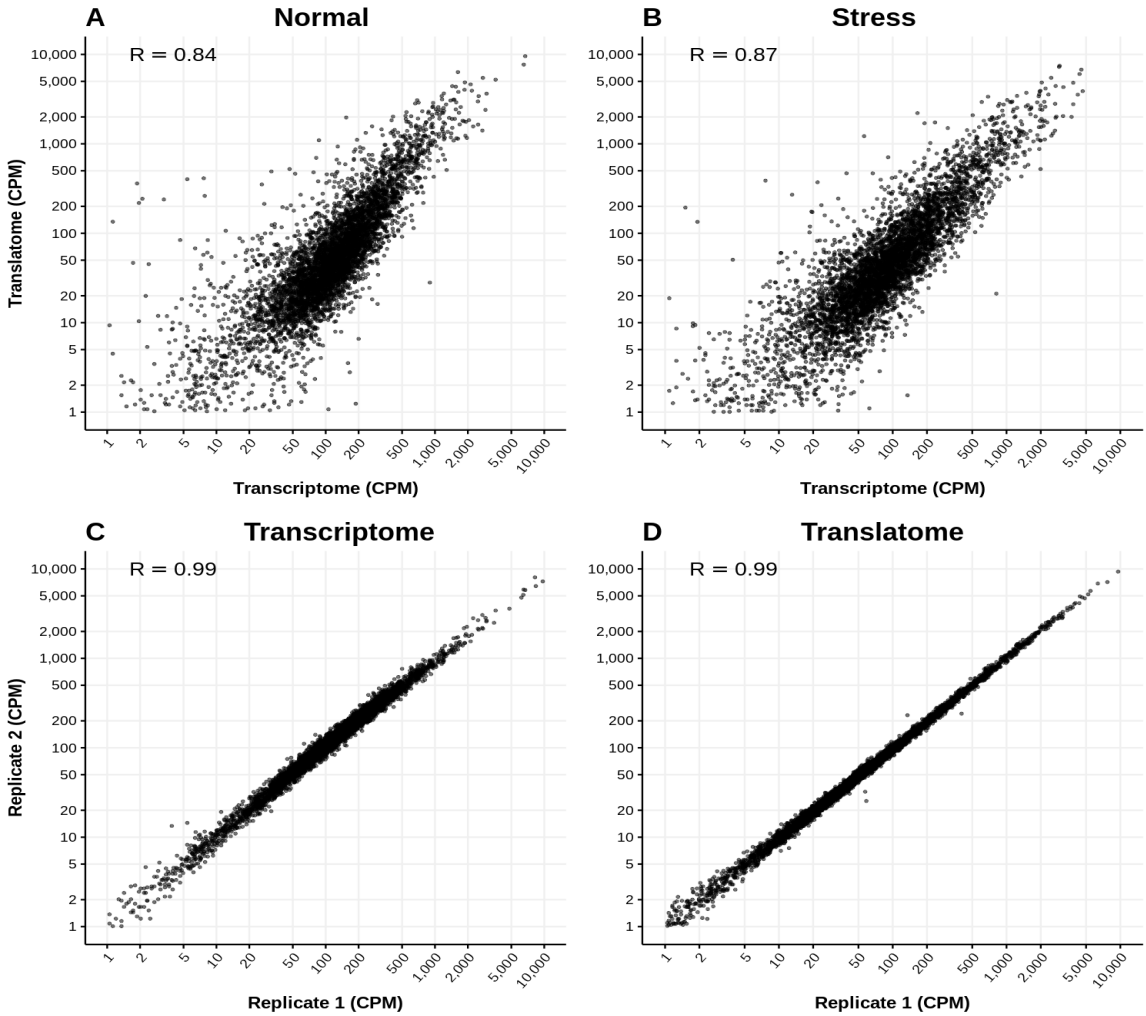
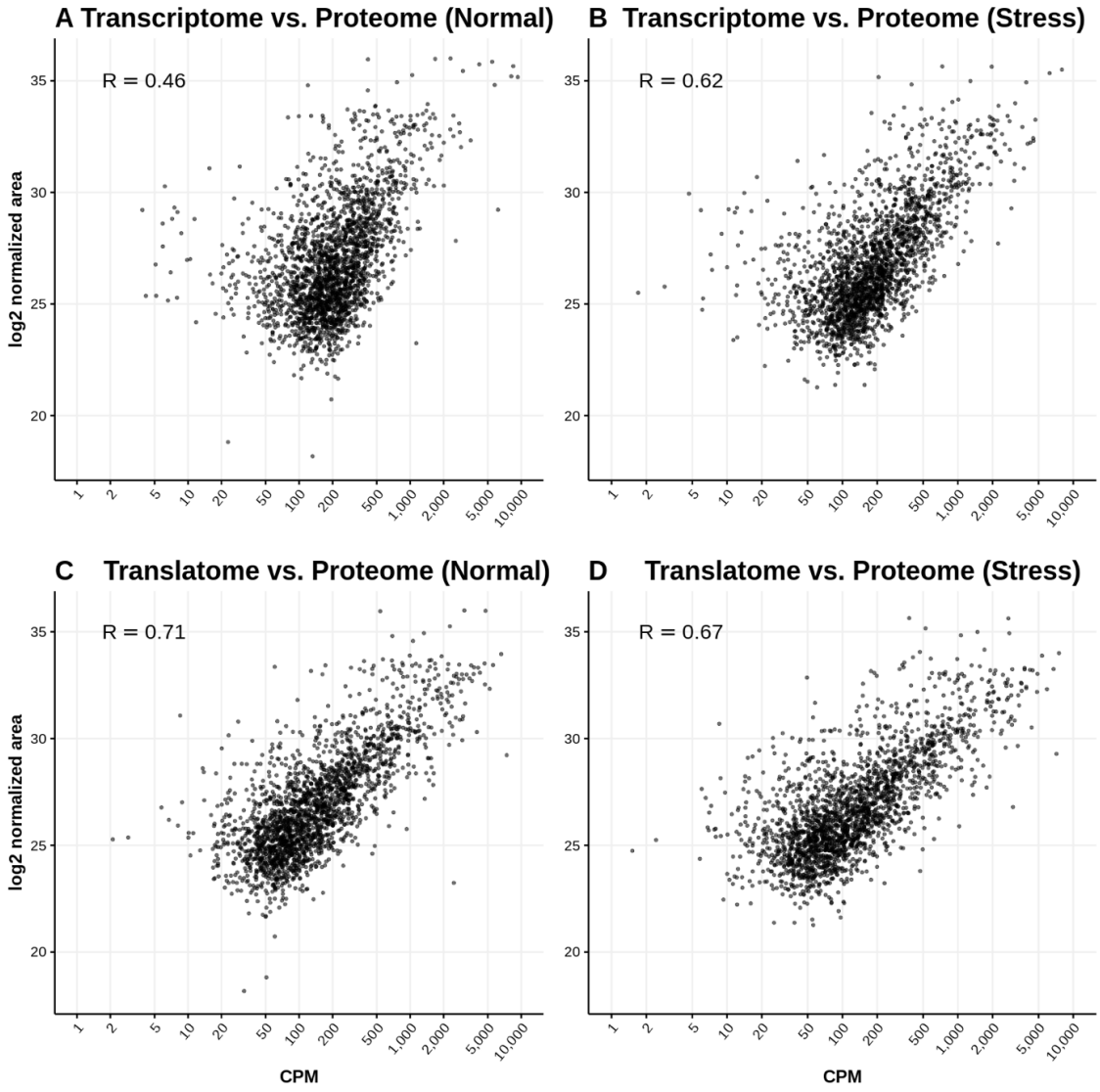
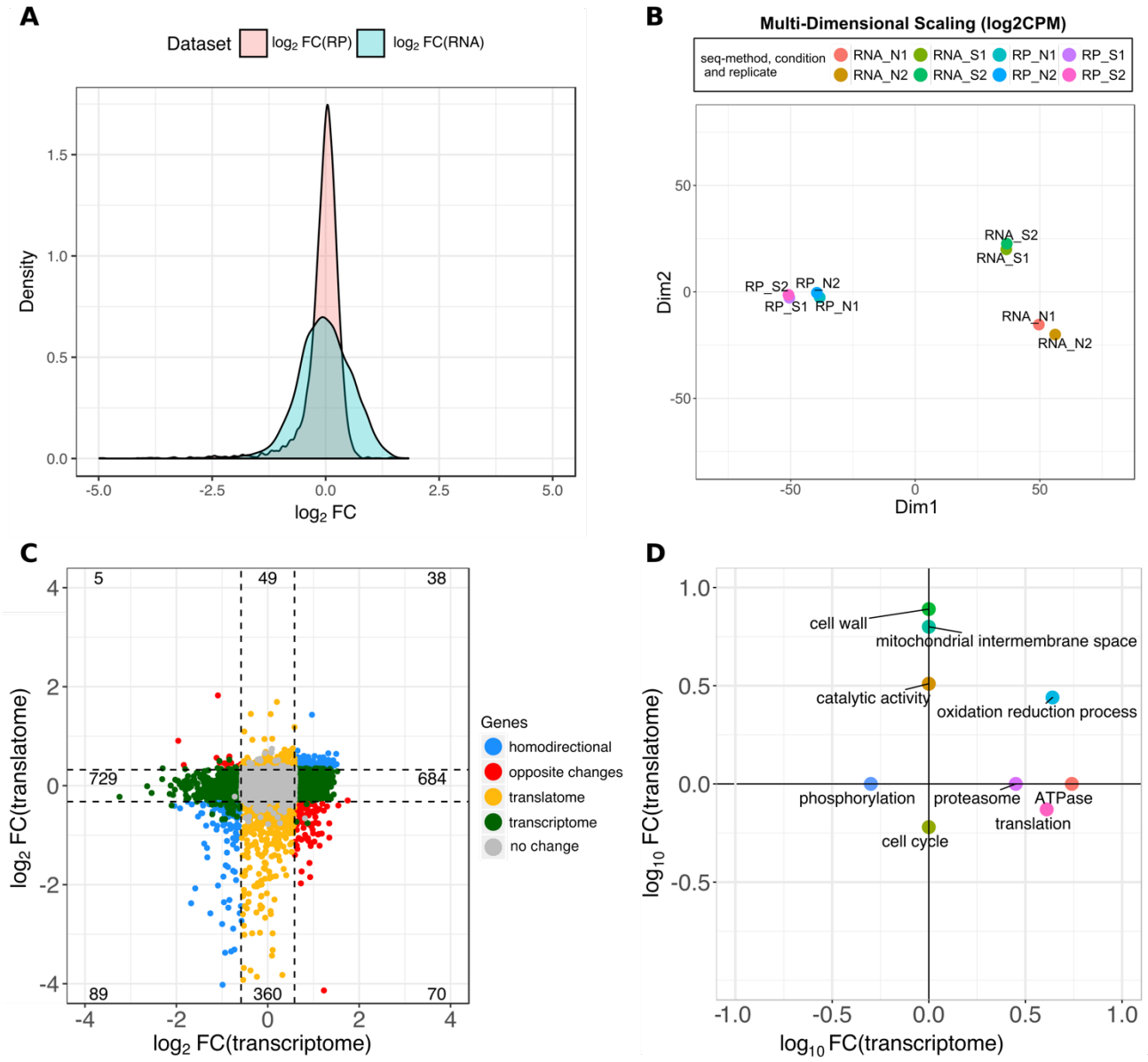

A
Normal
R = 0.84
Translatome (CPM)
Transcriptome (CPM)
B
Stress
R = 0.87
Transcriptome (CPM)
C
Transcriptome
R = 0.99
Replicate 2 (CPM)
Replicate 1 (CPM)
D
Translatome
R = 0.99
Replicate 1 (CPM)

Figure 3


A Transcriptome vs. Proteome (Normal)
R = 0.46
B Transcriptome vs. Proteome (Stress)
R = 0.62
C Translatome vs. Proteome (Normal)
R = 0.71
D Translatome vs. Proteome (Stress)
R = 0.67
log2 normalized area
CPM
35
30
25
20
1
2
5
10
20
50
100
200
500
1,000
2,000
5,000
10,000

646 **Figure 4**

647



648