Using ribosome profiling to quantify differences in protein expression: a case study in *Saccharomyces cerevisiae* oxidative stress conditions

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], M. Mar Albà[1,2,3,*]

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

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

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

[#]Shared first co-authorship

[*]To whom correspondence should be addressed.

Running title: differential gene translation

Keywords: differential gene expression, ribosome profiling, RNA-Seq, translation, oxidative stress

Abstract

Cells respond to changes in the environment by modifying the concentration of specific proteins. Paradoxically, the cellular response is usually examined by measuring variations in transcript abundance by high throughput RNA sequencing (RNA-Seq), instead of directly measuring protein concentrations. This happens because RNA-Seq-based methods provide better quantitative estimates, and more extensive gene coverage, than proteomics-based ones. However, variations in transcript abundance do not necessarily reflect changes in the corresponding protein abundance. How can we close this gap? Here we explore the use of ribosome profiling (Ribo-Seq) to perform differentially gene expression analysis in a relatively well-characterized system, oxidative stress in baker's yeast. Ribo-Seq is an RNA sequencing method that specifically targets ribosome-protected RNA fragments, and thus is expected to provide a more accurate view of changes at the protein level than classical RNA-Seq. We show that gene quantification by Ribo-Seq is indeed more highly correlated with protein abundance, as measured from mass spectrometry data, than quantification by RNA-Seq. The analysis indicates that, whereas a subset of genes involved in oxidation-reduction processes is detected by both types of data, the majority of the genes that happen to be significant in the RNA-Seq-based analysis are not significant in the Ribo-Seq analysis, suggesting that they do not result in protein level changes. The results illustrate the advantages of Ribo-Seq to make inferences about changes in protein abundance in comparison with RNA-Seq.

Introduction

In recent years high throughput RNA sequencing (RNA-Seq) has become the method of choice for comparing gene expression changes of cells grown under different conditions (Rapaport *et al.*, 2013). The relatively low cost of RNA-Seq, together with the availability of efficient computational methods to process information from millions of sequencing reads, has undoubtedly accelerated our understanding of gene regulation. However, a change in mRNA relative abundance does not always imply a change in the amount of the encoded protein (Schwanhäusser *et al.*, 2011). Filling this gap in understanding is essential to discern the functional changes in the cell upon a given stimulus.

Many studies have shown that mRNA levels only partially explain protein levels in the cell (de Sousa Abreu *et al.*, 2009; Schwanhäusser *et al.*, 2011; Payne, 2015; Ponnala *et al.*, 2014). In yeast, the correlation between mRNA and protein abundance is typically in the range 0.6-0.7 (de Sousa Abreu *et al.*, 2009). In addition, the ratio between protein and mRNA levels may vary across different conditions. For instance, substantial differences in this ratio have been observed during osmotic stress in yeast (Lee et al. 2011) or after the treatment of human cells with epidermal growth factor (Tebaldi *et al.*, 2012). This strongly suggests that measuring changes in mRNA levels may often be insufficient to identify the functional shifts taking place in the cell upon a given stimulus.

Protein quantification is often performed using whole proteome mass spectrometry-based methods (Gerber *et al.*, 2003; Edfors *et al.*, 2016). These methods provide a direct measurement of protein abundance but they also have limitations, especially for the detection of lowly expressed and/or small proteins (Slavoff *et al.*, 2013). An alternative

63 way to estimate protein levels is the sequencing of ribosome-protected mRNA fragments,
64 or ribosome profiling (Ribo-Seq) (Ingolia *et al.*, 2009, 2011; Aspden *et al.*, 2014; Ruiz-
65 Orera *et al.*, 2014). In contrast to RNA-Seq, which measures the total amount of mRNA
66 in the cell, Ribo-Seq only captures those mRNAs that are being actively translated.
67 Although Ribo-Seq measures translation, which is an indirect estimate of protein
68 abundance, it has the advantage over proteomics that virtually any mRNA can be
69 interrogated. In addition, Ribo-Seq reads can be quantified in the same manner as RNA-
70 Seq reads. This implies that we can use the same pipelines as for RNA-Seq to identify
71 differentially expressed genes.
72
73 It has been proposed that alterations in the ratio between the relative number of Ribo-Seq
74 and RNA-Seq reads mapping to a given locus, known as the translation efficiency (TE),
75 can be used to identify putative translation activation or repression events (Ingolia, 2016).
76 Numerous recent studies have used ribosome profiling data has been used to study
77 translation regulatory mechanisms (Jungfleisch *et al.*, 2017; Yordanova *et al.*, 2018) or
78 to discover new translated RNA sequences (Michel et al. 2012; Aspden et al. 2014;
79 Ingolia et al. 2014; Ruiz-Orera et al. 2014).
80
81 Here we perform differential gene expression analysis using RNA-Seq and Ribo-Seq data
82 during oxidative stress in *Saccharomyces cerevisiae*, a condition that is known to trigger
83 important regulatory changes both at the transcriptional and translational levels (Shenton
84 *et al.*, 2006; Gerashchenko *et al.*, 2012). We compare the results to proteomics data
85 obtained from the same samples. The results show that the dynamics of total mRNA and
86 translated mRNAs are very distinct, and that most changes in the relative amount of
87 mRNA do not appear to have any consequences at the protein level. The study opens a

door for a more generalized use of Ribo-Seq data to measure changes in protein expression across conditions.

Results and Discussion

Quantification of gene expression by Ribo-Seq and RNA-Seq

We extracted ribosome-protected RNA fragments, as well as total polyadenylated RNAs, from *Saccharomyces cerevisiae* grown in rich medium (normal) and in H_2O_2-induced oxidative stress conditions (stress). We then sequenced ribosome-protected RNAs (Ribo-Seq) as well as complete polyA+ mRNAs (RNA-Seq) using a strand-specific protocol. The Ribo-Seq data corresponded to the translated mRNA fraction (translatome), whereas the RNA-Seq data corresponded to total mRNAs (transcriptome). For comparison we also estimated protein concentrations (proteome) in the two conditions by mass spectrometry (Figure 1).

After quality control of the sequencing reads we obtained 31-36 million reads for Ribo-Seq and 12-15 million reads for RNA-Seq (Supplementary Table S1). We mapped the reads to the genome and generated a table of gene counts for each of the samples. After filtering out non-expressed genes (see Methods), the table contained data for 5,419 *S. cerevisiae* genes. Using mass spectrometry (mass spec) we could quantify the protein products of 2,200 genes (see Methods), representing about 40% of the genes quantified by RNA-Seq.

We normalized the RNA-Seq and Ribo-Seq-based table of counts by calculating counts per million (CPM) in logarithmic scale, or $\log_2$CPM (Supplementary Figure S1). The gene normalized expression values showed a very high correlation between biological replicates, with a correlation coefficient large than 0.99 between all pairs of Ribo-Seq or RNA-Seq replicas (Supplementary Table S2). In contrast, normalized protein abundances between pairs of proteomics replicates showed correlation coefficients between 0.83 and 0.93 (Supplementary Table S3), indicating that quantification by proteomics is less reproducible than quantification by RNA-Seq and Ribo-Seq.

Importantly, the Ribo-Seq data correlated better with the proteomics data than RNA-Seq; in the first case the correlation was 0.67-0.71 and in the second one 0.46-0.62 (Figure 3). This supports that notion that Ribo-Seq provides a more accurate view of protein expression than RNA-Seq (Ingolia *et al.*, 2009).

We next clustered the RNA-Seq and Ribo-Seq gene expression values using multidimensional scaling (MDS)(Borg and Groenen, 1997)(Supplementary Figure S2). Remarkably, the Ribo-Seq measurements for the two conditions (normal and stress) were more similar to each other than any of them was to the condition-matched RNA-Seq measurements, and the same thing happened with the RNA-Seq-based measurements. Thus, the sequencing approach employed is expected to have a strong impact in the results.

Next, we calculated the fold change (FC) gene expression difference between conditions, taking the average expression values between replicates of the same experimental condition. In agreement with the results obtained with MDS, the $\log_2$FC distribution based on the Ribo-Seq data had a lower variance than the $\log_2$FC distribution using RNA-

Seq data (Figure 4). We considered the possibility that this pattern was due to the number of Ribo-Seq reads being 2-3 times larger than the number of RNA-Seq reads (Supplementary Table S1). To test for this, we subsampled the mapped reads so as to have a similar number of reads in all the RNA-Seq and Ribo-Seq samples (Supplementary Tables S4 and S5). We again observed a lower $\log_2$FC variance for Ribo-Seq than for RNA-Seq (Supplementary Figure S3), indicating that the observed variance difference has a biological origin.

Differential gene expression analysis

We performed differential gene expression analysis, separately for Ribo-Seq and RNA-Seq data, using multivariable linear regression with the Limma package (Law *et al.*, 2014). Limma provides a list of differentially expressed genes with the corresponding adjusted p-values. We selected genes with an adjusted p-value < 0.05 and a $\log_2$FC larger than one standard deviation; the latter corresponded to a minimum FC of 1.49 for RNA-Seq data and 1.36 for Ribo-Seq data. We used the standard deviation instead of a fixed value to accommodate for the differences in the width of the $\log_2$FC distributions (Figure 4).

We obtained 817 up-regulated genes during oxidative stress using RNA-Seq data, compared to only 92 with Ribo-Seq data. Thus, the vast majority of the genes identified as up-regulated in stress with RNA-Seq data were not significantly up-regulated when using the Ribo-Seq data to do the same analysis. The number of down-regulated genes was 846 and 519 for RNA-Seq and Ribo-Seq, respectively. Overall, only a small fraction of the differentially expressed genes was common to both approaches (5-10%, see below).

The induction of oxidative stress by hydrogen peroxide (H_2O_2) results in an excess of reactive oxygen species (ROS) in the cell. This is known to activate the expression of several protein families including thioredoxins, hexoquinases, and heat shock proteins (Morano *et al.*, 2012). The set of up-regulated genes identified by both RNA-Seq and Ribo-Seq included several members of these families (e.g. HXK2, TDH1, CYC1, HSP10), consistent with transcriptional activation of genes directly involved in stress response.

Attempts to use the same pipeline to identify differentially expressed genes using the proteomics data did not yield significant results. The reproducibility of protein abundance estimates using mass spec data is not as high as the reproducibility of gene expression levels in the case of RNA sequencing data, which decreases the power of differential gene expression analysis using this kind of data (Supplementary Table S3).

Uncoupling between changes at the transcriptome and translatome levels

The correlation between RNA-Seq and Ribo-Seq gene $\log_2$FC values was quite low (0.18), supporting an important disconnect between the two kinds of data (Figure 5). We quantified the number of genes that showed a significant change in the same direction i.e. homodirectional changes. There were 38 genes that were up-regulated during stress using both RNA-Seq and Ribo-Seq data, this is a small number but still more than double the number expected by chance (15 genes). The number of homodirectional down-regulated genes was 89, compared to 55 be expected by chance. In summary, while there was a modest overlap between the stories told by RNA-Seq and Ribo-Seq data (test of proportions p-value $< 1.32 \times 10^{5}$), the majority of the differentially expressed genes were not concordant.

187

188 Dissecting differential regulation by functional class

189

190 To better understand the biological relevance of the above results, we investigated if
191 certain functional classes were significantly enriched among the sets of differentially
192 expressed genes. We used DAVID (Huang *et al.*, 2009) to identify significantly over-
193 represented functional clusters (Figure 4). Only one class, 'oxidation-reduction process',
194 was enriched among genes up-regulated during stress both using RNA-Seq and Ribo-Seq
195 data. This is consistent with transcriptional activation of this set of genes upon stress,
196 increasing the signal for both total mRNA and the translated fraction. Three other classes
197 – 'translation', 'ATPase' and 'proteasome' – showed increased mRNA levels during
198 stress, but this was not reflected in an increase in the translated fraction. Thus, it is likely
199 that an important part of these transcripts are stored in a translation inactive form during
200 stress, for example as P-bodies or stress granules (Zid and O'Shea, 2014; Khong *et al.*,
201 2017; Luo *et al.*, 2018). In this case, an accumulation of transcripts would be detected by
202 RNA-Seq but not by Ribo-Seq, as translation of the transcripts is impaired.

203

204 Interestingly, there were functions that only appeared when we performed differential
205 gene expression analysis with the Ribo-Seq data: 'cell wall', 'mitochondrial
206 intermembrane space' and 'catalytic activity' were enriched among up-regulated genes,
207 whereas 'cell cycle' was enriched among down-regulated genes (Figure 6). As these
208 classes are not detected by RNA-Seq, they are candidates to be regulated at the
209 translational level only. An alternative possibility is that the storage of some transcripts
210 in stress granules distorts the RNA-Seq patterns to such a degree that some truly up-

211 regulated genes become undetectable with RNA-Seq; they would only be detected when
212 examining actively translated mRNAs with Ribo-Seq.
213
214 Translation inhibition of cell cycle genes
215
216 In order to further identify possible translational regulatory events we compared the
217 translational efficiency (TE; Ribo-Seq reads divided by RNA-Seq reads) of the different
218 genes in the two conditions using the program Ribodiff (Zhong *et al.*, 2017). This
219 approach is based on the assumption that the number of Ribo-Seq reads is proportional to
220 the amount of translated protein. We detected 470 genes that showed increased TE, and
221 714 genes that showed decreased TE, in oxidative stress *versus* normal growth conditions
222 (adjusted p-value < 0.05; see Methods).
223
224 We reasoned that genes whose translation becomes more active during stress should have
225 increased TE values but also be classified as upregulated when using Ribo-Seq for
226 differential gene expression analysis. We only found 17 genes fulfilling both conditions
227 (3.6% of the genes with increased TE), indicating that activation of translation probably
228 has a relatively small impact in the response to oxidative stress. In the vast majority of
229 cases the increase in TE could be explained by a decrease in RNA-Seq signal during stress
230 (Supplementary Table S6).
231
232 By the same token, genes whose translation is repressed during stress are expected to
233 have decreased TE values but also be classified as down-regulated by Ribo-Seq. We
234 found 246 such genes (34.4% of the genes with decreased TE), suggesting that this
235 mechanism may be more prevalent. Among them there were 12 genes from the cell cycle

functional category (Supplementary Table S7). The putative translational repression of these genes did not appear to be mediated by increased translation of upstream ORFs (Gerashchenko *et al.*, 2012), as we did not detect any increase in the number of Ribo-Seq reads mapping to 5'UTR regions when compared to coding sequences in stress conditions.

Concluding remarks

The adaptation of organisms to variations in the environmental conditions is associated with the activation or repression of the expression of particular genes. These changes are usually studied at the level of complete mRNA molecules using microarrays or next generation sequencing. However, changes in mRNA concentration do not necessarily reflect changes in their encoded protein products; rather, uncoupling between total and polysomal mRNA levels has been observed in many different conditions (Tebaldi *et al.*, 2012; Shenton *et al.*, 2006).

Ribo-Seq specifically targets ribosome-protected mRNAs, providing a closer view to protein expression than RNA-Seq, which is for total mRNA sequences. Although Ribo-Seq data is more labour-intensive than RNA-Seq, the protocols are being simplified and its use is rapidly growing (Reid et al. 2015; Xie et al. 2016; Liu et al. 2018; Michel et al. 2018). Here we have used Ribo-Seq data to perform differential gene expression analysis during oxidative stress, and compared the results to RNA-Seq and to proteomics data.

We have shown that gene expression levels inferred from Ribo-Seq data correlate better with protein abundance than those inferred from RNA-Seq data. Remarkably, many of the genes that are classified as differentially regulated using RNA-Seq do not show a

similar effect when the Ribo-Seq data is analyzed, strongly suggesting that, for these genes, no significant changes at the protein level take place. The methodological framework we have developed here can be applied to other conditions and help advance our understanding of gene regulation.

Methods

Biological material

We grew *S. cerevisiae* (S288C) in 500 ml of rich medium (Tsankov *et al.*, 2010). In order to induce oxidative stress, 30 minutes before harvesting we added diluted H_2O_2 to the medium for a final concentration of 1.5 mM. The cells were harvested in log growth phase (OD600 of ~0.25) via vacuum filtration and frozen with liquid nitrogen.

Ribosome profiling

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

Cells were lysed using the freezer/mill method (SPEX SamplePrep); after preliminary preparations, lysates were treated with RNaseI (Ambion), and subsequently with SUPERaseIn (Ambion). Monosomal fractions were collected; SDS was added to stop any

286 possible RNAse activity, then samples were flash-frozen with N_2(l). Digested extracts
287 were loaded in 7%-47% sucrose gradients. RNA was isolated from monosomal fractions
288 using the hot acid phenol method. Ribosome-Protected Fragments (RPFs) were selected
289 by isolating RNA fragments of 28-32 nucleotides (nt) using gel electrophoresis. The
290 preparation of sequencing libraries for Ribo-Seq and RNA-Seq was based on a previously
291 described protocol (Ingolia *et al.*, 2012). Pair-end sequencing reads of size 35 nucleotides
292 (2x35bp) were produced for Ribo-Seq and RNA-Seq on MiSeq and NextSeq platforms,
293 respectively. The data has been deposited at NCBI Bioproject PRJNA435567
294 (https://www.ncbi.nlm.nih.gov/bioproject/435567).
295
296 Processing of the sequencing data
297
298 The RNA-Seq data was filtered using Trimmomatic with default parameters (version
299 0.36)(Bolger *et al.*, 2014). In the Ribo-Seq data we discarded the second read pair as it
300 was redundant and of poorer quality than the first read, and then used Cutadapt (Martin,
301 2011) to eliminate the adapters and to trim five and four nucleotides at 5' and 3' edges,
302 respectively. Ribosomal RNA was depleted from the Ribo-Seq data *in silico* by removing
303 all reads which mapped to annotated rRNAs. Ribo-Seq reads shorter than 25 nucleotides
304 were not used.
305
306 After quality check and read trimming, the reads were aligned against the *S. cerevisiae*
307 genome (S288C R64-2-1) using Bowtie 2 (Langmead *et al.*, 2009). For annotation we
308 used a previously generated *S. cerevisiae* transcriptome containing 6,184 annotated
309 coding sequences plus 1,009 non-annotated assembled transcripts (see Supplementary
310 data). SAMtools (Li *et al.*, 2009) was used to filter out unmapped reads.

We counted the number of reads that mapped to each gene with HTSeq-count (Anders *et al.*, 2015). We used the mode 'intersection strict' to generate a table of counts from the data; the procedure removed about 5% of the reads in the case of RNA-Seq, and 8% in the case of Ribo-Seq. Only genes in which the average read count of the two replicates was larger than 10 in all conditions (normal and stress, for RNA-Seq and for Ribo-Seq) were kept. The filtered table of counts contained data for 5,419 genes.

For subsampling the number of mapped reads we used SAMtools (Li *et al.*, 2009). We used the function 'samtools view' with option '-s 0.X', where X is the percentage of reads that we wish to keep.

Differential gene expression analysis

The table of counts was normalized to $\log_2$ Counts per Million ($\log_2$CPM), in order to account for the different number of total reads in each sample. Before performing differential gene expression analysis, we normalized the data using Trimmed Mean of M-values (TMM) as implemented is the package edgeR (Robinson *et al.*, 2010). Finally, we applied the Limma voom method (Law *et al.*, 2014) to identify differentially expressed genes, separately for RNA-Seq and Ribo-Seq data (adjusted p-value < 0.05 and $|\log_2 FC| >$ 1 SD($\log_2$FC)) .

We also performed the same kind of analysis for the proteomics data. We used genes which had at least 3 unique peptides and could be quantified in all 6 replicates (1,580

genes); the procedure did not identify any significantly up or down regulated genes, using an adjusted p-value < 0.05.

Quantification of protein abundance by mass spectrometry

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

Analysis of functional clusters

We identified significantly enriched functional clusters in differentially expressed genes using DAVID (Huang *et al.*, 2009). The analysis was done separately for over- and under-expressed genes and for RNA-Seq and Ribo-Seq derived data. Only clusters with

enrichment score ≥ 1.5 and adjusted p-val < 0.05 were retained. In each cluster we chose a representative Gene Ontology (GO) term (Ashburner *et al.*, 2000), with the highest number of genes inside the cluster. Figure 4 integrates the results obtained with the Ribo-Seq and the RNA-Seq data, the $\log_{10}$ fold enrichment of the significant GO terms is plotted.

Analysis of translational efficiency

We searched for genes with significantly increased or decreased translational efficiency (TE)(Ingolia et al., 2009) using the RiboDiff program (Zhong et al., 2017). We selected genes significant at an adjusted p-value < 0.05 and showing $\log_2(TE_{stress}/TE_{normal})$ higher than 0.67 or lower than -0.67 (plus or minus the standard deviation of the distribution).

We downloaded *S.cerevisiae* 5'UTR sequences from the Yeast Genome Database (https://downloads.yeastgenome.org/sequence/S288C_reference/SGD_all_ORFs_5prime_UTRs.fsa). We selected 5'UTR sequences longer than 30 nucleotides, removed identical sequences and took the longest 5'UTR per gene when several existed. The resulting annotation file contained the genomic coordinates of the 5'UTRs of 2,424 genes. We recovered 5'UTR sequences for 5 of the 12 cell cycle-related genes that were potentially repressed at the translational level (HTL1, SPC19, CDC26, BNS1, DIB1). In none of these cases the number of Ribo-Seq reads in the 5'UTR divided by the number of Ribo-Seq reads in the coding sequence increased in oxidative stress with respect to normal growth conditions.

Supplementary data

385

386 Supplementary data files have been uploaded to Figshare and can be accessed at
387 http://dx.doi.org/10.6084/m9.figshare.5809812. This includes the transcriptome genomic
388 coordinates, the filtered table of counts and the list of differentially expressed genes
389 obtained using RNA-Seq and Ribo-Seq. The RNA-Seq and Ribo-Seq original sequencing
390 data is available from https://www.ncbi.nlm.nih.gov/bioproject/435567 (NCBI
391 Bioproject PRJNA435567). Processed data, including annotation files and differentially
392 regulated genes can be found at http://dx.doi.org/10.6084/m9.figshare.5809812.

393

394 Acknowledgements

395

396 We acknowledge the Proteomics Unit of Center for Regulatory Genomics and Universitat
397 Pompeu Fabra for their lab support to isolate proteins from the yeast cultures. We are also
398 grateful to Jorge Ruiz-Orera and Robert Castelo for advice during this project. The work
399 was funded by grants BFU2015-65235-P, BFU2015-68351-P and BFU2016-80039-R,
400 from Ministerio de Economía e Innovación (Spanish Government) - FEDER (EU), and
401 from grant PT17/0009/0014 from Instituto de Salud Carlos III – FEDER. We also
402 received funding from the "Maria de Maeztu" Programme for Units of Excellence in
403 R&D (MDM-2014-0370) and from Agència de Gestió d'Ajuts Universitaris i de Recerca
404 Generalitat de Catalunya (AGAUR), grant number 2014SGR1121, 2014SGR0974,
405 2017SGR01020 and, predoctoral fellowship (FI) to W.B. We also acknowledge support
406 from the EU Erasmus Programme to T.T.

407

408 Figure legends

409

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

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

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

Figure 4. Distribution of gene expression fold change (FC) differences in logarithmic scale. FC was calculated as the ratio between the number of reads in oxidative stress and normal conditions. We took the average number of reads per gene among the replicates. The standard deviation of $\log_2$FC was 0.44 for Ribo-Seq (RP) and 0.57 for RNA-Seq (RNA).

Figure 5. Correlation between gene expression fold changes with RNA-Seq and Ribo-Seq data. Fold change (FC) gene expression values are shown in logarithmic scale. The X axis corresponds to the RNA-Seq data, or transcriptome, the Y axis to the Ribo-Seq data, or translatome. The number of down-regulated and up-regulated genes is indicated. Coloured dots correspond to differentially expressed genes. In the legend

442 homodirectional means up-regulated, or down-regulated, at the transcriptome and
443 translatome levels; opposite_change is up-regulated at one level and down-regulated at
444 the other one.
445
446 **Figure 6. Significant gene functional classes among differentially expressed genes.**
447 Shown is a 2-D plot of the enrichment score values, in logarithmic scale, provided by the
448 software DAVID for differentially expressed genes using RNA-Seq (transcriptome) or
449 Ribo-Seq (translatome) data. Significant enrichment scores are associated with a p-val <
450 0.05. Functional classes associated with positive values are significantly enriched among
451 up-regulated genes, and functional classes with negative values are significantly enriched
452 among down-regulated genes. Non-significant enrichment scores are given a value of 0
453 in the plot.
454
455 ## References
456
457 Anders,S. *et al.* (2015) HTSeq--a Python framework to work with high-throughput
458 sequencing data. *Bioinformatics*, **31**, 166–9.
459 Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene
460 Ontology Consortium. *Nat. Genet.*, **25**, 25–9.
461 Aspden,J.L. *et al.* (2014) Extensive translation of small Open Reading Frames revealed
462 by Poly-Ribo-Seq. *Elife*, **3**, e03528.
463 Bolger,A.M. *et al.* (2014) Trimmomatic: a flexible trimmer for Illumina sequence data.
464 *Bioinformatics*, **30**, 2114–20.
465 Borg,I. and Groenen,P.J.F. (1997) Modern multidimensional scaling Springer.
466 Edfors,F. *et al.* (2016) Gene-specific correlation of RNA and protein levels in human
467 cells and tissues. *Mol. Syst. Biol.*, **12**, 883.
468 Gerashchenko,M. V. *et al.* (2012) Genome-wide ribosome profiling reveals complex
469 translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci.*, **109**,
470 17394–17399.
471 Gerber,S.A. *et al.* (2003) Absolute quantification of proteins and phosphoproteins from
472 cell lysates by tandem MS. *Proc. Natl. Acad. Sci.*, **100**, 6940–6945.
473 Huang,D.W. *et al.* (2009) Systematic and integrative analysis of large gene lists using
474 DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.

475 Ingolia,N.T. *et al.* (2009) Genome-wide analysis in vivo of translation with nucleotide
476 resolution using ribosome profiling. *Science*, **324**, 218–23.
477 Ingolia,N.T. (2016) Ribosome Footprint Profiling of Translation throughout the
478 Genome. *Cell*, **165**, 22–33.
479 Ingolia,N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the
480 complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.
481 Ingolia,N.T. *et al.* (2014) Ribosome Profiling Reveals Pervasive Translation Outside of
482 Annotated Protein-Coding Genes. *Cell Rep.*, **8**, 1365–79.
483 Ingolia,N.T. *et al.* (2012) The ribosome profiling strategy for monitoring translation in
484 vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.*, **7**,
485 1534–50.
486 Jungfleisch,J. *et al.* (2017) A novel translational control mechanism involving RNA
487 structures within coding sequences. *Genome Res.*, **27**, 95–106.
488 Khong,A. *et al.* (2017) The Stress Granule Transcriptome Reveals Principles of mRNA
489 Accumulation in Stress Granules. *Mol. Cell*, **68**, 808–820.e5.
490 Langmead,B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA
491 sequences to the human genome. *Genome Biol.*, **10**, R25.
492 Law,C.W. *et al.* (2014) voom: precision weights unlock linear model analysis tools for
493 RNA-seq read counts. *Genome Biol.*, **15**, R29.
494 Lee,M. V. *et al.* (2014) A dynamic model of proteome changes reveals new roles for
495 transcript alteration in yeast. *Mol. Syst. Biol.*, **7**, 514–514.
496 Li,H. *et al.* (2009) The Sequence Alignment/Map format and SAMtools.
497 *Bioinformatics*, **25**, 2078–2079.
498 Liu,W. *et al.* (2018) TranslatomeDB: a comprehensive database and cloud-based
499 analysis platform for translatome sequencing data. *Nucleic Acids Res.*, **46**, D206–
500 D212.
501 Luo,Y. *et al.* (2018) P-Bodies: Composition, Properties, and Functions. *Biochemistry*,
502 **57**, 2424–2431.
503 Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput
504 sequencing reads. *EMBnet.journal*, **17.1**.
505 Michel,A.M. *et al.* (2018) GWIPS-viz: 2018 update. *Nucleic Acids Res.*, **46**, D823–
506 D830.
507 Michel,A.M. *et al.* (2012) Observation of dually decoded regions of the human genome
508 using ribosome profiling data. *Genome Res.*, **22**, 2219–29.

509 Morano,K.A. *et al.* (2012) The response to heat shock and oxidative stress in
510 Saccharomyces cerevisiae. *Genetics*, **190**, 1157–95.
511 Payne,S.H. (2015) The utility of protein and mRNA correlation. *Trends Biochem. Sci.*,
512 **40**, 1–3.
513 Ponnala,L. *et al.* (2014) Correlation of mRNA and protein abundance in the developing
514 maize leaf. *Plant J.*, **78**, 424–440.
515 Rapaport,F. *et al.* (2013) Comprehensive evaluation of differential gene expression
516 analysis methods for RNA-seq data. *Genome Biol.*, **14**, R95.
517 Reid,D.W. *et al.* (2015) Simple and inexpensive ribosome profiling analysis of mRNA
518 translation. *Methods*, **91**, 69–74.
519 Robinson,M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression
520 analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
521 Ruiz-Orera,J. *et al.* (2014) Long non-coding RNAs as a source of new peptides. *Elife*, **3**,
522 e03523.
523 Schwanhäusser,B. *et al.* (2011) Global quantification of mammalian gene expression
524 control. *Nature*, **473**, 337–342.
525 Shenton,D. *et al.* (2006) Global translational responses to oxidative stress impact upon
526 multiple levels of protein synthesis. *J. Biol. Chem.*, **281**, 29011–21.
527 Slavoff,S.A. *et al.* (2013) Peptidomic discovery of short open reading frame-encoded
528 peptides in human cells. *Nat. Chem. Biol.*, **9**, 59–64.
529 de Sousa Abreu,R. *et al.* (2009) Global signatures of protein and mRNA expression
530 levels. *Mol. Biosyst.*, **5**, 1512–26.
531 Tebaldi,T. *et al.* (2012) Widespread uncoupling between transcriptome and translatome
532 variations after a stimulus in mammalian cells. *BMC Genomics*, **13**, 220.
533 Tsankov,A.M. *et al.* (2010) The Role of Nucleosome Positioning in the Evolution of
534 Gene Regulation. *PLoS Biol.*, **8**, e1000414.
535 Xie,S.-Q. *et al.* (2016) RPFdb: a database for genome wide information of translated
536 mRNA generated from ribosome profiling. *Nucleic Acids Res.*, **44**, D254–D258.
537 Yordanova,M.M. *et al.* (2018) AMD1 mRNA employs ribosome stalling as a
538 mechanism for molecular memory formation. *Nature*, **553**, 356–360.
539 Zhong,Y. *et al.* (2017) RiboDiff: detecting changes of mRNA translation efficiency
540 from ribosome footprints. *Bioinformatics*, **33**, 139–141.
541 Zid,B.M. and O'Shea,E.K. (2014) Promoter sequences direct cytoplasmic localization
542 and translation of mRNAs during starvation in yeast. *Nature*, **514**, 117–121.

543

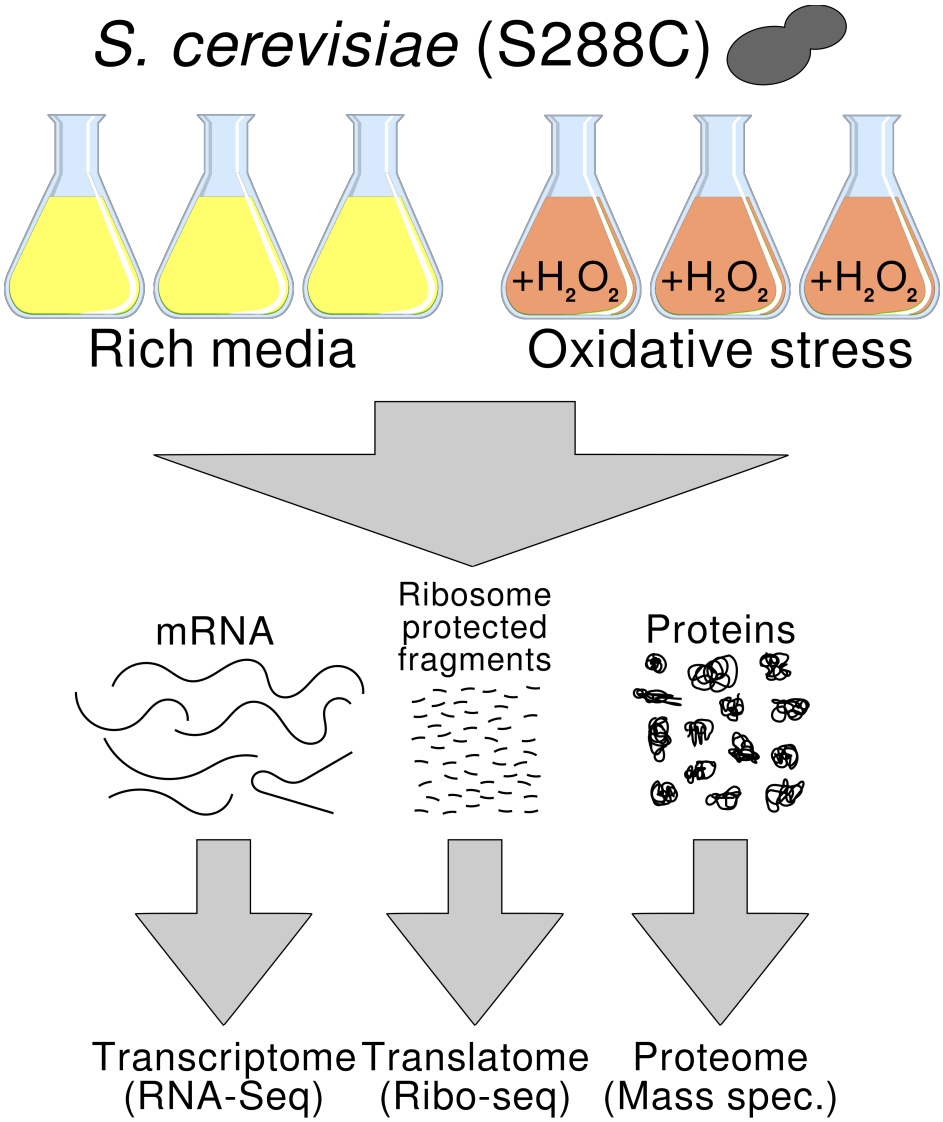
544 **Figures**

545

546 **Figure 1**

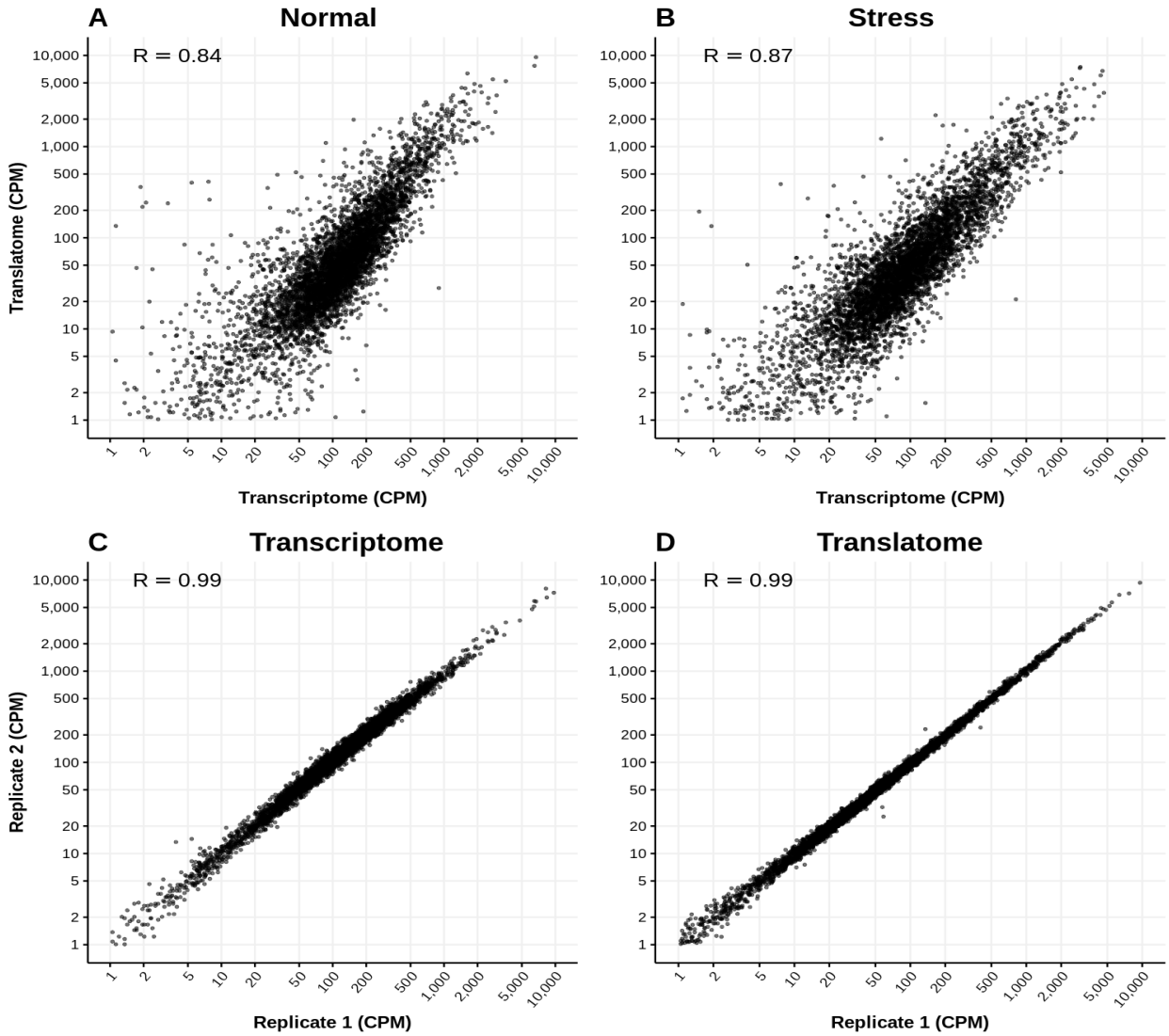
547

548



549 **Figure 2**

550

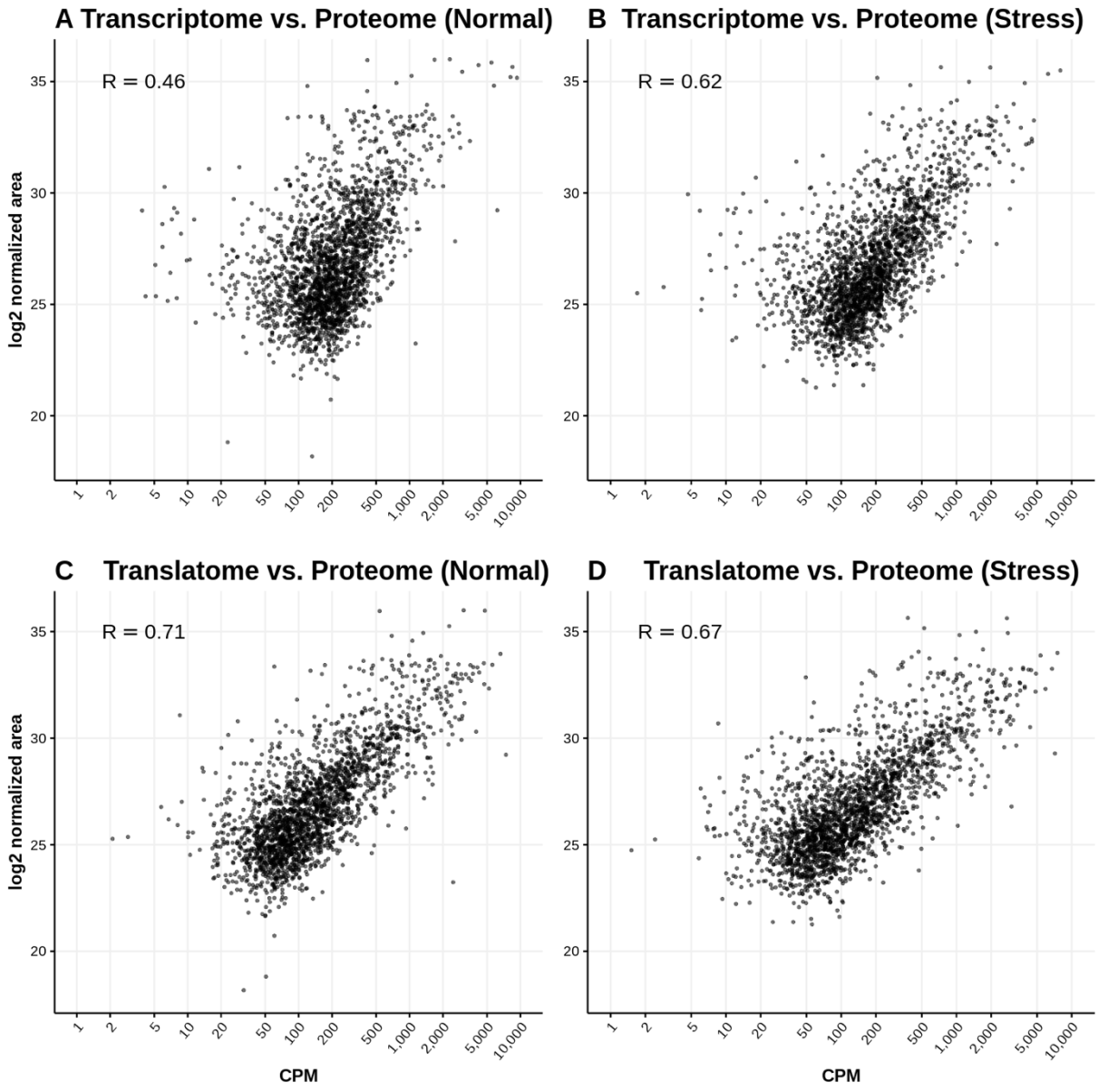


551

552

553 **Figure 3**

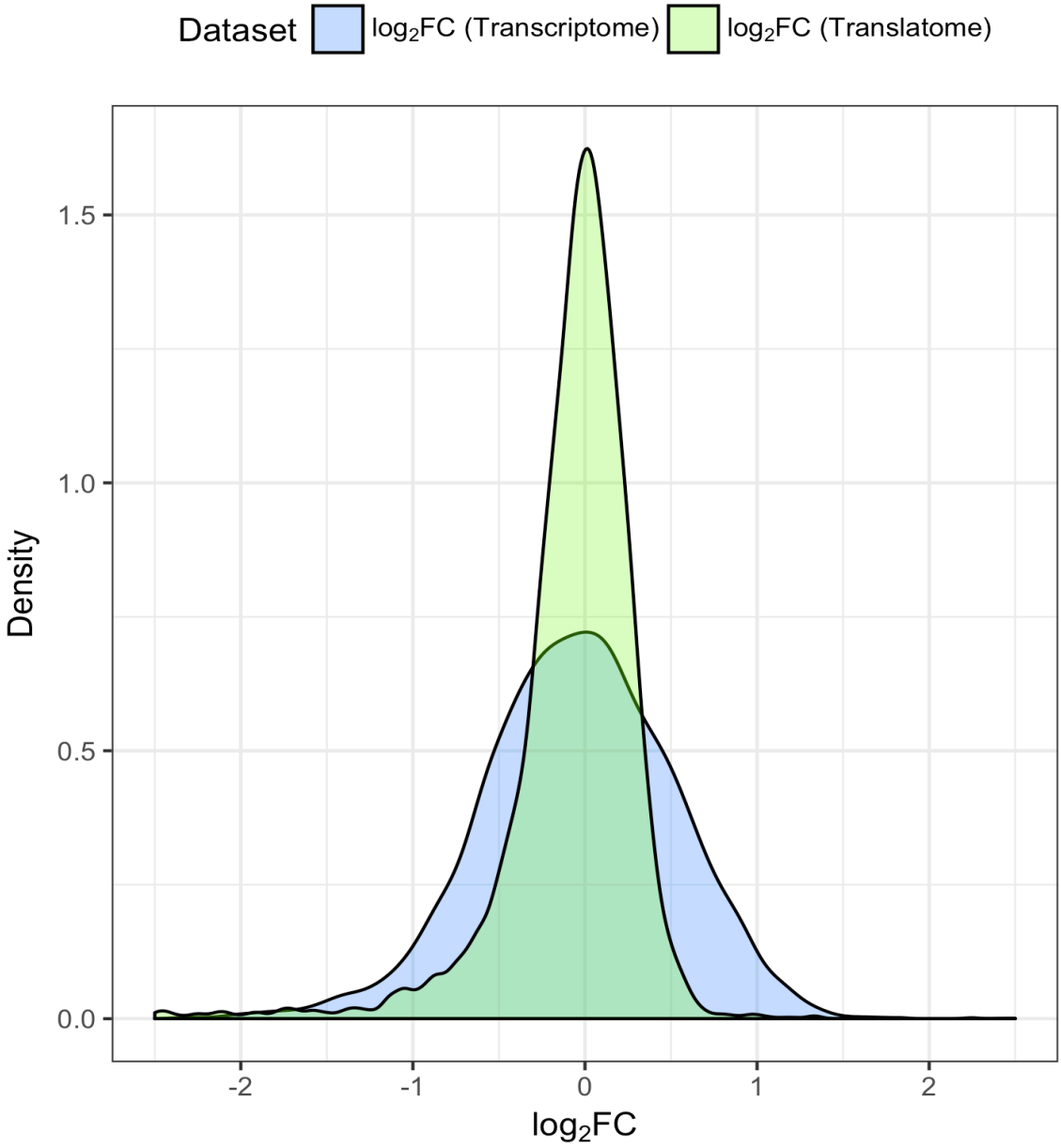
554



555

556 **Figure 4**

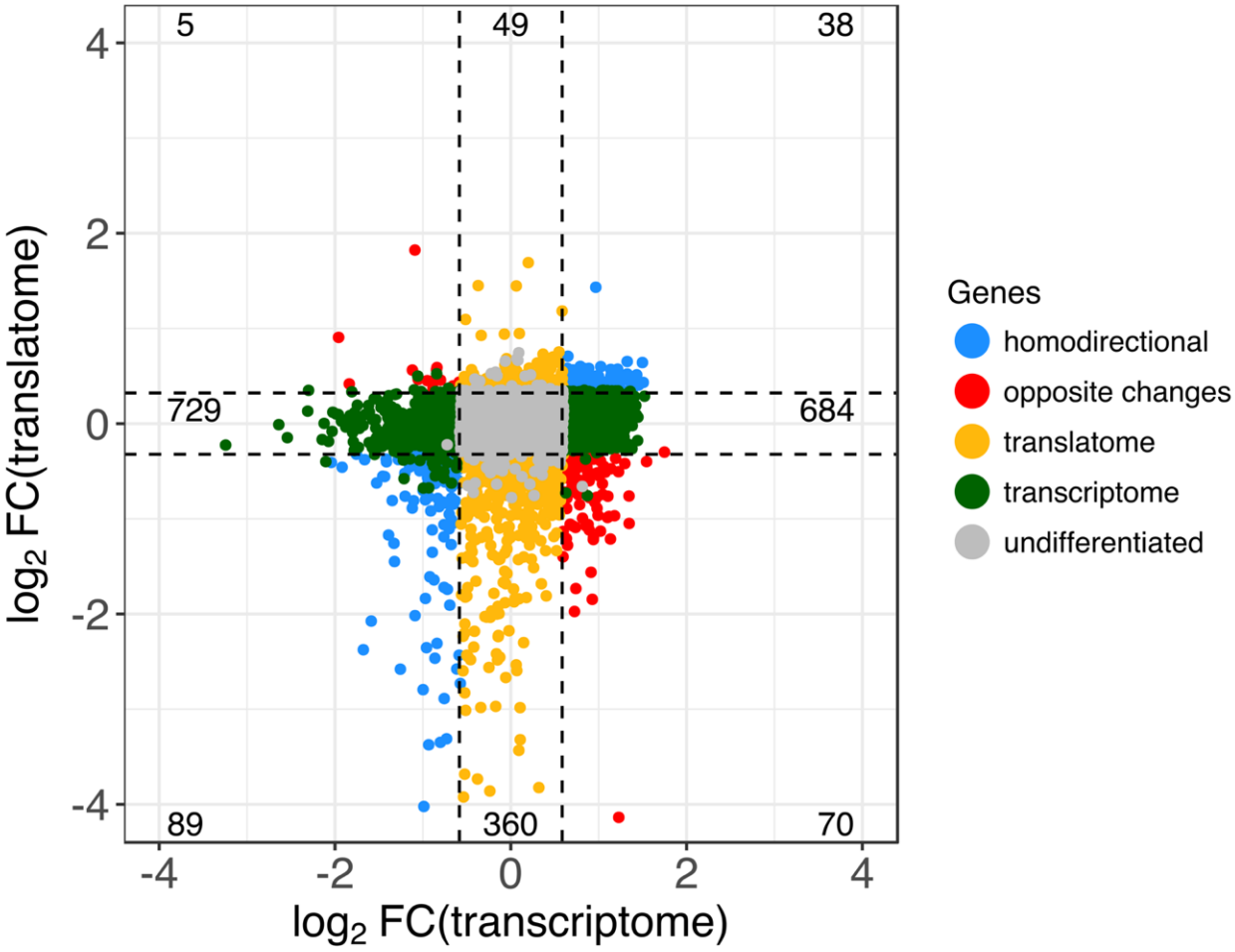
557



558

559 **Figure 5**

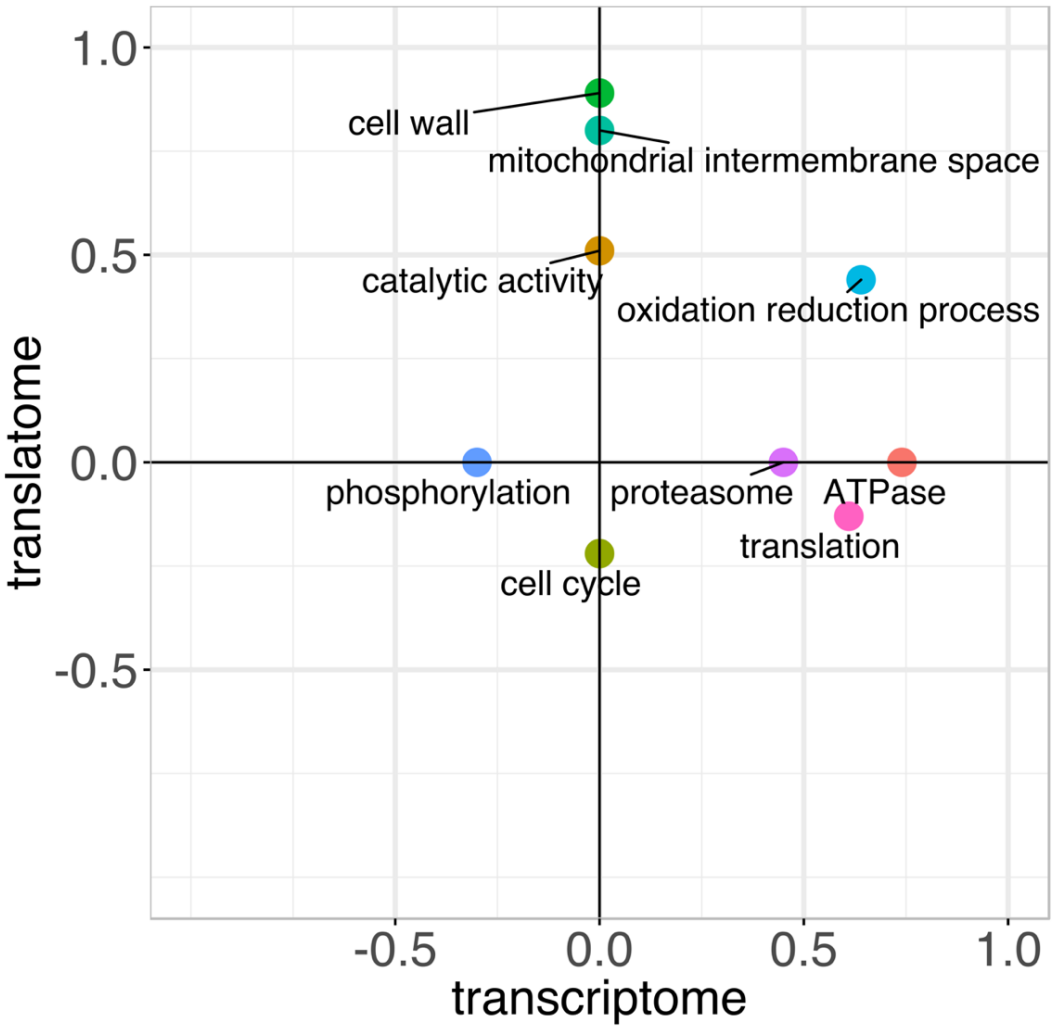
560



561

562 **Figure 6**

563



564