DATA NOTE

Ribo-Seq and RNA-Seq of TMA46 (DFRP1) and GIR2 (DFRP2) knockout yeast strains [version 1; peer review: 3 approved]

Artyom A. Egorov1-3, Desislava S. Makeeva2,3, Nadezhda E. Makarova2-4, Dmitri A. Bykov2,5, Yanislov S. Hrytseniuk2,6, Olga V. Mitkevich7, Valery N. Urakov7, Alexander I. Alexandrov7, Ivan V. Kulakovskiy2,3,8, Sergey E. Dmitriev2-5

1Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology (State University), Dolgoprudny, 141700, Russian Federation
2Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, 119234, Russian Federation
3Sirius University of Science and Technology, Sochi, 354340, Russian Federation
4Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, 119234, Russian Federation
5Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991, Russian Federation
6Department of Chemistry, Lomonosov Moscow State University, Moscow, 119234, Russian Federation
7Bach Institute of Biochemistry, FRC of Biotechnology of the Russian Academy of Sciences, Moscow, 119071, Russian Federation
8Institute of Protein Research, Russian Academy of Sciences, Pushchino, 142290, Russian Federation

Abstract
In eukaryotes, stalled and collided ribosomes are recognized by several conserved multicomponent systems, which either block protein synthesis in situ and resolve the collision locally, or trigger a general stress response. Yeast ribosome-binding GTPases RBG1 (DRG1 in mammals) and RBG2 (DRG2) form two distinct heterodimers with TMA46 (DFRP1) and GIR2 (DFRP2), respectively, both involved in mRNA translation. Accumulated evidence suggests that the dimers play partially redundant roles in elongation processivity and resolution of ribosome stalling and collision events, as well as in the regulation of GCN1-mediated signaling involved in ribosome-associated quality control (RQC). They also genetically interact with SLH1 (ASCC3) helicase, a key component of RQC trigger (RQT) complex disassembling collided ribosomes. Here, we present RNA-Seq and ribosome profiling (Ribo-Seq) data from S. cerevisiae strains with individual deletions of the TMA46 and GIR2 genes. Raw RNA-Seq and Ribo-Seq data as well as gene-level read counts are available in NCBI Gene Expression Omnibus (GEO) repository under GEO accession GSE185458 and GSE185286.

Keywords
Transcriptome, translome, ribosome profiling, ribosome stalling, ribosome collision, Saccharomyces cerevisiae, TMA46, GIR2, GCN1/GCN20, eIF2A, YGR054W, STM1, PUB1

Open Peer Review

Reviewer Status 🟢🟢🟢

Invited Reviewers

1. Anton A. Komar, Cleveland State University, Cleveland, USA
2. Olivier Namy, Université Paris-Saclay, Gif-sur-Yvette, France
3. Nikolay E. Shirokikh, The Australian National University, Canberra, Australia
Katrina Woodward, The Australian National University, Canberra, Australia

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding authors: Ivan V. Kulakovskiy (ivan.kulakovskiy@gmail.com), Sergey E. Dmitriev (sergey.dmitriev@belozersky.msu.ru)

Author roles: Egorov AA: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Makeeva DS: Investigation, Methodology, Validation; Makarova NE: Investigation, Methodology; Bykov DA: Investigation; Hrytseniuk YS: Investigation; Mitkevich OV: Investigation, Methodology; Urakov VN: Investigation; Alexandrov AI: Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; Kulakovskiy IV: Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; Dmitriev SE: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Analyses of the tma46Δ, gir2Δ, stm1Δ, and ygr054wΔ strains were supported by the Russian Foundation for Basic Research (grant 19-34-51047 to S.E.D.) and the Ministry of Science and Higher education of the Russian Federation (to A.I.A, V.N.U, O.V.M). The pub1Δ strain analyses were supported by the Russian Science Foundation (grant 18-14-00291 to S.E.D.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Egorov AA et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Egorov AA, Makeeva DS, Makarova NE et al. Ribo-Seq and RNA-Seq of TMA46 (DFRP1) and GIR2 (DFRP2) knockout yeast strains [version 1; peer review: 3 approved] F1000Research 2021, 10:1162 https://doi.org/10.12688/f1000research.74727.1

First published: 16 Nov 2021, 10:1162 https://doi.org/10.12688/f1000research.74727.1
Introduction

Here, we present Riboseq and RNA-Seq data for *S. cerevisiae* strains lacking translation-associated proteins Tma46 and Gir2, as well as for the wild type BY4742 parent strain. Tma46 and Gir2 are yeast orthologs of two mammalian DRG family regulatory proteins: DFRP1 and DFRP2, respectively.1–5 Mammalian DFRP1 and DFRP2 are binding partners of two closely related proteins, developmentally regulated GTPases DRG1 and DRG2,1,4 while yeast Tma46 and Gir2 dimerize with their orthologs, the ribosome-binding GTPases Rbg1 and Rbg2, respectively.2,6,7 Thus, in both yeast and mammals, two distinct heterodimers exist, RBG1 (DRG1) and RBG2 (DRG2) complexes.10 These results suggest that GIR2 is a physical linker between RBG2 and GCN1 and that this interaction could prevent excessive activation of the GCN2 pathway upon incidental ribosome stalling.

Interestingly, neither the yeast *rbg1Δ* or *rbg2Δ* knockout strains nor the double *rbg1Δrbg2Δ* mutants display any defects in translation or cell growth.1 However, a genetic screen for triple synthetic interaction demonstrates that RBGs have redundant function with SLH1,8 an RNA helicase involved in ribosome-associated quality control (RQC). SLH1 is an ortholog of mammalian ATCC3, a component of the ASC-1 complex that disassembles collided ribosomes (see11 and references therein).

Taken together, the above data suggest that the RBG1*TMA46* (DRG1*DFRP1*) and RBG2*GIR2* (DRG1*DFRP1*) complexes play a role in elongation processivity and resolution of ribosome stalling and collision events, as well as in control of GCN1-mediated signaling accompanying these processes. However, many questions remain unanswered. In particular, the individual roles of the two distinct complexes are still unclear. To improve our understanding of their functions, we systematically characterized translational defects in *S. cerevisiae* strains with individual deletions of the *TMA46* or *GIR2* genes using ribosome profiling.3 We present RNA-Seq and Riboseq data for the yeast *tma46Δ* and *gir2Δ* knockout strains. For comparison, we also provide corresponding data for three strains bearing deletions of other translation-related genes: *STM1*, *PUB1* and *YGR054W* (encoding translation factor eIF2A), as well as for the wild type BY4742 parent strain. Raw sequencing data are available online in the NCBI Gene Expression Omnibus (GEO accession: GSE185458 and GSE185286).

Materials and methods

Yeast strains, cell maintenance, RNA-Seq and Ribo-Seq library preparation

RNA-Seq and Ribo-Seq cDNA libraries were prepared from total RNA samples or ribosome-bound RNA samples, respectively, for the wild-type BY4742 (*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) yeast strain and five knockout strains. The data were obtained in two independent series. The first one included the wild-type BY4742 (wt) and strains with individually deleted *YGR054W* (eIF2A), *STM1*, and *TMA46* genes, which were obtained from the Yeast Gene Deletion Collection.13 The second one included the wild-type BY4742 yeast strain and strains lacking *GIR2* and *PUB1*, which were created during this study. KanMX disruption cassettes were obtained via polymerase chain reaction (PCR) using the primers (GIR2 - 5'-CAATTGTAAAATCAGCAGGCA-3'; 5'-AACTTGTCTATTTCTCTTC-3'; PUB1 - 5'-TTGTCCTTCATTCTCTG-3'; 5'-AGGCCCTTTGTTTTTCGAGC-3') and genomic DNA of the corresponding deletion collection strains as a template. Phusion HF polymerase (TFS F-530L) was used as recommended by the supplier. PCR protocol: 95°C – 5 min, [95°C – 30 sec, 64°C – 30 sec, 72°C – 60 sec] x 28 cycles, 72°C – 90 sec. The resulting cassettes were used to transform the BY4742 strain. The correctness of the integration was checked using pairs of verification primers (GIR2 5'-GAAAAAGAAAGAAGAAAAATTTGGG-3'; PUB1 5'-ACGACCACAAAGGATCCAGGGCCTT-3'; Universal primer inside the KanMX cassette - 5'-CTGCAGCGAGGAGCCGTAAT-3').

Here we focus on wt, *tma46Δ*, and *gir2Δ* strains. The data from the other strains were used to correct for batch effects within each series. The libraries were sequenced, resulting in 31 RNA-Seq and 28 Ribo-Seq data sets, including 18 RNA-Seq and 16 Ribo-Seq data sets for wt, *gir2Δ*, and *tma46Δ* strains. Supplementary Table 1 in the Extended data summarizes information about the sequencing experiments.

The experimental procedure followed the ribosome profiling protocol described in.15 Briefly, yeast cells were grown to exponential phase (OD = 0.5-0.6) in yeast extract peptone dextrose (YPD) media (1% yeast extract, 2% peptone, 2% glucose). Cells were harvested by filtration, scraped into liquid nitrogen, and ground using a liquid nitrogen-cooled
Sequencing data processing and analysis

Reads were trimmed using cutadapt v. 2.10 with the following parameters for RNA-Seq (-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC minimum-length 20 -q 20) and Ribo-Seq samples (-a CTGTAGGCACCATCAAGTTAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC trimmed-only -q 20). Additionally, for Ribo-Seq, the reads were deduplicated with seqkit rmdup v. 0.10.1, and unique barcodes were then removed with cutadapt v. 2.10 –TAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC –minimum-length 20 -u -4). Afterwards, reads were aligned against eukaryotic rRNA sequence set obtained from silica-euk18 and rfam19 databases using bowtie2 v. 1.2.3. Only unmapped non-rRNA reads were used in the further processing.

Read mapping and counting against the Saccharomyces_cerevisiae.R64-1-1.95 (Ensembl) genome assembly was performed with STAR v. 2.7.9a. We estimated the position of the P-site for each dataset from the 5′ analysis. Read mapping and counting against the Saccharomyces_cerevisiae.R64-1-1.95 (Ensembl) genome assembly was performed with STAR v. 2.7.9a. Further Ribo-Seq and RNA-Seq library preparations were performed as described previously.12

Figure 1. Differential gene expression upon deletions of GIR2 and TMA46. (A) Total number of differentially expressed genes (passing FDR < 0.05) in each test. (B) and (C) Scatter plots illustrating expression changes at the level of transcription (X-axis, RNA-Seq) and translation (Y-Axis, Ribo-Seq) in the gir2Δ and tma46Δ strains, respectively. Translationally upregulated or downregulated (FDR < 0.05) genes marked in pink/blue, respectively. (D-G) Results of Gene Ontology (GO) enrichment analysis of RNA-Seq (D-E) and Ribo occupancy (RO) (F-G) considering upregulated (D, F) and downregulated (E, G) genes in gir2Δ vs. wt comparison. The numbers of differentially expressed genes with particular GO terms are shown in labels. X-axes show the enrichment P-value in log-scale.
on the basis of the length of each footprint using plastid v0.5.1. Fraction of reads in each phase and read length distribution were also obtained with plastid, see Figure S1 in the Extended data. The results show that almost 90% 28nt reads are in 0 phase through the annotated coding sequences (CDSs). Then we produced BedGraph profiles from SAM data with samtools v. 1.10 and bedtools v2.27.1. Coverage profiles were normalized using normalization factor and library size estimates from differential expression analysis (see below) separately for each bedGraph profile. Finally, we visualized coverage tracks in the modified genomic loci using svist4get. Figure S2 in the Extended data shows that the read counts originating from the mRNA encoded by the knockout gene in the corresponding strain are negligible. The coverage of the neighboring genes remains unaltered, i.e. there are no indications of the so-called neighboring gene effect (NGE).

**Differential expression and Gene Ontology (GO) enrichment analysis**

Statistical analyses were performed in R v. 4.1.2 using edgeR Bioconductor package. As mentioned above, the data were produced in two independent series which were analyzed separately. Genes not reaching 10 read count per million (CPM) in at least 4 RNA-Seq and 4 Ribo-Seq libraries were excluded from the analysis. Then, we performed the batch correction using ComBat-seq R package. Principal component analysis (PCA) plots of the raw and batch corrected expression profiles are shown in Figure S3 in the Extended data. A generalized linear model (glmQLFit, glmQLFTest of the edgeR package) was used to detect differentially expressed genes (for RNA-Seq, Ribo-Seq, and ribosome occupancy (RO) defined as the Ribo-Seq coverage of a CDS normalized to its RNA-Seq coverage) with the strain as a categorical variable. The false discovery rate (FDR < 0.05) was used for identification of differential expressed genes. We also performed Gene Ontology (GO) enrichment analysis for upregulated and downregulated genes with yeast-mine. The results are shown in Figure 1.

**Data availability**

**Underlying data**

NCBI Gene Expression Omnibus: RNA Sequencing and Ribosome profiling of TMA46, STM1 and YGR054W knockout yeast strains. Accession number GSE185458; https://identifiers.org/geo:GSE185458.

NCBI Gene Expression Omnibus: RNA Sequencing and Ribosome profiling of GIR2 and PUB1 knockout yeast strains. Accession number GSE185286; https://identifiers.org/geo:GSE185286.

**Extended data**

Figshare: Supplementary Table 1.csv https://doi.org/10.6084/m9.figshare.16818505. This project contains the following extended data:

- Supplementary Table 1.csv (Table with information about sequencing experiments)

Figshare: Supplementary Figures. https://doi.org/10.6084/m9.figshare.16818610.v1.

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

**Acknowledgments**

A.A.E., N.E.M., D.A.B., and S.E.D. are members of the Interdisciplinary Scientific and Educational School of Moscow University “Molecular Technologies of the Living Systems and Synthetic Biology”.

**References**

1. Ishikawa K, Azuma S, Ikawa S, et al.: Identification of DRG family regulatory proteins (DFRPs): specific regulation of DRG1 and DRG2. *Genes Cells.* 2005; 10: 139-150.
2. Fleischer TC, Weaver CM, McAffee KJ, et al.: Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Dev.* 2006; 20: 1294-1307.
3. Daugeron MC, Prouteau M, Lazorute F, et al.: The highly conserved eukaryotic DRG factors are required for efficient translation in a manner redundant with the putative RNA helicase Slh1. *Nucleic Acids Res.* 2011; 39: 2221-2233.
Open Peer Review

Current Peer Review Status: ✓ ✓ ✓

Version 1

Reviewer Report 07 December 2021

https://doi.org/10.5256/f1000research.78509.r100404

© 2021 Shirokikh N et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nikolay E. Shirokikh
Division of Genome Sciences and Cancer, The John Curtin School of Medical Research, The Australian National University, Canberra, Australia

Katrina Woodward
Division of Genome Sciences and Cancer, The John Curtin School of Medical Research, The Australian National University, Canberra, Australia

The work by Artyom Egorov and others touches upon proteins implicated in the processivity of translation of mRNA into polypeptides by eukaryotic ribosomes, specifically in resolving the stalled (or ‘collided’) translational intermediates. Research in this area is accelerating with many new studies uncovering unforeseen depths of the cells' mechanisms of translation fidelity control. Not too many datasets exist in this space to date and few are multi-target studies that provide a package of data obtained in a highly similar fashion for several translation-related genes, which is the rationale that justifies well the generated datasets.

In a series of yeast knock-out strains, the authors accurately employ common approaches of RNA sequencing and RNA-sequencing-based ribosome (translation) footprint profiling, to provide high-throughput short-read sequencing data resulting from the individual gene knock-out effects. The data are provided for the wild-type yeast of BY4742 strain used as a control, the GIR2 and TMA46 knock-out mutant strains which are in the focus of the dataset, and additionally the YGR054W (eIF2A), STM1, and PUB1 knock-out mutant strains, either commercially purchased or generated in the work.

The methods are written succinctly but capture most of the relevant detail where it matters. Perhaps, expanding on the particulars of the method used to isolate mRNA with oligo(dT) beads could add to the better understanding of the RNA-seq data (including any information about the other treatments and purifications used, such as RNA fragmentation and size cut-offs). Regarding the ribosomal material isolation, although not strictly necessary, it would add to the value of the datasets by outlining their applicability if the authors could provide more detail on the intensity of the digestion with RNase I to generate the footprints, the homogeneity of the resultant sedimentation profiles, and the boundaries used to define the monosomal fraction. For the lysate
clarification, the second spin uses very high acceleration; comments about whether the spin does or does not result in any loss of the heavy polysomes (e.g., if a suitably long particle sedimentation path was employed) could be beneficial. Although the authors are referring to the prior works, to bolster the applicability of the datasets and make their uptake convenient, any information about the additional size selection of the ribosome-protected fragments used (biochemical and/or bioinformatic) would be very helpful.

The data appears as accurate. The authors provide some differential abundance and GO term enrichment information for the more and less abundant gene categories within the pairwise comparisons, to characterise the datasets. This is very useful information; more information regarding the general properties of the datasets (such as total read/mappable/unique read number tables or histograms, gene coverage saturation plots etc.) could be further attractive in the main figure. The replicated datasets are validated through principal component analysis plots in Supplementary Figure 3, which show generally good groupings between the replicated sample types. The validity of the ribo-seq data is confirmed by Supplementary Figure 1 through demonstrating triplet periodicity of the dominant-length footprint fraction usually attributed to the cycloheximide-stabilised ribosomal protection extent (28 nt; although this is shown only for the wild-type, GIR2 and TMA46 knock-out strains). The datasets are most definitely usable, presented accessibly including uploads to the NCBI Gene Expression Omnibus, and are of a high utility.

Altogether, the authors provide an excellent resource to the field, which perfectly fits the scope of F1000Research Data Note publication, and will likely be cited.

Is the rationale for creating the dataset(s) clearly described?  
Yes

Are the protocols appropriate and is the work technically sound?  
Yes

Are sufficient details of methods and materials provided to allow replication by others?  
Yes

Are the datasets clearly presented in a useable and accessible format?  
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular biology, biochemistry, cell biology, genetics and gene expression, RNA, protein biosynthesis, ribosome profiling

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 01 December 2021

https://doi.org/10.5256/f1000research.78509.r100401
Olivier Namy
CNRS, I2BC, Université Paris-Saclay, Gif-sur-Yvette, France

In this manuscript Egorov A.A and colleagues aim to understand the translational consequences of the inactivation of two genes TMA46 and GIR2 in yeast *S. cerevisiae*. The experiments are clearly detailed and basic quality controls have been done for the RiboSeq analysis. It is not clear for me why the authors provided data for STM1, PUB1, and YGR054W without showing or analyzing them. I have several comments:

- The authors should justify why they have used a generalized linear model to calculate differential expression instead of a robust and well established negative binomial distribution (DESeq2). Why the authors used two different databases (silva-euk and rfam) to align reads to the rRNA? Did they use sequences specific to *S. cerevisiae*?

- Legend of Supplementary Fig1 needs to be improved. In the actual format, it is not obvious how the authors calculate the phasing of their RiboSeq data. Moreover, it is a pity that the analyses were not pursued a little further. For example, triplet periodicity could be shown, ribosome collision, and/or ribosome occupancy calculated. These additional qualitative analyses would strengthen the manuscript.

**Is the rationale for creating the dataset(s) clearly described?**
Yes

**Are the protocols appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and materials provided to allow replication by others?**
Partly

**Are the datasets clearly presented in a useable and accessible format?**
Yes

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 30 November 2021

https://doi.org/10.5256/f1000research.78509.r100403
Anton A. Komar

Center for Gene Regulation in Health and Disease, Cleveland State University, Cleveland, OH, 44115, USA

In this manuscript, Egorov and co-authors presented transcriptome and ribosome profiling data for *Saccharomyces cerevisiae* strains with individual deletions of the *TMA46* and *GIR2* genes. The interest in the protein products encoded by these genes has increased substantially in recent years, as these protein products have been shown to be involved in translation elongation and ribosome stalling. However, to date, there were limited “omics” data available from the respective knockout strains that could help shed additional light on the function of these genes. Particularly, ribosomal profiling of the knockout strains has not been performed. The present study fills in this gap and presents comparative ribosome profiling data obtained using the wild-type and knockout yeast strains. The ribosome profiling was followed by the standard quality checks and some basic downstream analysis, i.e. analysis of the differentially expressed genes and Gene Ontology (GO) enrichment analysis. Overall, these data will be useful for further research on *TMA46* and *GIR2* genes’ functions.

The sample preparation was performed using the standard protocols. The raw and processed data were deposited in the relevant open access repositories such as the Gene Expression Omnibus (GEO). Data processing has been described with reasonable detail (including the description of the command-line tools options and adapter sequences). The reliability of Ribo-Seq data has been demonstrated by read length distribution and the corresponding fraction of reads in each frame (Supplementary figure 1). All in all, the data seem to be technically sound and accurately described, but I would suggest adding illustrative plots showing triplet periodicity of the ribosome footprints.

**Is the rationale for creating the dataset(s) clearly described?**
Yes

**Are the protocols appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and materials provided to allow replication by others?**
Yes

**Are the datasets clearly presented in a useable and accessible format?**
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Translational control of gene expression
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com