RNA BIOLOGY
2021, VOL. 18, NO. 12, 2605–2616
https://doi.org/10.1080/15476286.2021.1935116

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER



# Comprehensive characterization of mRNAs associated with yeast cytosolic aminoacyl-tRNA synthetases

Shahar Garin[a], Ofri Levi[a], Megan E. Forrest[b], Anthony Antonellis[b,c], and Yoav S. Arava[a]

[a]Faculty of Biology, Technion - Israel Institute of Technology, Haifa, Israel; [b]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA; [c]Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, USA

**ABSTRACT**

Aminoacyl-tRNA synthetases (aaRSs) are a conserved family of enzymes with an essential role in protein synthesis: ligating amino acids to cognate tRNA molecules for translation. In addition to their role in tRNA charging, aaRSs have acquired non-canonical functions, including post-transcriptional regulation of mRNA expression. Yet, the extent and mechanisms of these post-transcriptional functions are largely unknown. Herein, we performed a comprehensive transcriptome analysis to define the mRNAs that are associated with almost all aaRSs present in *S. cerevisiae* cytosol. Nineteen (out of twenty) isogenic strains of GFP-tagged cytosolic aaRSs were subjected to immunoprecipitation with anti-GFP beads along with an untagged control. mRNAs associated with each aaRS were then identified by RNA-seq. The extent of mRNA association varied significantly between aaRSs, from MetRS in which none appeared to be statistically significant, to PheRS that binds hundreds of different mRNAs. Interestingly, many target mRNAs are bound by multiple aaRSs, suggesting co-regulation by this family of enzymes. Gene Ontology analyses for aaRSs with a considerable number of target mRNAs discovered an enrichment for pathways of amino acid metabolism and of ribosome biosynthesis. Furthermore, sequence and structure motif analysis revealed for some aaRSs an enrichment for motifs that resemble the anticodon stem loop of cognate tRNAs. These data suggest that aaRSs coordinate mRNA expression in response to amino acid availability and may utilize RNA elements that mimic their canonical tRNA binding partners.

**ARTICLE HISTORY**
Received 2 March 2021
Revised 18 May 2021
Accepted 21 May 2021

**KEYWORDS**
Aminoacyl tRNA synthetases; RNA-binding proteins; RIP-seq; post-transcriptional regulation; mRNA; yeast

## Introduction

Aminoacyl-tRNA synthetases (herein referred to as aaRS and specifically by replacing 'aa' with the three-letter code of each amino acid) are a family of proteins that recognize an amino acid and its cognate tRNAs to perform ATP-dependent, aminoacylation of tRNA. All organisms have a set of aaRSs that are responsible for charging the 20 canonical amino acids to their respective isoacceptor tRNAs [1,2].

tRNA recognition by aaRSs involves specific identity elements that are spread along the tRNA, predominantly at the tRNA anticodon loop and acceptor stem [3,4]. These elements can induce specific charging by either enhancing binding of cognate tRNA or hindering binding of non-cognate tRNA. Extensive studies on these tRNA elements revealed that different aaRSs may utilize different identity elements, and the extent of protein–RNA interaction varies enormously [5,6].

In addition to their canonical interaction with tRNA, certain aaRSs have been shown to interact with mRNA and exert a post-transcriptional regulatory function [7–9]. Initially identified in *E.coli*, binding of ThrRS to its own mRNA was found to serve a translation autoregulatory role. Interestingly, this recognition is made through RNA structural elements that are similar to the recognition elements within ThrRS cognate tRNA [10,11]. Recently, we and others found a similar structural RNA mimicry for yeast HisRS and mammalian ThrRS, suggesting a conserved mode of regulation [12,13]. Furthermore, we found that mRNA recognition may exploit another tRNA feature, the abundant pseudouridine modification [14], thus expanding the palette of mimicry modes. aaRSs binding to mRNA and pre-mRNA may regulate steps other than translation, including mRNA stability [15] and splicing [16]. Interestingly, aaRSs have been implicated in myriad dominant and recessive human diseases, and it is unclear if disease mechanisms are fully explained by their canonical charging activity [17,18]. Non-canonical, disease-related functions of aaRSs are suggested by their non-cytosolic localizations [19,20], mutations that are outside catalytic domains [21,22], and binding to molecules other than tRNA [8,23]. These data raise the possibility that roles other than tRNA binding and aminoacylation underlie the development of these diseases [24]. Thus, mRNA recognition and regulation by aaRSs may impose important functional roles.

Herein we aimed to obtain a comprehensive catalogue of mRNA-aaRS binding for the entire family of cytosolic aaRSs from a single organism (*S. cerevisiae*). We used a collection of isogenic yeast strains in which 19 of the 20 aaRSs were fused to GFP. Each strain, as well as a control untagged strain, was subjected to RNA binding protein-immunoprecipitation (RIP) by bead-conjugated GFP antibodies, followed by RNA-seq to identify bound RNAs. A significant number of bound mRNAs were detected in each sample, with the majority of

**CONTACT** Yoav S. Arava arava@technion.ac.il Faculty of Biology, Technion - Israel Institute of Technology, Haifa 3200003, Israel
Supplemental data for this article can be accessed here

mRNAs associated with all 19 aaRSs studied. Enrichment analysis identified tens of mRNAs that are significantly associated with each aaRS compared to controls. Gene Ontology (GO) term analyses revealed an enrichment for mRNAs encoding components of amino acid metabolism pathways and ribosome biogenesis. Sequence and structure motif analysis for target mRNAs revealed similarities between the tRNA recognition element and enriched mRNA motifs for some aaRSs. These data expand our knowledge of mRNA associations with aaRSs and their possible mechanisms of interaction. Furthermore, it suggests coordination between tRNA and mRNA recognition in response to amino acid levels.

## Results

### A collection of GFP-tagged aaRSs

We aimed to identify mRNAs that are bound by yeast (*S. cerevisiae*) cytosolic aaRSs in an unbiased manner. To that end, we applied an RNA-binding protein immunoprecipitation (RIP) followed by RNA-sequencing approach (RIP-seq) (Figure 1(a)). We used a collection of isogenic yeast strains that each express a single aaRS fused to GFP, which provides a uniform and efficient handle for the RIP step. Of the 20 cytosolic yeast aaRSs, 12 were previously tagged at their C-terminus and are available [25] (Table 1). While the remaining eight were also suggested to be tagged by this approach, western analyses did not reveal signals at their expected sizes (data not shown). We therefore regenerated these strains using the same homologous recombination scheme [25] and the same parental strain (*BY4741*). Unfortunately, despite extensive efforts, we were unable to tag ProRS at the C-terminus; this is likely because the C-terminus is folded into the protein's active core and is important for enzyme function [26]; GFP insertion probably leads to unviable cells. Altogether, a set of 19 isogenic strains each expressing C-terminally GFP-tagged cytosolic aaRS was used herein (Table 1).

### Identification of mRNAs associated with each aaRS

To identify mRNA targets for each aaRS in our library, each strain (and matched untagged control), was grown to mid-logarithmic phase, subjected to formaldehyde crosslinking to maintain protein–RNA interactions and harvested. Tagged aaRSs were immunoprecipitated with anti-GFP beads (Chromotek), and the quality of isolation was verified by western analysis of protein samples from key steps of the protocol (Figure 1(b)). When an anti-GFP antibody was used, a band of the expected size is apparent in the Input samples, confirming proper tagging and expression. This signal is usually decreased in the flow-through sample, indicating efficient binding to the beads. Concomitantly, a clear signal is apparent in the Bound sample, attesting to the efficiency of the isolation. Non-specific binding was tested with anti-Hxk1 antibody and did not reveal any signal in the 'Bound' sample. These results confirm proper expression of the fusion protein and purification that is both specific and efficient.

To examine the quality of eluted RNA, we used automated electrophoresis (using the 2100 Bioanalyzer system). Cognate tRNAs could be readily identified by a signal in the region of 90 nucleotides (Supplementary Figure 1). Additional signals, of a longer size, were observed in most cases. These signals varied between aaRSs (Supplementary Figure 1(b)), indicating different tendencies to bind non-tRNA transcripts.

For global mRNA analysis, RNA-seq was performed on Bound samples of each tagged aaRS experiment. To account for differences in mRNA abundance, RNA from input samples was also subjected to RNA-seq (a mix of all input samples). In addition, to account for non-specific association with the beads, RNA-seq was also performed on elution samples from a parental, untagged strain (*BY4741*). Two biological repeats were performed for each experiment. Overall, 42 RNA samples were subjected to library preparation, deep sequencing, read annotation, and read counting. Full lists of bound mRNAs per aaRS are provided in Supplementary Table 1.

The overall outcomes of the RNA-seq are presented in the heatmap in Supplementary Table 2. The two biological repeats of each aaRS show a high Pearson correlation ($r_p$), usually greater than 0.95. Interestingly, the correlation between different aaRSs is also high, indicating that many mRNAs are bound by multiple aaRSs. Of note, the correlation with the Input RNA data is much lower, confirming that signals are not merely a reflection of mRNA abundance. Furthermore, correlation with the eluted RNA from the untagged strain is also low, indicating that identified RNAs are not due to non-specific binding to the beads.

We next identified mRNAs whose association with an aaRS is enriched relative to other mRNAs (dubbed 'cognate mRNAs'). We employed the DEseq2 pipeline [27] for each aaRS dataset and utilized a double-sieve approach to select for positive hits: enriched mRNAs were selected both against the Input RNA data set (to select for mRNAs that are enriched irrespective of their abundance) and against the untagged dataset (to remove non-specific interactors with the beads). Cut-offs of count per million greater than 0.5, fold change higher than 2, and $q$-value of biological repeats lower than 0.05 were used in each sieve. While this stringent selection may lower the number of positive hits, it provides a high-confidence dataset of enriched targets.

As is evident from Figure 2 and Supplementary Table 3, there are significant differences in the number of cognate mRNAs each aaRS binds, ranging from 176 (bound by PheRS) to 0 (MetRS). A small number of statistically confidant targets are apparent for GlyRS, ThrRS, LeuRS and IleRS, presumably due to low similarity between biological repeats. Intriguingly, MetRS has zero cognate mRNAs despite its relatively similar biological repeats ($r_p$ 0.95). MetRS has been shown to bind tRNA while in a multi subunit complex (MSC) with GluRS and Arc1 [28]. Nevertheless, our data indicate a different mRNA binding preference between

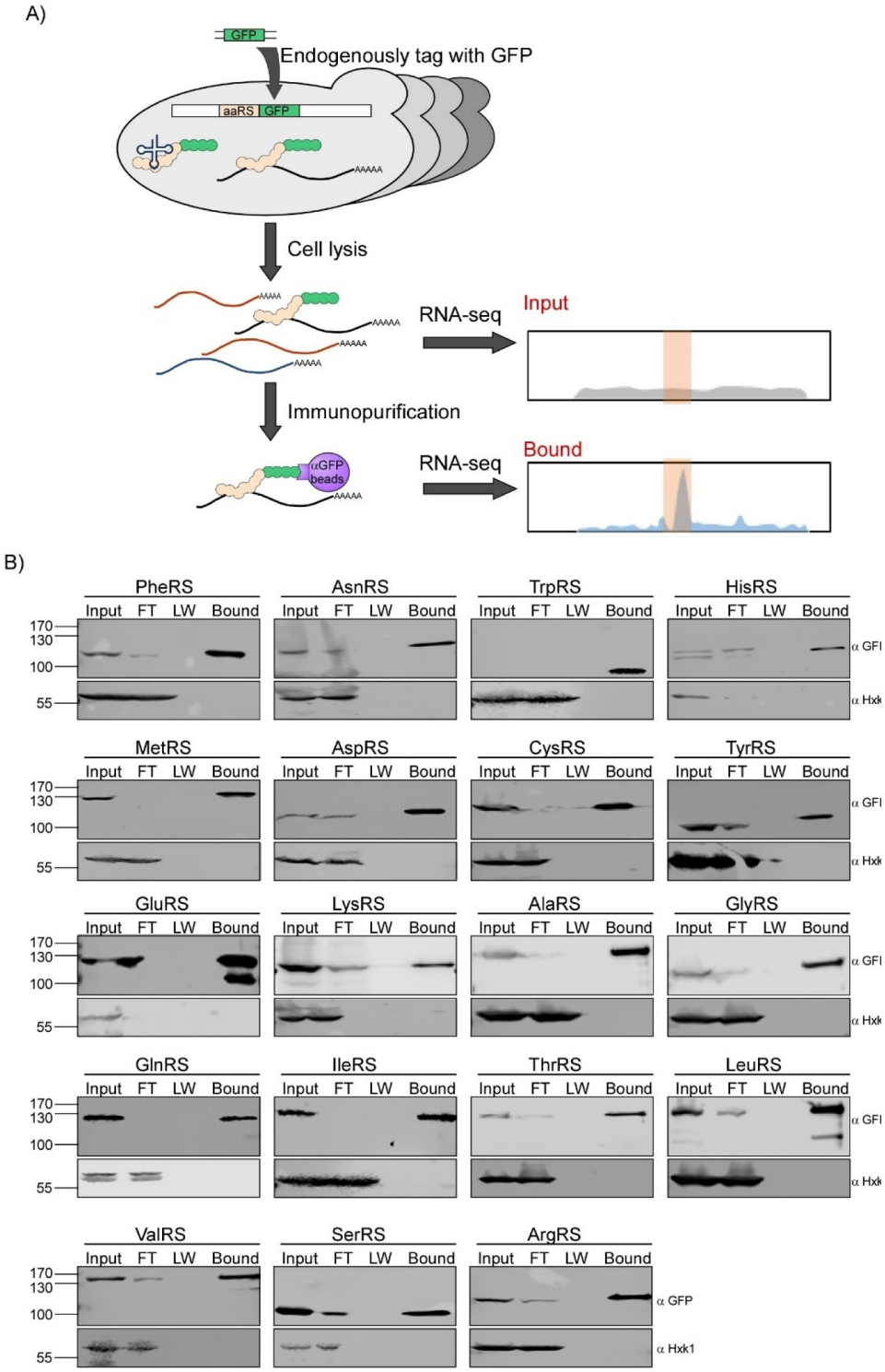


**Figure 1. Comprehensive identification of mRNAs bound by yeast cytosolic aaRS**. (a) Scheme of the experimental procedure. aaRSs were tagged with GFP by genomic integration. aaRSs were isolated with their bound RNA through GFP_TRAP columns (Chromotek), and mRNAs enriched in the Bound samples compared to the Input samples were identified by RNA-seq. (b) Protein samples from either the cellular lysate (Input), the unbound flow-through (FT), the last wash (LW) or the eluted sample from the beads (Bound) were subjected to western analyses with the indicated antibodies.

GluRS and MetRS, suggesting that mRNA binding occurres by monomeric MetRS and GluRS. MetRS was previously found to be localized in the nucleus while it is outside of the MSC [29], which may explain its low mRNA binding yield.

aaRSs differ in the number of distinct tRNA isoacceptors that they bind [30]. Nevertheless, as can be seen in Figure 2, there is no relationship between the number of anticodons recognized by an aaRS and the number of cognate mRNAs. Thus, promiscuity in tRNA binding is unlikely to explain diversity in mRNA binding.

To validate the RNA-seq results by an alternative approach, we repeated the RIP protocol for three

**Table 1.** List of GFP-tagged strains.

| Lab Number | Amino Acid | Gene Name | Common Name | Number of recognized anticodons | Reference |
|---|---|---|---|---|---|
| YA1536 | Aspartic Acid | YLL018C | DPS1 | 1 | [25] |
| YA1537 | Isoleucine | YBL076 C | ILS1 | 2 | [25] |
| YA1538 | Glycine | YBR121C | GRS1 | 3 | [25] |
| YA1539 | Methionine | YGR264 C | MES1 | 1 | [25] |
| YA1541 | Cysteine | YNL247W | CRS1 | 1 | [25] |
| YA1543 | Leucine | YPL160W | CDC60 | 4 | [25] |
| YA1544 | Alanine | YOR335C | ALA1 | 2 | [25] |
| YA1545 | Threonine | YGR264C | THS1 | 3 | [25] |
| YA1546 | Tryptophan | YOL097C | WRS1 | 1 | [25] |
| YA1547 | Tyrosine | YGR185C | TYS1 | 1 | [25] |
| YA1554 | Asparagine | YHR019C | DED81 | 1 | This study |
| YA1457 | Glutamine | YOR168W | GLN4 | 2 | [25] |
| YA1458 | Histidine | YPR033C | HTS1 | 1 | [25] |
| YA1569 | Arginine | YDR341C | N/A | 4 | This study |
| YA1570 | Glutamic Acid | YGL245W | GUS1 | 2 | This study |
| YA1571 | Serine | YDR037W | SES1 | 4 | This study |
| YA1572 | Lysine | YDR037W | KRS1 | 2 | This study |
| YA1573 | Valine | YGR094W | VAS1 | 3 | This study |
| YA1574 | Phenylalanine | YLR060W | FRS1 | 1 | This study |

representative aaRSs and quantified several candidates by RT-qPCR (Figure 2(b)). These data indicate that all tested mRNAs had much higher association values with all aaRSs compared to the untagged control.

### Preferred association with functional groups

Eleven aaRSs bind forty or more cognate mRNAs, which allows reliable analyses of features shared among them. Interestingly, a significant number of these targets are bound by multiple aaRSs (Figure 3(a)). For example, 114 of the 176 mRNAs bound by PheRS are also bound by ArgRS (65%). The extent of overlap does not seem to be related to the number of isoacceptors that these aaRSs bind, nor to the structural class they belong to. Clustering of cognate mRNAs to six groups revealed cohorts that are enriched with interesting cellular processes (Figure 3(b), Supplementary Table 4): mRNAs encoding proteins that are involved in ribosome biogenesis (cluster A) and rRNA processing (cluster B) are bound by the majority of aaRSs. Furthermore, mRNAs encoding enzymes with roles in amino acids biosynthesis (cluster F) are bound preferentially by three aaRSs (PheRS, SerRS and AspRS).

As a complementary approach, we performed Gene Ontology (GO) enrichment analysis to each list of cognate mRNAs (using the PANTHER GO-Slim Biological Process data set [31]). Of the 11 aaRSs with 40 or more targets, no significant enrichment was observed for ValRS. Nevertheless, two groups of pathways dominated the lists of the remaining aaRSs: those that relate to amino acid metabolism and to ribosome biogenesis (Figure 4(a), Supplementary Table 5). Some aaRSs (i.e. PheRS, SerRS) bind mRNAs that encode proteins of both groups, yet some (ArgRS, GluRS, LysRS, TyrRS and CysRS) bind almost exclusively those encoding ribosome biogenesis and others (AspRS and GlnRS) preferably bind mRNAs encoding amino acid metabolism factors. For example, AspRS binds mRNAs encoding enzymes that catalyse reactions in many stages of isoleucine, valine, and leucine biosynthesis pathways (Figure 4(b)). CysRS, on the other hand, binds mRNAs that encode factors involved in ribosome biogenesis (Figure 4(c)). These data suggest that aaRSs serve as regulators of amino acid metabolism and ribosome biosynthesis.

Importantly, our validation of association of representative mRNAs by RT-qPCR (Figure 2(b)) partially supported this observation. We observed consistent results for mRNAs encoding enzymes involved in amino acid biosynthesis (ILV5, ILV3 and ILV6) (i.e. high association with PheRS and AspRS and low in CysRS). However, representative mRNAs for ribosome biogenesis pathway (UTP4, UTP21 and NAN1) appeared relatively low in PheRS and CysRS, inconsistent with the enrichment of this functional group among these aaRSs. Therefore, this functional enrichment should be interpreted with caution.

### Association with tRNA-like elements

To gain insight into the mechanisms that underlie associations between aaRSs and mRNAs, we subjected the mRNAs associated with each aaRS to sequence motif enrichment analysis (DREME [32]). Unfortunately, this straightforward analysis provided many motifs of high resemblance (Supplementary Figure 2). This is probably because many of our input sequences are redundant among several aaRSs and motifs that are common to these mRNAs dominate our lists. While association with multiple aaRSs may represent a true biological feature, it hinders identification of motifs that are specific to each aaRS. To overcome this, we removed cognate mRNAs that are shared with other aaRSs for further analysis; we grouped aaRSs by the number of distinct cognate tRNA isoacceptors presented in Figure 2, where each aaRS was compared to those that are shared within its group. PheRS appeared to bind 49 mRNAs that are not associated with AsnRS, TyrRS, CysRS or AspRS. This group was enriched with mRNAs encoding tRNA modification enzymes (in particular methyltransferases) (GO:0008173), suggesting a possible regulatory role on this family of enzymes (Figure 5(a)). Interestingly, analysis for sequence motifs that are shared among these 49 mRNAs revealed that 34 include the sequence UGAAG, which is identical to the sequence at $tRNA^{Phe}$ anticodon loop (Figure 5(b)). Similar analysis for other aaRSs did not detect additional motifs of interest, suggesting that a better approach is needed to sift specific motifs from common ones.

Considering the limited output of a simple sequence motif search, we instead sought to subject aaRS cognate mRNAs to an alternative analysis, focusing on identification of structural similarities with cognate tRNAs. We therefore exploited the patteRNA algorithm [33] to scan for 17 nt anticodon stem loop (ASL)-like structures within an experimentally determined, genome-wide database of *S. cerevisiae* RNAs secondary structures [34]. This scan revealed ~39,000 such structures, spread along 3,140 mRNAs. For each aaRS, we further filtered this list for those that contain trinucleotide

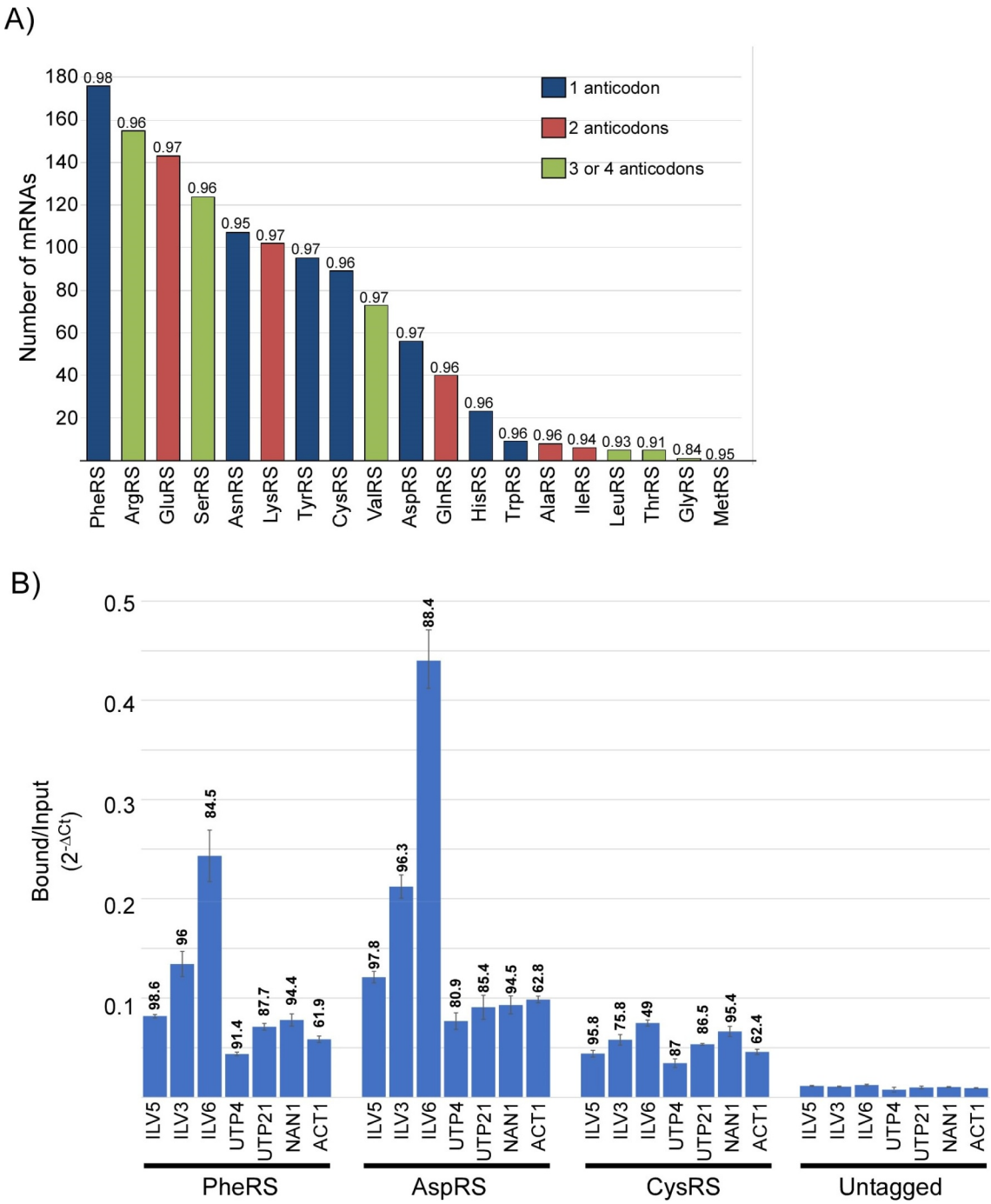


**Figure 2. aaRS vary significantly in the number of mRNAs bound**. (a) mRNAs that were enriched by > 2 fold compared to the input samples and to the untagged control, with q-value of biological repeats lower than 0.05, were defined as positive hits. Numbers above each bar are Pearson correlation of biological repeats, and colours indicate the number of tRNA bound by each aaRS. (b) Validation of representative mRNAs was performed by RIP followed by RT-qPCR for the indicated aaRSs. Results are from two independent biological repeats and presented as $2^{-\Delta Ct}$ of Bound to Input samples. Percentile ranks of each mRNA in the Bound vs Input Log2Foldchange values of the RIP-seq data are indicated on top of each bar.

sequences identical to their respective cognate tRNA isoacceptors (Supplementary Table 6). Comparison of the overall abundance of such structures in the transcriptome with their abundance among target mRNA identified by RIP revealed that ASL-like structures are statistically enriched among RIP mRNAs in some cases (Figure 6(a)). For example, GlnRS binds 40 mRNAs, 90% of which contain an ASL-like structure, significantly higher than expected based on the abundance of $tRNA^{Gln}$ ASL-like within the transcriptome (59.5%) (*p value* $2.98*10^{-6}$, hypergeometric test). Significant enrichment of ASL-like elements (*p value <0.05*) in cognate mRNAs was also apparent for AspRS, AsnRS and GluRS. For the multi-isoacceptor aaRSs (ArgRS, SerRS, LysRS and GlnRS), we refined this enrichment analysis to determine if specific tRNA isoacceptors are over-represented among cognate mRNAs (Supplementary Figure 3(a)). Indeed, some isoacceptors were found to be more enriched than others, regardless of their abundance in the genome. For example, ASL-like

A)

| | PheRS$^{II}$ (176) | ArgRS$^{I}$ (155) | GluRS$^{I}$ (143) | SerRS$^{II}$ (124) | AsnRS$^{II}$ (107) | LysRS$^{II}$ (102) | TyrRS$^{I}$ (95) | CysRS$^{I}$ (89) | ValRS$^{I}$ (76) | AspRS$^{II}$ (56) | GlnRS$^{I}$ (40) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PheRS$^{II}$ (176) | 100 | 65 | 59 | 56 | 48 | 44 | 42 | 36 | 30 | 26 | 15 |
| ArgRS$^{I}$ (155) | 74 | 100 | 59 | 53 | 54 | 48 | 45 | 41 | 30 | 23 | 16 |
| GluRS$^{I}$ (143) | 72 | 64 | 100 | 58 | 49 | 59 | 55 | 39 | 41 | 22 | 15 |
| SerRS$^{II}$ (124) | 79 | 66 | 67 | 100 | 60 | 60 | 51 | 50 | 39 | 31 | 17 |
| AsnRS$^{II}$ (107) | 79 | 78 | 65 | 70 | 100 | 59 | 49 | 54 | 43 | 28 | 22 |
| LysRS$^{II}$ (102) | 76 | 74 | 82 | 73 | 62 | 100 | 62 | 50 | 52 | 25 | 20 |
| TyrRS$^{I}$ (95) | 78 | 73 | 83 | 66 | 55 | 66 | 100 | 48 | 56 | 28 | 17 |
| CysRS$^{I}$ (89) | 71 | 72 | 63 | 70 | 65 | 57 | 52 | 100 | 43 | 22 | 17 |
| ValRS$^{I}$ (76) | 73 | 64 | 79 | 66 | 63 | 73 | 73 | 52 | 100 | 27 | 18 |
| AspRS$^{II}$ (56) | 82 | 63 | 57 | 68 | 54 | 46 | 48 | 36 | 36 | 100 | 27 |
| GlnRS$^{I}$ (40) | 65 | 63 | 55 | 53 | 60 | 50 | 40 | 38 | 33 | 38 | 100 |

B)

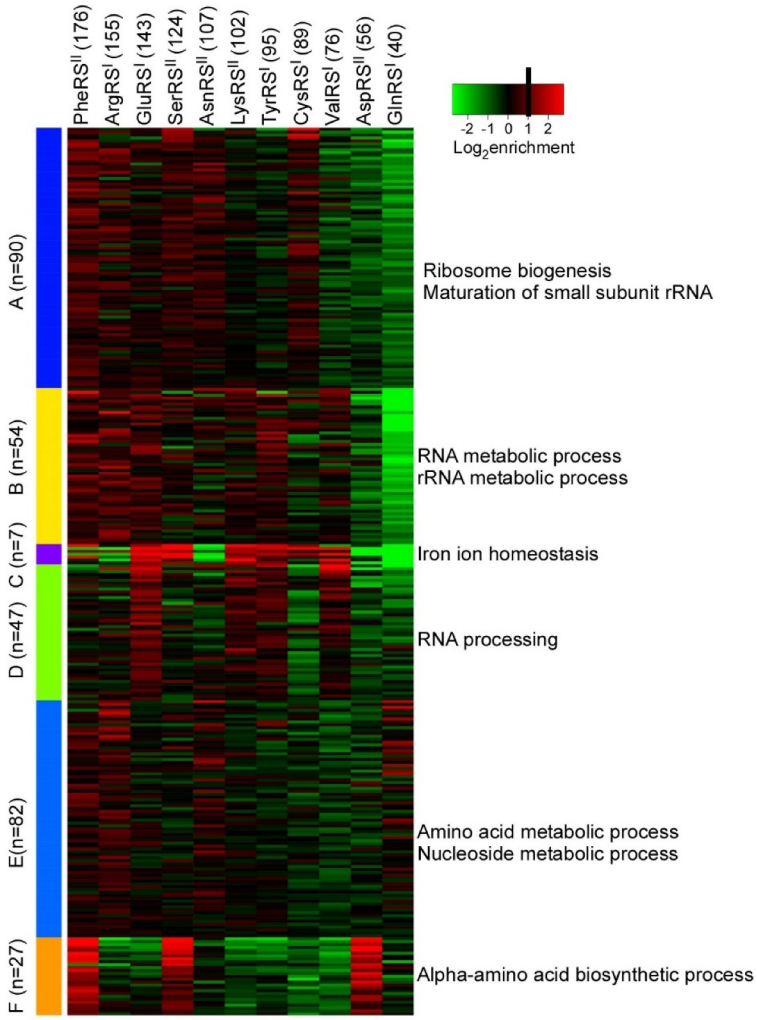


**Figure 3. mRNAs of similar functions are associated with multiple aaRSs**. (a) Heatmap of cognate mRNAs that are also bound by other aaRSs. Numbers in the table are the percent of mRNAs bound to the aaRS on the left, which are also included in the list of the aaRS at the top. Superscript I and II indicate belonging to class I or class II structural groups, respectively. Numbers in parentheses are the number of mRNAs defined as bound to each aaRS. (b) Clustering analysis (using k-means of 6) for all cognate mRNAs, bound by one or more aaRS (total of 308 mRNAs, complete lists of mRNAs are available in Supplementary Table 4). Prominent cellular process in each cluster is indicated to the right. n indicate the number of mRNAs in each cluster. Colour scale is shown at the top, and the vertical bar indicates the value from which colours denote our assignment as bound by an aaRS.

structures containing the Arg-CCG trinucleotide sequence are found in 19.9% of mRNAs transcriptome-wide, yet 27.1% of ArgRS cognate mRNAs contain it (*p-value 0.0097*, hypergeometric test). Intriguingly, sequence similarity appears to play only a partial role, as the identified ASL-like structures showed a wide-range of sequence identity to their respective anticodon loop sequences, usually less than 50% identity (overall median 47.1%) (Figure 6(b)). Further, analysis of the location of ASL-like elements within mRNA revealed that these elements are predominantly located within the coding

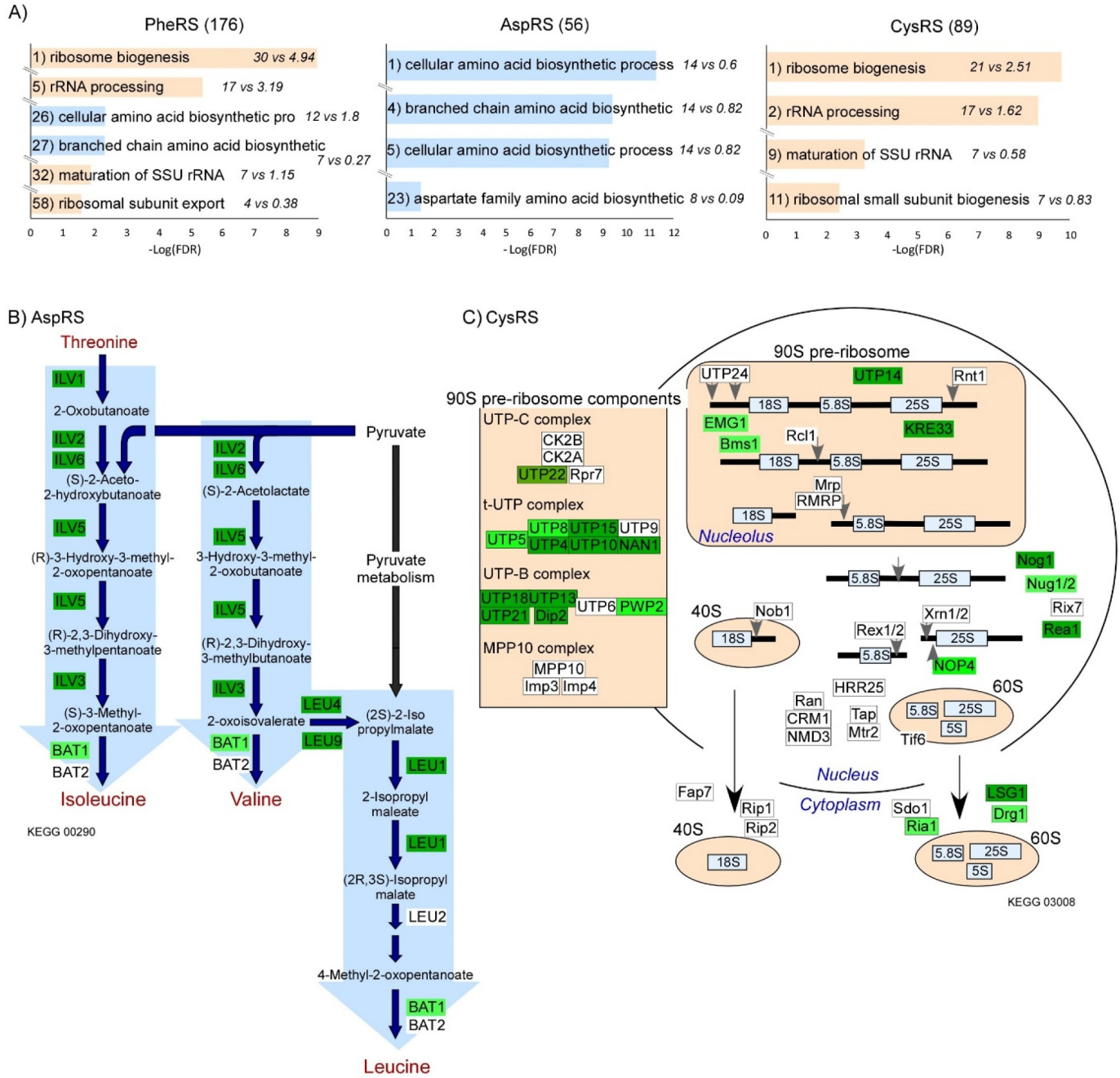


**Figure 4. aaRSs preferentially bind mRNAs encoding proteins involved in amino acids metabolism and ribosome biogenesis.** (a) Enriched GO terms for representative aaRSs. mRNAs enriched per aaRS were subjected to GO term analyses in PANTHER. Presented here are terms with -logFDR <0.05, that are related to ribosome biosynthesis (light brown) or amino acids biosynthesis (light blue). Terms are indicated within each bar, and numbers in italics indicated the number of detected vs. expected genes per group of that size. Numbers preceding each term is its rank within all significant terms. Full dataset is in Supplementary Table 5. (b) Simplified scheme of Isoleucine, Valine and Leucine biosynthesis pathway (following KEGG pathway 00290). Enzymes with green background are encoded by mRNAs bound by AspRS. Lighter green indicates cases of lower enrichment; *i.e.* apparent only compared to the Input and not the untagged data. (c) Scheme of ribosome biogenesis and complexes involved in the process (following KEGG pathway 03008). Colour coding of mRNAs bound by CysRS is as in (b).

region of the transcript (Figure 6(c)). Taken together, these data are consistent with the notion that tRNA anticodon loop structural similarity underlie binding of some aaRSs to cognate mRNAs.

As noted above (Figure 3), many mRNAs are associated with multiple aaRSs. Notably, these mRNAs also tended to contain multiple ASL-like elements (Supplementary Figure 3 (b), Supplementary Table 7). These putative binding sites showed different position patterns–while some mRNAs have ASL-like spread along the transcript, others have clusters at specific sites (Supplementary Figure 3(c)). Thus, some

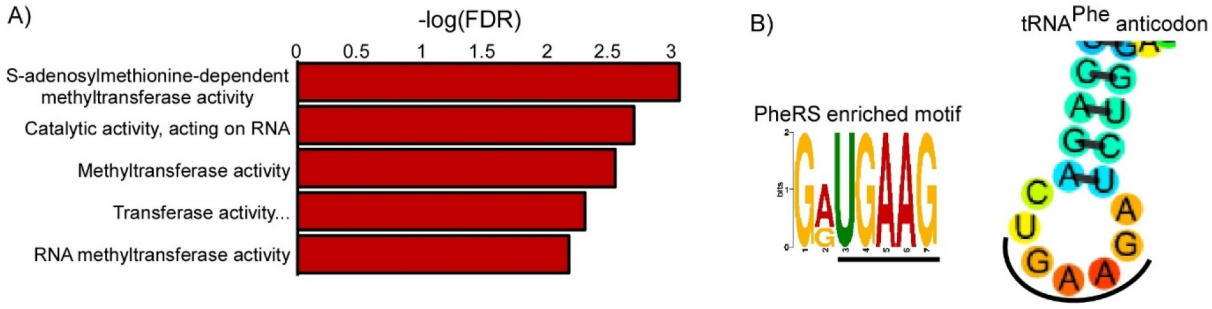


**Figure 5. Enriched sequence motif in PheRS targets**. (a) mRNAs that are cognate to PheRS, yet not to other aaRSs that recognize a single anticodon (AsnRS, TyrRS, CysRS or AspRS), were subjected to GO term analysis and key groups are presented. (b) PheRS-specific mRNAs were subjected to motif-enrichment analysis, and the top hit is presented (*e value 1.7e-0.02*). $tRNA^{Phe}$ anticodon loop is presented for comparison.

mRNAs might be hubs for aaRSs interaction and serve as relays for downstream regulation.

## Discussion

In this study, we complemented the publicly available 12 strains of GFP-tagged aaRSs with seven new strains, thus generating a collection of isogenic strains, each endogenously expressing one of 19 cytosolic aaRSs tagged with GFP at its C-terminus. While we used these strains herein for comparative RIP-seq analyses, these strains may be useful for various other applications. Recent studies indicate unexpected localization for many aaRSs, with intriguing physiological implications [20]. Of interest is the observation that some aaRSs are localized to the nucleus [35]: The GFP tagging is useful for imaging confirmation and time-lapse microscopy of transport kinetics. Moreover, the GFP tag can be used as a handle for other experimental approaches, such as ChIP, to determine DNA binding sites by aaRSs.

A key observation from our global view of aaRSs association with mRNA is the apparent enrichment of mRNAs encoding enzymes involved in amino acid metabolism and ribosome biogenesis (Figures 3 and 4). While these initial observations need further establishment, they raise intriguing possibilities for aaRSs in sensing of amino acid levels and subsequent regulation of their production. The first step in tRNA charging involves amino acid binding with high affinity. It is therefore possible that when amino acid levels change, a vacant aaRS binds the relevant mRNAs and regulates their expression to maintain proper production of the necessary amino acid. Interestingly, our data suggests that aaRS binding is not restricted to mRNAs involved in production of their cognate amino acid. For example, AspRS binds mRNAs encoding enzymes involved in Ile, Val and Leu metabolism (Figure 4(b)). This suggests a general response mechanism that affects multiple pathways. Another group that is highly represented is the group of mRNAs encoding proteins involved in ribosome biogenesis, implying a regulatory loop that connects tRNA charging to the ribosome content of a cell.

Another prominent observation is that many mRNAs are bound by multiple aaRSs. This is apparent from the overall high correlation between RIP-seq data of different aaRSs (Figure 2), and from the overlap in ASL-like structures among bound mRNAs (Figure 6). In mammals, multiple aaRSs are part of a large multi subunit complex [36], which may lead to similar cohorts of bound mRNAs. In yeast, however, only MetRS and GluRS are known to be in complex, through Arc1 protein [28]. Therefore, most mRNAs in yeast are expected to bind several aaRSs in an independent manner. mRNAs that are bound by multiple aaRSs may serve as regulatory hubs that coordinate downstream post-transcriptional regulatory impacts of aaRSs. For example, mRNAs encoding tRNA modifiers Pus1, Pus7 and TRM44 are bound by 4, 6 and 8 aaRSs, respectively (Supplementary Table 7). Presumably, by affecting the translation of these modifiers, aaRSs control modification of their cognate tRNAs and consequently their availability. Such models are yet to be explored.

Our analyses for enriched sequences among cognate mRNAs revealed similarity to cognate tRNA, particularly in the case of PheRS (Figure 5(b)). This finding is consistent with previous observations of similarities between the target motif within an mRNA and the cognate tRNA (reviewed in [8,37]) consistent with models for a role for tRNA sequences in mRNA evolution [38]. Moreover, structural analyses revealed an enrichment of anticodon stem loop (ASL)-like structures within bound mRNAs for tested aaRSs (Figure 6(a)). These may serve as binding sites for an aaRS in a manner that mimics its association with cognate tRNAs, thus leveraging the essential tRNA recognition domains of these enzymes. The presence of such tRNA-like sequences in aaRS-bound mRNAs offers the opportunity to alternate between tRNA charging and mRNA regulation. For example, increasing tRNA levels may direct aaRS from cognate mRNAs to cognate tRNAs to meet increased charging demands. Conversely, increased expression of vacant mRNAs may re-acquire the associated aaRS to regulate mRNA expression to meet other cellular needs. Indeed, we previously presented such a regulatory loop for *S. cerevisiae* HisRS, where increased $tRNA^{His}$ levels reduced association of HisRS with its own mRNA and led to increased HisRS protein levels [13]. Surprisingly, while we previously found the mRNA encoding HisRS to be the top ranked target for HisRS binding [13], herein it was ranked much lower and did not pass our filtration criteria. This suggests that mRNA association is of low affinity and highly regulated, rendering it sensitive to the applied experimental conditions.

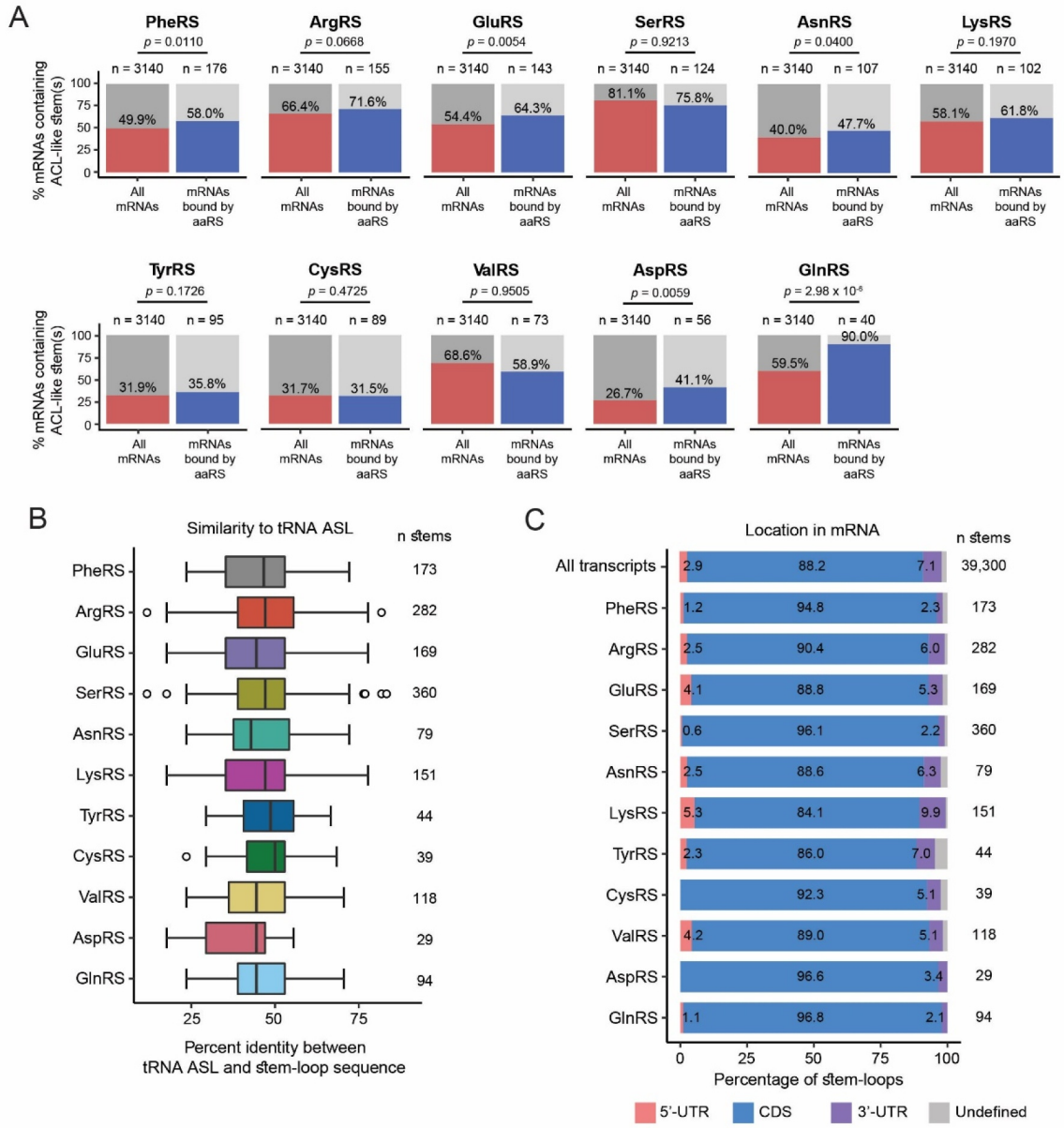


**Figure 6. Anticodon stem loop (ASL)-like structures within aaRS-bound mRNAs**. (a) 17 nt stem-loop sequences matching the tRNA ASL structure (39,300 total) were identified in 3,140 yeast transcripts using yeast transcriptome-wide secondary structure data. A subset of these ASL-like structures were found to contain a trinucleotide sequence matching the anticodon sequence for one or more cognate tRNA isoacceptors. For each aaRS, left bar represents the percentage of all yeast transcripts that contain all cognate ASL-like motifs. Right bar indicates percentage of aaRS-bound mRNAs that were found to contain tRNA ASL-like sequences. P-value indicates result of cumulative hypergeometric distribution. Total number for each category is indicated above each bar. (b) ASL-like sequences show varying sequence similarity to cognate tRNAs. Shown for each aaRS are boxplots for percent identity between cognate tRNA anticodon loop sequence(S) and ASL-like motif sequences in RIP-seq mRNAs, as determined by pairwise alignments. Total number of unique ASL-like element within RIP-seq mRNAs for each aaRS is indicated on the right. (c) Position of ASL-like within mRNA regions. n indicates the total number of ASL-like elements that were analysed. 'Undefined' indicates missing position data or that the stem-loop contains the annotated start or stop codon for a given mRNA.

Positional analysis revealed that ASL-like structures are predominantly located within the coding region of the mRNA. Protein binding within coding regions is usually associated with an impact on ribosome transit, suggesting that aaRSs have a regulatory role over protein synthesis rate [39,40]. We note that our experimental approach does not allow identification of exact aaRS binding sites along an mRNA. We therefore cannot unequivocally assert that aaRSs bind these mimics directly; this awaits the establishment of single-base resolution approaches in yeast cells. Nevertheless, similarity between tRNA and mRNA motifs suggests the aaRS shuttles between tRNA and mRNA binding [8]. Such a shuttling will allow communication between tRNA charging and mRNA expression regulation. Thus, aaRSs may represent a new layer of interaction between seemingly distant RNA processes.

## Materials and methods

### Yeast strains and growth conditions

Parental yeast strain for all studies is BY4741 (*MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*). Cells were grown in liquid or on plates of YPD (1% Yeast extract 2% Peptone, 2% Dextrose) at 30□.

### Yeast transformation

Yeast in early exponential phase were collected (50 ml) by centrifugation, washed once with sterile water and re-suspended in 0.4 ml of 0.1 M LiAc, of which 100 µl were used per transformation. Each 100 µl fraction was pelleted and suspended in 40 µl sterile water, 36 µl 1 M LiAc, 25 µl salmon sperm DNA (2 mg/ml), 100 µl donor DNA from PCR amplification (using primers indicated in Supplementary Table 8 and pFA6a-GFP(S65T)-His3MX6 as template) and 240 µl 50% PEG. Samples were incubated for 30 minutes at 30□ followed by 15 minutes at 42□. Transformation mixtures were plated on appropriate selection medium. Positive colonies were collected after 2 days and re-plated on selection medium for verification by PCR and western analysis.

### RNA binding proteins ImmunoPrecipitation (RIP)

Yeast in exponential phase (250 mL) were collected by centrifugation at 3000 g for 4 min at 25□. Cells were re-suspended in 25 ml PBS supplemented with 0.05% formaldehyde and incubated for 10 minutes at room temperature. Glycine solution was added to 0.125 M final concentration, and cells were incubated at room temperature for 3 minutes. Cells were washed twice with 20 mL Buffer A (20 mM Tris pH 7.5, 140 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 0.1% SDS), re-suspended in 1 mL of buffer B (Buffer A supplemented with 1 mM DTT, 2 mM PMSF, 10 µg/mL Leupeptin, 14 µg/mL Pepstatin, 0.02 U/µL RQ1 RNase-free DNase (Promega), 0.24 U/µL RNasin (Thermo Scientific)) and lysed using glass beads. The lysate was cleared by centrifugation at 10,000 RPM for 10 minutes at 4□, and the supernatant was transferred to clean tubes. For protein sample, 25 µL was set aside and for RNA Input control 100 µL was set aside. The remaining lysate was mixed with GFP_Trap agarose beads from Chromotek (20 µL) and incubated for 2 hours at 4□ with gentle rotation. Beads were spun down at 200 g and 25 µL of the supernatant was set aside for 'Flow Through' protein control. Bound material was washed four times with 900 µL of buffer C (20 mM Tris pH 7.4, 1 M NaCl, 0.5% NP-40, 0.5 mM DTT, 10 U/µL RNasin (Thermo Scientific), 0.5 mM EDTA, 0.1% SDS) by centrifugation at 200 g at 4□ and 25 µL of the last wash protein sample was kept for washing efficiency control (Last Wash). Bound material was eluted by two rounds of addition of 100 µL 0.2 M Glycine pH 2.5 that were neutralized with 60 µL Tris 1 M pH 10.4 after collection. A 25 µL was set aside for Bound protein sample. RNA samples (Input and Bound) were mixed with 250 µL 2X Reverse Crosslinking buffer (150 mM Tris pH 7.5, 15 mM EDTA, 30 mM DTT, 3% SDS) and RNasin 0.5 µL (Thermo Scientific) was added to each sample. Samples were incubated for 2 hours in 65□ and RNA was extracted by acidic phenol-chloroform [41].

### RNA sequencing

Library preparations, sequencing and gene annotation were performed by the Nancy and Stephen Grand Israel National Center for Personalized Medicine, at the Weizmann Institute of Science. Libraries of all RIP samples excluding the total RNA samples were performed without poly-A selection, to avoid losses. The samples were sequenced on a single lane on Illumina HiSeq 2500 platform, yielding 140–400 thousand reads per Bound sample and 2.5 million reads per Input sample. The reads were mapped to the S288c Saccharomyces cerevisiae version R64-1-1 genome and uniquely mapped reads were counted. Read count files were uploaded to the iDEP server [42] for further analysis. Pearson coefficient was calculated between each data set, and a correlation matrix was prepared in Excel. DEseq2 method was used to determine the log2fold change and *p-values* of Bound vs. Input and Bound vs Untagged of each RIP-Seq for genes with count per million >0.5. Enrichment of genes in each tagged aaRS data set, against untagged and total RNA controls, was set to fold enrichment >2, adjusted *p-value* <0.05. An enriched gene list for each aaRS was constructed from the intersection of enrichments against both controls (the Input sample and the Bound of the untagged strain) (Supplementary Table 3). The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB44410 (https://www.ebi.ac.uk/ena/browser/view/PRJEB44410).

### RIP-qPCR analysis

RIP was done as described in RIP-seq method section, with two independent biological repeats. RNA from Input and Bound samples (500 ng) was reverse-transcribed using SuperScript™ II kit (Invitrogen) according to the manufacturer's instructions. Gene-specific levels were determined in a 20 µl reaction volume in triplicate using a Fast SYBR Green PCR Master Mix® (Applied Biosystems) two-step RT-PCR method following the manufacturer's instructions using primers for the indicated genes (Supplementary Table 8). All qPCR reactions used the following parameters: 95°C for 3 min, and then 3 sec at 95°C and 30 sec at 60°C for 40 cycles. Results were analysed with Quantstudio™ Design & Analysis. Enrichment was calculated using $2^{-(\Delta Ct)}$ between the Bound and Input of each aaRS.

### Data analyses

Clustering was done using iDEP webserver [42], by K-means of 6 (using different number of groups did not yield significant differences). GO terms enrichment analysis was performed in PANTHER (analysis type: PANTHER Overrepresentation Test (Release 2020–07-28), Annotation Version: PANTHER version 16.0 (Released 2020–12-01)) using the PANTHER GO-Slim Biological Process annotation

data set [31]. Sequence motif enrichment was performed by using DREME sequence alignment tool [32].

### *Identification of tRNA anticodon stem loop (ASL)-like structures in yeast mRNAs*

Publicly available nucleotide-level secondary structure data and associated FASTA sequences were obtained for >3000 *S. cerevisiae* transcripts [34]. Secondary structure data was used as input to develop a model for structural motif analysis using patteRNA [33] with log transformation option (-l). The resultant model was used to search for the tRNA anticodon loop consensus motif '(((((. . . . .)))))' within yeast transcripts. Motif sequences containing only 'N' were filtered from the motif search results. For a given tRNA isodecoder, the full list of motifs was searched to identify motif sequences containing the appropriate anticodon sequence (e.g. 'GAA' for $tRNA^{GAA}$-Phe) found within position 7–11 using the stringR str_locate function [43,44]. Anticodon loop trinucleotide sequences ± 7 nt (17 nt total) were obtained from sacCer3 mature tRNA isodecoder sequences [30] and aligned to the motif sequences using R Biostrings pairwiseAlignment function with default settings [45]. Alignments were cross-referenced with cognate aaRS RIP-seq gene hits.

## Acknowledgments

We thank members of our labs for advice and support, and members of the Nancy and Stephen Grand Israel National Centre for Personalized Medicine, at the Weizmann Institute of Science for help in RNA-seq analyses.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was funded by Israel Science Foundation (258/18) (Y.S.A.) and by Israel Michigan University Partnership (Y.S.A. and A.A). O.L. is a recipient of the Jacobs fellowship of outstanding students. A.A. is supported by a grant from the National Institute of General Medical Sciences [GM136441].

## References

[1] Perona JJ, Gruic-Sovulj I. Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. Top Curr Chem. 2014;344:1–41.
[2] Gomez MAR, Ibba M. Aminoacyl-tRNA synthetases. RNA. 2020;26:910–936.
[3] Giegé R, Eriani G. Transfer RNA recognition and aminoacylation by synthetases. eLS. Chichester, UK: John Wiley & Sons, Ltd.; 2014.
[4] Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. 1998;26. DOI:10.1093/nar/26.22.5017
[5] Beuning PJ, Musier-Forsyth K. Transfer RNA recognition by aminoacyl-tRNA synthetases. Biopolymers. 1999;52:1–28.
[6] Kaiser F, Krautwurst S, Salentin S, et al. The structural basis of the genetic code: amino acid recognition by aminoacyl-tRNA synthetases. Sci Rep. 2020;10. DOI:10.1038/s41598-020-69100-0
[7] Yao P, Poruri K, Martinis SA, et al. Non-catalytic regulation of gene expression by aminoacyl-tRNA synthetases. Top Curr Chem. 2014;344:167–187.
[8] Levi O, Garin S, Arava Y. RNA mimicry in post-transcriptional regulation by aminoacyl tRNA synthetases. Wiley Interdiscip Rev RNA. 2020;11. DOI:10.1002/wrna.1564
[9] Pang YLJ, Poruri K, Martinis SA. tRNA synthetase: TRNA aminoacylation and beyond. Wiley Interdiscip Rev RNA. 2014;5:461–480.
[10] Romby P, Brunel C, Caillet J, et al. Molecular mimicry in translational control of E. coli threonyl-tRNA synthetase gene. Competitive inhibition in tRNA aminoacylation and operator-repressor recognition switch using tRNA identity rules. Nucleic Acids Res. 1992;20:5633–5640.
[11] Ryckelynck M, Giegé R, Frugier M. tRNAs and tRNA mimics as cornerstones of aminoacyl-tRNA synthetase regulations. Biochimie. Elsevier; 2005:835–845. DOI:10.1016/j.biochi.2005.02.014
[12] Jeong SJ, Park S, Nguyen LT, et al. A threonyl-tRNA synthetase-mediated translation initiation machinery. Nat Commun. 2019;10:1357.
[13] Levi O, Arava Y. mRNA association by aminoacyl tRNA synthetase occurs at a putative anticodon mimic and autoregulates translation in response to tRNA levels. PLoS Biol. 2019;17. DOI:10.1371/journal.pbio.3000274
[14] Levi O, Arava YS. Pseudouridine-mediated translation control of mRNA by methionine aminoacyl tRNA synthetase. Nucleic Acids Res. 2021;49:432–443.
[15] Magrath C, Hyman LE. A mutation in GRS1, a glycyl-tRNA synthetase, affects 3′-end formation in Saccharomyces cerevisiae. Genetics. 1999;152:129–141.
[16] Lamech LT, Mallam AL, Lambowitz AM. Evolution of RNA-protein interactions: non-specific binding led to RNA splicing activity of fungal mitochondrial Tyrosyl-tRNA synthetases. PLoS Biol. 2014;12:e1002028.
[17] Motzik A, Nechushtan H, Foo SY, et al. Non-canonical roles of lysyl-tRNA synthetase in health and disease. Trends Mol Med. 2013;19:726–731.
[18] Sissler M, González-Serrano LE, Westhof E. Recent advances in mitochondrial aminoacyl-tRNA synthetases and disease. Trends Mol Med. 2017;23:693–708.
[19] Bervoets S, Wei N, Erfurth ML, et al. Transcriptional dysregulation by a nucleus-localized aminoacyl-tRNA synthetase associated with Charcot-Marie-Tooth neuropathy. Nat Commun. 2019;10. DOI:10.1038/s41467-019-12909-9
[20] Yakobov N, Debard S, Fischer F, et al. Cytosolic aminoacyl-tRNA synthetases: unanticipated relocations for unexpected functions. Biochim Biophys Acta - Gene Regul Mech. 2018;1861:387–400.
[21] Antonellis A, Green ED. The role of aminoacyl-tRNA synthetases in genetic diseases. Annu Rev Genomics Hum Genet. 2008;9:87–107.
[22] Wei N, Zhang Q, Yang XL. Neurodegenerative Charcot–Marie–Tooth disease as a case study to decipher novel functions of aminoacyl-tRNA synthetases. J Biol Chem. 2019;294:5321–5339.
[23] Garin S, Levi O, Cohen B, et al. Localization and RAN binding of mitochondrial aminoacyl tRNA synthetases. Genes (Basel). 2020;11:1–20.
[24] Guo M, Yang X-L, Schimmel P. New functions of aminoacyl-tRNA synthetases beyond translation. Nat Rev Mol Cell Biol. 2010;11:668–674.
[25] Huh WK, Falvo JV, Gerke LC, et al. Global analysis of protein localization in budding yeast. Nature. 2003;425:686–691.
[26] Crepin T, Yaremchuk A, Tukalo M, Cusack S. Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain. Structure. 2006 Oct;14(10):1511–25. doi: 10.1016/j.str.2006.08.007.PMID: 17027500
[27] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15. DOI:10.1186/s13059-014-0550-8
[28] Simos G, Sauer A, Fasiolo F, et al. Domain within Arc1p delivers tRNA to aminoacyl-tRNA synthetases. Mol Cell. 1998;1:235–242.
[29] Frechin M, Enkler L, Tetaud E, et al. Expression of nuclear and mitochondrial genes encoding ATP synthase is synchronized by

disassembly of a multisynthetase complex. Mol Cell. 2014;56:763–776.
[30] Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 2016;44:D184–9.
[31] Mi H, Ebert D, Muruganujan A, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2021;49: D394–403.
[32] Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics. 2011;27:1653–1659.
[33] Ledda M, Aviran S. PATTERNA: transcriptome-wide search for functional RNA elements via structural data signatures. Genome Biol. 2018;19:28.
[34] Kertesz M, Wan Y, Mazor E, et al. Genome-wide measurement of RNA secondary structure in yeast. Nature. 2010;467:103–107.
[35] Shi Y, Wei N, Yang XL. Studying nuclear functions of aminoacyl tRNA synthetases. Methods. 2017;113:105–110.
[36] Mirande M. The aminoacyl-tRNA synthetase complex. Subcell Biochem. 2017;83:505–522.
[37] Levi O, Arava YS. RNA modifications as a common denominator between tRNA and mRNA. Curr Genet. 2021. DOI:10.1007/s00294-021-01168-1
[38] Brosius J. Transmutation of tRNA over time. Nat Genet. 1999;22:8–9.
[39] Richter JD, Coller J. Pausing on polyribosomes: make way for elongation in translational control. Cell. 2015;163:292–300.
[40] Neelagandan N, Lamberti I, Carvalho HJF, et al. What determines eukaryotic translation elongation: recent molecular and quantitative analyses of protein synthesis. Open Biol. 2020;10:200292.
[41] Melamed D, Eliyahu E, Arava Y. Exploring translation regulation by global analysis of ribosomal association. Methods. 2009;48:301–305.
[42] Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinformatics. 2018;19:1–24.
[43] R Core Team. R: a language and environment for statistical computing. 2020;
[44] Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4:1686.
[45] Pages H, Aboyoun P, Gentleman R, et al. Biostrings: efficient manipulation of biological strings. 2020;