Europe PMC Funders Group
Author Manuscript
Nat Struct Mol Biol. Author manuscript; available in PMC 2018 January 09.

Published in final edited form as:
Nat Struct Mol Biol. 2015 December ; 22(12): 1027–1033. doi:10.1038/nsmb.3128.

Conserved mRNA-binding proteomes in eukaryotic organisms

Ana M. Matia-González, **Emma E. Laing**, and **André P. Gerber**
Department of Microbial and Cellular Sciences, School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

Abstract

RNA-binding proteins (RBPs) are essential for the post-transcriptional regulation of gene expression. Recent high-throughput screens have dramatically increased the number of experimentally identified RBPs; however, comprehensive identification of RBPs within living organisms is elusive. Here we describe the repertoire of 765 and 594 proteins that reproducibly interact with polyadenylated mRNAs in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, respectively. Furthermore, we report the differential association of mRBPs upon apoptosis induction in *C. elegans* L4 stage larvae. Strikingly, most proteins comprising mRNA-binding proteomes (mRBPomes) are evolutionarily conserved between yeast and *C. elegans*, including components of early metabolic pathways and the proteasome. Based on our evidence that glycolytic enzymes bind to distinct glycolytic mRNAs, we speculate that enzyme-mRNA interactions relate to an ancient mechanism for post-transcriptional coordination of metabolic pathways, perhaps established during the transition from the early RNA to the protein 'world'.

RNA-binding proteins (RBPs) are key players in post-transcriptional gene regulation, performing essential functions to maintain cellular fitness1,2. In particular, mRNA binding proteins (mRBPs) mediate the processing of mRNA-precursors (pre-mRNAs) in the nucleus, the export and localization of mRNAs to distinct subcellular regions in the cytoplasm, and the translation and eventual decay of mRNAs1. RBPs often bear characteristic RNA-binding domains (RBDs), such as the RNA-recognition motif (RRM), the hnRNP K homology (KH) domain or zinc-finger (ZnF) domains; many of them believed to originate at early stages of evolution1,3. Based on these RBDs, hundreds of RBPs were predicted to exist in eukaryotic organisms *in silico*3. However, recent systematic approaches to experimentally map mRBPs in yeast and human cultured cells, using protein microarrays4,5 or capturing of *in vivo* crosslinked mRNA-protein complexes followed by mass-spectrometry (MS)6–9, suggested that many proteins lacking canonical RBDs bind RNAs, among them proteins with other

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author: André P. Gerber, Department of Microbial and Cellular Sciences, University of Surrey, Guildford, United Kingdom.

Accession Codes

Proteomics data are deposited at the ProteomeXchange Consortium43 with project accession numbers PXD002293 and PXD002226 corresponding to the *S. cerevisiae* and *C. elegans* mRBPomes, respectively.

Author Contributions

A.M.M.G. and A.P.G. conceived and designed the experiments. A.M.M.G. performed laboratory experiments. E.E.L. performed bioinformatics analyses. All authors analyzed data, discussed the results, and wrote the manuscript.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

well-established cellular functions, such as metabolic enzymes. Thus, to gain a full system-wide understanding of post-transcriptional regulation, it is critical that mRBPs are characterized *in vivo*. Although unicellular eukaryotes such as yeast comprise many of the regulatory pathways seen in higher eukaryotes, identification of mRBPs in a living animal is vital, since it could reveal additional key regulatory factors that mediate post-transcriptional control essential for development and physiology.

We have thus developed a technique for the *in vivo* capture of mRNA-protein complexes to provide the first comprehensive catalogue of mRBPs in the unicellular yeast *Saccharomyces cerevisiae* and in the multicellular nematode *Caenorhabditis elegans,* respectively; and we show dynamic changes of the mRNA binding proteome (mRBPome) upon apoptotic stress in *C. elegans*. We find that mRBPomes are remarkably conserved between organisms, including many, if not all, components of early metabolic pathways and protein complexes, pointing to an ancient origin of the mRBPome. Finally, we show that yeast glycolytic proteins can bind to their own and other glycolytic mRNAs, potentially establishing an early RNA regulon, which, we hypothesize, could create an additional layer for coordinated control of pathway activity to achieve metabolic control.

Results

Defining the *Saccharomyces cerevisiae* mRBPome

To identify the sets of proteins interacting with mRNAs within living unicellular (yeast) and multicellular (*C. elegans*) eukaryotic organisms, we adapted a protocol recently applied to cultured human and starved yeast cells6–8. *S. cerevisiae* cells were grown in rich medium to mid-log phase and proteins were crosslinked *in vivo* to nucleic acids by UV-irradiation at 254 nm under conditions that preserved the integrity of RNA (Supplementary Fig. 1a); polyadenylated (poly(A)) RNAs were subsequently purified from cell lysates via oligo$[dT]_{25}$ beads under stringent washing conditions (see Methods). Since UV-exposure also crosslinks proteins to DNA10, we controlled for RNA-dependent binding of proteins by treating extracts with RNase ONE, and selective binding to poly(A) RNAs was tested by the addition of excess competitor polyadenylic acids to extracts prior to mRNA isolation. A silver-stained gel showed selective enrichment of proteins in UV-crosslinked poly(A) mRNA eluates compared to the input extract, but no proteins or RNA were detectable in our control samples (Fig. 1a). Likewise, immunoblot analysis showed that Scp160p, a well-known yeast RBP11, was present in the eluates; but actin (Act1), a protein not thought to bind RNA, was not detectable (Fig. 1b). Neither protein was evident in any of the equally sized negative control eluates. Taken together, these results showed that we were able to selectively capture mRBPs *in vivo*.

To comprehensively identify the proteins bound to poly(A) mRNAs, defining the yeast mRBPome, we subjected samples from three independent experiments (biological replicates) to MS analysis. At the same time, we analyzed respective control samples to demarcate non-specific binders. We identified 765 proteins in at least two out of three replicates with a false-discovery rate (FDR) of less than 1% and represented by at least 2 different peptides (Fig. 1c; for details see Methods; raw data is given in **Supplementary Table 1**).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Our yeast mRBPome contains 86 (72 %) of the 120 RBPs that were previously identified in starved yeast cells8 ($P < 8\times10^{-55}$, hypergeometric test), and it significantly overlaps with the repertoire of yeast RBPs identified in screens for RBPs using protein microarrays or poly(A) purification from non-crosslinked cell extracts4,5(overlap of data is displayed in the Supplementary Fig. 2a). In agreement with a previous observation12, which is based on a list of 561 known and predicted proteins with RNA-related functions13 (205 of them present in our mRBPome, $P < 2\times10^{-60}$, hypergeometric test), the mRNAs coding for proteins of our mRBPome were significantly less stable, but highly expressed with a tendency towards higher mRNA copy numbers and ribosome occupancies, compared to proteins not present in our mRBPome (Supplementary Fig. 3a). Moreover, the proteins comprising the mRBPome tended to be more stable and abundant. In contrast, protein expression noise, which is a measure for the extent of the variation of protein expression from cell-to-cell, was significantly lower for the mRBPome compared to all other proteins, suggesting that mRBPs are uniformly expressed across a homogenous population of cells12. Finally, our experimentally determined mRBPome consisted of proteins that were localized in all cellular compartments (Supplementary Fig. 3b). The observed slight overrepresentation of cytoplasmic proteins is in keeping with our approach where we specifically captured mRNAs that are mainly associated with translation in the cytoplasm in rapidly grown yeast cells14.

We further categorized the 765 proteins based on Gene Ontology annotation (GO) retrieved from the *Saccharomyces* Genome Database15 (SGD) (Fig. 1d; complete analysis of significantly enriched GO terms from SGD at FDR < 5% is provided in **Supplementary Table 2a-c**). We identified approximately half (47%) of previously annotated RBPs (175 out of 370 proteins annotated as 'RNA binding', $P < 3\times10^{-70}$), with most of the proteins (59%) assigned to 'mRNA-poly(A) RNA binding' (102 out of 172 annotated proteins, $P < 8\times10^{-55}$), and all six proteins with known poly(A) binding function (Nab2, Npl3, Mtr4, Pab1, Sgn1, Tpa1). Nevertheless, most of the proteins in the mRBPome (73%) have not been previously linked to RNA-binding functions and thus likely represent new RBPs or proteins with dual functions: this includes 325 enzymes, proteins with an annotated enzyme classification (EC) number in UniProt (42% of the mRBPome, $P < 0.001$ at FDR < 5%, hypergeometric test, see Methods), most of them with functions in metabolic processes such as hydrolases, oxidoreductases, and ligases.

We analyzed the yeast mRBPome for the occurrence of protein domains as annotated in the UniProt database16. The mRBPome was significantly enriched in prominent classical RNA-binding domains (RBDs) such as the hnRNPK homology domain (KH_dom), the RNA-recognition motif (RRM_dom) and the Sm domain (Ribonucl_LSM) (Fig. 1e). Less frequent domains, such as the S1 domain (4 of 7 the annotated proteins) and the La domain (all 3 annotated proteins), were also enriched in the mRBPome but did not reach statistical significance by the chosen statistical test (complete domain analysis is given in **Supplementary Table 2d**). Interestingly, this analysis also revealed significant enrichment of additional domains that have not been previously linked to RNA-binding or that occur in proteins devoid of classical RNA-binding domains. This includes the P-loop NTPase domain, a common motif in ATP and GTP binding proteins; the heat shock protein 70

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

domain and the chaperonin Cnp60–TCP-1 domain (Fig. 1e). Finally, we searched the yeast mRBPome for the overrepresentation of short amino-acid sequence stretches: 707 4-mer and 64 5-mer amino-acid sequences were significantly enriched in the yeast mRBPome compared to the reference proteome ($P < 0.05$ at FDR < 5%, hypergeometric test), excluding statistical bias from repeated motifs in the same protein (complete motif analysis is given in the **Supplementary Table 2e,f**). Most of these motifs contain arginine (R), lysine (K), tyrosine (Y) and glycine (G) residues (47 of 64 5-mers, i.e. GGRGG, RGGRG, GTGKT), which are also present in known RBPs and could relate to disordered regions that interact with RNA6.

Identification of *C. elegans* mRNA-interacting proteins

We next sought to decipher the mRBPome in a living animal, the nematode *C. elegans*. We thus applied UV-crosslinking and poly(A) purification to capture the *in vivo* mRBPome of mixed-stage animals, representing all stages of the *C. elegans* life cycle, and in synchronized animals at the fourth (final) larval stage (L4). At the latter stage, important cellular processes such as spermatogenesis and apoptosis (programmed cell death) take place, which are controlled substantially by post-transcriptional events17. Hence, we also investigated the mRBPome upon induction of germline apoptosis in synchronized L4 stage larvae with 5 mM N-nitroso-N-ethylurea (ENU), a potent alkylating mutagen that induces apoptosis18. The successful induction of apoptosis was monitored by the expression of *egl-1* mRNA (Supplementary Fig. 4a), a marker of germline apoptosis19.

Analogous to yeast, the animals were UV-exposed under conditions that preserved the integrity of the RNA (Supplementary Fig. 1b), and poly(A) mRNAs were isolated (see Methods). We found that the composition of proteins in the eluates of poly(A) selected mRNAs was dissimilar to that of cell lysates, indicating the selective enrichment of mRBPs (Fig. 2a). To further confirm this, we used GLD-1, a KH domain protein that modulates a number of key decisions in the *C. elegans* germline20, as a positive control and cytochrome C (CYC-1), a highly expressed mitochondrial protein not thought to bind RNA, as a negative control. As expected, GLD-1 was enriched in our mRNA isolates but not in the control samples and CYC-1 was not present in eluates (Fig. 2b).

We subjected three independent experiments (comprised of ~120,000 animals each) that constitute three biological replicates of mixed-staged, L4 and matched L4+ENU treated animals to MS analysis and analyzed respective control samples to demarcate non-specific binders (see Methods). Data was highly correlated between replicates of mixed-staged (Pearson correlation, $r = 0.89\text{-}0.97$) and in L4 and L4+ENU matched samples ($r = 0.95\text{-}0.98$); but somewhat less preserved between biological replicates of the L4 and L4+ENU treated animals ($r = 0.75\text{-}0.92$), which is possibly due to slight differences in the populations of L4 staged larvae (a scatter plot comparing samples is given in the Supplementary Fig. 4b). Applying likewise refinements we used for yeast, we identified 594 proteins in at least two out of the nine samples with a FDR < 1% and at least 2 different peptides (see Methods, raw data is given in the **Supplementary Table 3a**). We considered this entire set of 594 proteins to be part of the *C. elegans* mRBPome, a significant proportion (136 proteins, 23%, $P < 1\times10^{-64}$, hypergeometric test) of which was found to overlap with a

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

recent computationally predicted *C. elegans* RBPome21 (for comparison see Supplementary Fig. 2b).

Considering our stringent selection criteria, 93.5% of the experimentally captured *C. elegans* mRBPome was identifiable in mixed-stage animals (555 proteins), reflecting a comprehensive repertoire of mRBPs expressed during the entire nematode life cycle (Fig. 2c, a Venn diagram showing occurrence of proteins across samples is provided in the Supplementary Fig. 4c). The remaining 39 proteins (6.5% of mRBPome) were exclusively confined to L4-staged animals and include known RBPs with roles in germline development, such as FBF-2, a PUF family RNA-binding protein that negatively regulates *fem-3* and *gld-1* expression; ALG-4 a member of the argonaute proteins involved in male germline fertility; and CPB-2, a cytoplasmic polyadenylation dependent binding protein homolog specifically expressed in the germatogenic germline22. Moreover, GST-5 has known roles in germline apoptosis23 and was exclusively recorded in apoptosis induced animals. To further search for differences between L4 and L4+ENU treated animals, we performed relative quantification of MS data (see Methods). 96 proteins were differentially enriched in these mRBPomes at an FDR < 5% (Fig. 2d; complete dataset is provided in the **Supplementary Table 3b**). Of these, 41 proteins were more predominant in L4 and 55 proteins in apoptosis induced L4-stage animals, respectively. The former group was enriched in proteins involved in ‘cellular response to stress’ (Westfall and Young familywise error rate (FWER) adjusted $P < 0.01$, see Methods) and ‘eukaryotic translation elongation’ (Westfall and Young FWER adj. $P < 0.05$), including EEF-1B.1 and EEF-1B.2. Conversely, the latter group contained proteins linked to apoptosis, germline development or stress response, such as EIF-3.K, a translation initiation factor which positively regulates *ced-3* to promote programmed cell death24; EIF-6 required for control of cell division in the germ line25; and CTL-1, PRDX-3, both involved in the oxidative stress response22. Remarkably, six translation initiation factors were prevalent in ENU treated animals, namely EGL-45, EIF-3.I, EIF-3.K, EIF-6, IFE-2, and CLU-1 (Westfall and Young FWER adj. $P < 0.01$), indicating potential for selective remodeling of translation during apoptotic stress.

GO enrichment analysis revealed that, as one would expect, a significant fraction of the proteins comprising the *C. elegans* mRBPome had roles in RNA metabolism, i.e. ‘mRNA metabolic process’ ($P < 0.001$; all *P*-values are Westfall and Young FWER adj.), ‘RNA binding’ ($P < 0.001$), ‘ribonucleoprotein complex biogenesis’ ($P < 0.001$), and ‘translation’ ($P < 0.001$) (Fig. 2e, the full list of enriched GO terms is given in the **Supplementary Table 4a-c**). Furthermore, many mRBPs were linked to the control of development, i.e. ‘developmental process’ ($P < 0.001$), ‘embryo development’ ($P < 0.05$) and ‘regulation of cell death’ ($P < 0.001$), reminiscent of the increasingly recognized role of post-transcriptional control in these processes. As seen in yeast, many proteins of the *C. elegans* mRBPome were EC annotated enzymes (227 proteins, 38 % of the mRBPome, $P < 0.001$ at 5% FDR, hypergeometric test). They preferentially act in metabolic processes, e.g. ‘primary metabolic process’, ‘carbohydrate derivate binding’, ‘glycolytic process’ and ‘tricarboxylic acid (TCA) or Krebs cycle’ (all terms with Westfall and Young FWER adj. $P < 0.001$), further supporting the idea that metabolic enzymes could have dual functions.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Protein domain analysis revealed significant over-representation of classical RNA-related protein domains, such as the DEAD-DEAH-box helicase or zinc finger (e.g. Znf_CCHC) domains (e.g. Znf_CCHC) (Fig. 2f, full domain analysis is given in the **Supplementary Table 4d**). However, certain RBDs were less well represented, such as KH and RRM domains. The reason for the absence of many proteins bearing theses domains is not known, but the respective domains could target other types of RNA and (or) have other functions, e.g. acting as protein-protein interaction sites26. A search for short amino acid motifs that are significantly enriched in the mRBPome identified 184 4-mers and 228 5-mer motifs ($P <$ 0.05 at FDR < 5%, hypergeometric test) most of them bearing G, R, K, or Y residues (extended motif analysis is provided in **Supplementary Table 4e,f**). Of note, 33 4-mer motifs and seven 5-mer motifs (RGGRG, GGRGG, DEAVA, TITND, GTGKT, LGGGT, QATKD) were also significantly enriched in the yeast mRBPome (**Supplementary Table 5a,b**).

Evolutionary conservation of the mRBPome

We next assessed whether our experimentally determined mRBPomes are evolutionarily conserved. Therefore, protein homology information for *S. cerevisiae* and *C. elegans* were retrieved from InParanoid27, which revealed 1,841 orthologous protein pairs. Thereby, 7% of all *C. elegans* proteins in the reference proteome comprising 26,165 proteins matched a homologous yeast protein and inversely, 28% of all *S. cerevisiae* proteins in the 6,621 protein reference proteome had an orthologous protein in *C. elegans* (Fig. 3a). Strikingly, more than half of the *C. elegans* mRBPome matched a *S. cerevisiae* ortholog (330 proteins, 56%), and conversely, almost two-thirds of the yeast mRBPome had a homolog in *C. elegans* (476 proteins, 62.3%). Furthermore, we found a highly significant overlap of both mRBPomes orthologous pairs of proteins, suggesting high conservation of mRBPomes (Fig. 3b). For instance, the glycolytic pathway, which is significantly enriched in yeast ($P < 7\times10^{-27}$) and *C. elegans* ($P < 2\times10^{-7}$, *P*-values from Wikipathway analysis, see Methods) mRBPomes was contained in the intersection representing conserved orthologous mRBPs (Fig. 3c, complete analysis of orthologous proteins in mRBPomes is given in the **Supplementary Table 5c-f**). Moreover, distinct proteins of large macromolecular complexes, such as the proteasome are conserved and significantly enriched in the yeast ($P < 7\times10^{-11}$) and in the *C. elegans* ($P < 2\times10^{-19}$) mRBPomes (Wikipathway analysis). Interestingly, we found that the components of the 26S proteasome interact with mRNAs aligned along the surface and cover the cavity through which ubiquitinated proteins are degraded (Fig. 3d).

Validation of novel mRBPs

To validate novel and conserved mRBPs we adapted a fluorescence based mRNA-protein interaction assay to yeast28 (see Methods). Proteins were UV-crosslinked to mRNAs *in vivo* and endogenously GFP-tagged proteins were immunopurified (IP) from cell lysates under stringent conditions. The abundance of proteins and bound mRNAs was then assessed via GFP and hybridization with fluorescently labeled oligo$[dT]_{25}$, respectively (Fig. 4). As expected, we detected poly(A) mRNAs in the IP of three known mRBPs (Gis2, Khd1 and Pab1) but no RNA-dependent associations were seen with untagged wild-type (WT) control cells nor Ras1p, a GTPase not known to bind RNA. We further confirmed mRNAs bound to

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Sbp1p and YGR250p, two predicted mRBPs bearing classical RBDs; the glycolytic proteins Eno1 and Pfk2; the pentose phosphate pathway enzymes Tal1p and Tkl1p; and the conserved proteasome components Pre10p and Rpt1p. Of note, no statistically significant correlation was seen between the abundance of proteins (GFP) and the fluorescent signals inferred from the bound mRNAs (untreated: $r^2 = 0.11$; $P = 0.29$; RNase treated: $r^2 = 0.05$, $P = 0.48$; r^2 following fitting of linear model), showing that protein abundance is not a predictor for the amount of mRNAs bound. Intrigued by our finding that all yeast glycolytic enzymes were part of the mRBPome, and what their specific mRNA targets could be, we speculated that they may bind to glycolytic mRNAs, providing potential for post-transcriptional pathway coordination4. Indeed, we detected mRNAs coding for several glycolytic proteins in the IPs of GFP and (or) tandem affinity purification (TAP)-tagged glycolytic enzymes (Pfk2, Gpm1, Eno1), whereas unrelated mRNAs (Act1, Ash1) were absent (Fig. 5). No mRNAs were detected in control isolates from untagged strains (mock). Furthermore, association of Pfk2p and Eno1p with glycolytic mRNAs was recapitulated in affinity isolates obtained from non-crosslinked cells; the observed interactions were not ribosome dependent due to potential co-translational assembly29, as evidenced by the treatment of extracts with puromycin, a compound that disassembles ribosomes (Supplementary Fig. 5). In conclusion, these results strongly suggest that some glycolytic enzymes bind to their own mRNAs as well as other selected mRNAs for glycolytic proteins (no physical or genetic interactions are reported between Eno1, Gpm1 and Pfk215).

Discussion

Our experimentally identified mRBPomes notably enhance and broaden the repertoire of known mRBPs in nature, particularly as we describe the first experimentally captured mRBPomes of living whole-organisms, and the deployment of different mRBPs in apoptosis. Whilst a significant proportion of our identified mRBPs are in agreement with previously reported data, we identified highly novel mRBPs, characteristics (e.g. enzymes), RNA-binding related protein domains and amino acid motifs, providing annotation for hundreds of new mRBPs (Fig. 1,2).

We also observed the differential arrangement of mRBPs upon apoptosis induction in *C. elegans* L4 stage larvae, suggesting stress-specificity within mRBPomes of animals: 16 % (96 proteins) of the 583 detectable mRBPs in L4 staged animals exhibited significantly altered abundance within the mRBPome upon induction of apoptotic stress (Fig. 2d). Interestingly, apoptotic stress decreased the association of translation ‘elongation’ factors within the mRBPome, whereas several ‘initiation’ factors became enriched, which suggests active remodeling of translation. Whether this relates to translation regulation of a subset of mRNAs by one or several initiation factors 30, or to changes of translation elongation rates, remains to be investigated. Monitoring of the dynamic responses of the mRBPome may thus establish a valuable new approach to identify potential post-transcriptional regulators in apoptosis or other physiological conditions, and in disease.

In addition to many of the annotated canonical RNA binding domains, we found strong enrichment of other motifs in both mRBPomes, some of these may, or may not, act as RBDs. For example, the significant enrichment of the heat shock protein 70 domain and the

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

chaperonin Cnp60–TCP-1 domain in both mRBPomes is reminiscent of previous observations that certain heat shock proteins could act as RNA-binding entities *in vivo* and guide the folding of RNA substrates for subsequent degradation and translation31,32. Conversely, domain analysis also revealed limitations to our chosen UV-crosslinking approach. For instance, proteins bearing double-stranded RBDs (dsRBDs) were underrepresented in both mRBPomes: of the 6 annotated proteins containing dsRBDs in yeast, only Rps2p was identified; and only 3 of the 18 annotated dsRBDs proteins in *C. elegans* (RPS-2, MRPS-5 and RHA-1), all of them containing additional RBDs, were reproducibly detected. These results may recapitulate inefficiency of UV-crosslinking of proteins interacting with extended dsRNA in A-helical structures, and other methods may be used to capture dsRBPs with high efficiency33. On this line, UV light crosslinks proteins to bases in RNA, preferentially uridine, if they are in close proximity for a certain time. This could explain the capturing of RNA catalytic enzymes, such as *C. elegans* GLD-2, which is an enzyme that polyadenylates *gld-1* mRNA but is not reported to directly bind to this mRNA as it needs GLD-3 as a bridge34. It is thus possible that we also identified proteins that, in current terms, ‘indirectly’ interact with mRNA due to the proximity of reactive centers to RNA.

We found significant conservation of experimentally determined mRBPs, suggesting elementary functions for the underlying protein-mRNA interactions in cell homeostasis (Fig. 3). Bioinformatics analysis of all 1,841 orthologous *S. cerevisiae* and *C. elegans* proteins showed that proteins associated with the UniProt keywords ‘RNA binding’, ‘mRNA processing’ or ‘proteasome’ are significantly overrepresented ($P < 0.001$ at FDR < 5%, hypergeometric test) whereas ‘DNA binding’ proteins were not particularly enriched (complete analysis of orthologs is provided in **Supplementary Table 5g,h**). Our experimental data is thus in agreement with these *in silico* predictions, which probably reflect early establishment and sustainability of post-transcriptional regulators during evolution. In particular, we believe that conserved mRNA-binding properties of proteins comprising one of the oldest metabolic pathways such as glycolysis or protein complexes, such as the proteasome, could have emerged during the proposed transition from the RNA to protein world, exerting activity that may or may not have been lost in present cells.

The presence of proteasome subunits within both mRBPomes (Fig. 3d) is reminiscent of early findings that 19S prosomes, renamed to proteasome after its role in proteolysis was discovered35, may constitute an RNP complex involved in the negative control of mRNA translation36. Moreover, it was postulated that the 20S and 26S proteasome exhibits RNA endoribonuclease activity37, and that alpha subunits (α1, α5, α6, α7), which are present in both of our mRBPomes, exhibit endoribonuclease activity38. Inspired by the conceptual analogy between the proteasome and the RNA exosome39, we thus speculate that the proteasome may still play a role in mRNA degradation, possibly via the catalytic activity of the 20S.

The significant proportion of metabolic enzymes within both mRBPomes indicates the possibility of cross-talk between metabolism and RNA biology. It is known that several human enzymes selectively bind mRNAs, including the glycolytic proteins glyceraldehyde-3-phosphate dehydrogenase (GAPDH; termed Tdh1, Tdh2, Tdh3 in yeast),

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

enolase (ENOL-1; Eno1,2), aldolase (ALDO, Fba1) or phosphoglycerate kinase (PGK)40,41. However, we were particularly intrigued by our finding that all 16 enzymes that mediate the 10 steps of glycolysis bound mRNAs in yeast as do some of the orthologous proteins in *C. elegans* (Fig. 3c). Moreover, since we found that glycolytic proteins could bind both their own mRNA and other glycolytic mRNAs, we postulate that glycolysis may be additionally coordinated and (or) controlled by specific enzyme-mRNA interactions within the pathway. In one scenario, mRNA-protein interactions could activate or modify enzymatic activities; however, this may appear less likely in light of the tremendous excess of glycolytic enzymes (~100,000 copies) to mRNAs (~10 copies) in cells. In another scenario, metabolic enzymes may directly control glycolytic mRNAs in the cytoplasm, possibly affecting the localization, translation or decay. If so, we hypothesize that it could relate to an ancient mechanism for the coordination of pathway activity to achieve metabolic control – perhaps establishing an early RNA regulon, a structure defined by mRBPs to coordinate the expression of mRNAs encoding functionally related proteins42.

Online Methods

In vivo capture of mRBPs in *S. cerevisiae*

Strain BY4741 (MAT**a** *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*) was grown in 500 ml YPD media (1% yeast extract, 2% peptone, 2% D-glucose) at 30 °C with constant shaking at 220 r.p.m. Cells were collected at mid-log phase (OD_{600} ~ 0.6) by centrifugation and washed three times with phosphate-buffered-saline (PBS). For UV-crosslinking, the cells were resuspended in 25 ml of PBS and exposed to 1,200 mJ/cm^2 of 254 nm UV light in a Stratalinker 1800 (Stratagene) with two 2-min breaks on ice and gentle mixing. Cells were resuspended in 4 ml lysis buffer (100 mM Tris-HCl, pH 7.5, 500 mM LiCl, 10 mM EDTA, 1% Triton X-100, 5 mM DTT, 20 U ml^{-1} DNase I (Promega, M6101), 100 U ml^{-1} RNasin (Promega, N2611), complete EDTA-free protease-inhibitor cocktail (Roche, 11836170001)) and mechanically broken with glass beads in a Tissue Lyser (RETSCH MM200, Qiagen) for 10 min at 30 Hz at 4 °C. The cooled lysate was cleared by three sequential centrifugations at 4 °C at 3,000 *g* for 3 min, and 5,000 *g* and 10,000 *g* for 5 min each. Two negative controls were introduced at this stage: the extract was either supplemented with 100 U RNase ONE (Promega, M4265) for 2 h at 37 °C to digest RNA; or 20 mg of poly(A) (Sigma, P9403) was added to the extract for competition experiments. To control for the integrity of RNA, total RNA was isolated from 50 μl of extracts with the ZR RNA MiniPrep kit (Zymo Research, R1065). RNA was quantified with a Nanodrop ND-2000 device and visualized on a 1% agarose gel stained with Red-safe (iNtRON BIOTECHNOLOGY, 21141). The presence of mRNAs in poly(A) eluates was controlled by reverse-transcription (RT) PCR (data not shown). To capture polyadenylated RNAs for mRBPome analysis, one milligram of oligo$[dT]_{25}$ Dynabeads (Life Technologies, 61011) was equilibrated in lysis buffer, mixed with the extracts (~5 mg) and incubated on a shaker for 10 min at room temperature (RT). The beads were collected with a magnet and the supernatant recovered for repeat incubations (see below). The beads were washed once with 500 μl of wash buffer A (10 mM Tris-HCl, pH 7.5, 600 mM LiCl, 1 mM EDTA, 0.1% Triton X-100) and twice with 500 μl wash buffer B (10 mM Tris-HCl, pH 7.5, 600 mM LiCl, 1 mM EDTA). The poly(A) RNA was eluted from beads in 60 μl of 10 mM Tris-HCl, pH 7.5 at 80 °C for 2 min and collected.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

The entire procedure was repeated twice by reapplying the supernatant to the oligo[dT]$_{25}$ beads, which were recovered after elution and washed three times with lysis buffer prior to reapplication. The three sequential eluates were combined and concentrated to 70 µl in a 0.5 ml Microcon-10 kD Centrifugal Filter Unit with Ultracel-10 membrane (Millipore, MRCPRT010). In total, we subjected 6 samples for MS analysis: three biological replicates of the mRBPome, two RNase ONE treated and one poly(A) competition experiment.

C. elegans cultures and apoptosis induction

Bristol N2 worms were cultured at 20 °C on NGM plates (0.3% NaCl, 1.7% agar, 0.25% peptone, 1 mM $CaCl_2$, 5 µg/ml cholesterol, 1 mM $MgSO_4$, 25 mM KPO_4 buffer, pH 6.0) seeded with OP50 strain *E.coli* according to standard procedures (http://www.wormbook.org). Synchronization was reached by bleaching gravid adults. To induce germ cell death, ~120,000 worms were synchronized in late L4 and treated with 5 mM of N-nitroso-N-ethylurea (ENU, Sigma, N3385) in 4 ml M9 buffer (0.3% KH_2PO_4, 0.6% Na_2HPO_4, 0.5% NaCl, 1 mM $MgSO_4$) in 15 ml falcon tubes for 4 h with constant rotation18.

In vivo capture of mRBPs in *C. elegans*

~120,000 worms were pelleted and washed three times in M9 buffer, followed by 15 min incubation with 15 ml M9 buffer on a rotatory wheel. Worms were transferred to a NGM plate and exposed to UV-light (254 nm) at 300 mJ/cm^2 in a Stratalinker 1800 (Stratagene) and collected in M9 buffer and resuspended in lysis buffer (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.75% IGEPAL, 1 mM DTT, 20 U ml^{-1} DNase I (Promega), 100 U ml^{-1} RNasin (Promega), complete EDTA-free protease-inhibitor cocktail (Roche)). To prepare extracts, the worms or larvae were grinded in a mortar filled with liquid nitrogen, and the lysate including the fat layer was subsequently clarified at 14,000 *g* for 10 min and passed through a 0.45 µm filter (Millipore). For the controls, worm extracts were either treated with 100 U RNase ONE (Promega) or 20 mg of poly(A) (Sigma) was added for competition experiments. The poly(A) RNA and crosslinked RBPs were isolated from 1.25 ml (~12 mg) of mixed-stage and 1 ml (~10 mg) of synchronized L4 worm extracts in three sequential rounds as described for yeast with minor modification: the volumes of beads and buffers were up-scaled in relation to the amount of input extract; and 500 mM LiCl was used in wash buffers. In total, we submitted 18 samples for MS analysis: three independent experiments (three biological replicates) containing ~120,000 animals of mixed staged worms as well as two RNase ONE treated and one poly(A) competitor control respectively; three independent experiments (three biological replicates) of synchronized larvae stage 4 (L4) and the three matched experiments of synchronized L4 treated with 5 mM ENU along with their corresponding three poly(A) competition control experiments.

LC-MS/MS and protein identification

MS analysis was performed at the Proteomics Facility, University of Bristol. 50 µl of the samples (~ 8-10 µg of protein) were run on a 4–12% NuPAGE Novex acrylamide gel (Life Sciences). The gel lane was cut into one slice and subjected to in-gel tryptic digestion using a ProGest automated digestion unit (Digilab UK). The resulting peptides were fractionated using a Dionex Ultimate 3000 nanoHPLC system in line with an LTQ-Orbitrap Velos mass

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

spectrometer (Thermo Scientific). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Dionex). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid peptides were resolved on a 250 mm × 75 μm Acclaim PepMap C18 reverse phase analytical column (Dionex) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1min., 6-15% B over 58 min, 15-32% B over 58min, 32-40% B over 3 min, 40-90% B over 1 min, held at 90% B for 6 min and then reduced to 1% B over 1 min) with a flow rate of 300 nl min^{-1}. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.1 kV using a stainless steel emitter with an internal diameter of 30 μm (Thermo Scientific) and a capillary temperature of 250 °C. Tandem mass spectra were acquired using an LTQ- Orbitrap Velos mass spectrometer controlled by Xcalibur 2.1 software (Thermo Scientific) and operated in data-dependent acquisition mode. The Orbitrap was set to analyze the survey scans at 60,000 resolution (at m/z 400) in the mass range m/z 300 to 2,000 and the top twenty multiply charged ions in each duty cycle selected for MS/MS in the LTQ linear ion trap. Charge state filtering, where unassigned precursor ions were not selected for fragmentation and dynamic exclusion (repeat count, 1; repeat duration, 30s; exclusion list size, 500), was used. Fragmentation conditions in the LTQ were as follows: normalized collision energy, 40%; activation q, 0.25; activation time 10 ms; and minimum ion selection intensity, 500 counts.

The raw data files were processed and quantified using Proteome Discoverer software v1.2 (Thermo Scientific) and searched against the SwissProt SPECIES database using the Mascot algorithm (Version 2.4). Peptide precursor mass tolerance was set at 10 ppm, and MS/MS tolerance was set at 0.8 Da. Search criteria included carbamidomethylation of cysteine (+57.0214) as a fixed modification and oxidation of methionine (+15.9949) as a variable modification. Searches were performed with full tryptic digestion and a maximum of one missed cleavage was allowed. The reverse database search option was enabled and all peptide data was filtered to satisfy an FDR threshold of < 1%.

To define a list of proteins referred to as the yeast mRBPome, we initially considered the 1,425 proteins that were enriched in at least one of the three biological replicates with a FDR of at least 1% at the protein identification and quantification level (raw data is given in **Supplementary Table 1**). Of these, we selected 973 proteins that were identified with at least two peptides, and for which data was obtained in two out of three biological replicates, leaving 765 proteins. 116 of these proteins were also identified in at least one of our three control samples (63 and 111 proteins in two RNase ONE treated controls, 16 proteins in poly(A) competition). However, the respective proteins were not excluded since peptides numbers were at least 2-fold higher in mRBP samples compared to the controls (except Pab1p, 1.6-fold).

To define the *C. elegans* mRBPome, we initially considered the 1,016 proteins that were enriched in at least one of the nine UV-crosslinked mRBPome samples with a FDR < 1% (raw data is given in **Supplementary Table 3a**). We further selected those proteins that were represented with at least two peptides (691 proteins) in at least two samples of the nine samples, revealing 594 proteins. 43 of them were also seen in at least one of our 9 negative control samples but not excluded from the mRBPome because the number of peptides for

respective proteins was at least 2-fold lower in the negative controls compared to mRBP samples.

C. elegans MS data processing and label-free quantification

Only data for proteins identified as having at least two peptides across samples was submitted for processing and analysis (peak areas for proteins not meeting our criteria were set to zero). For each protein within the experiment, the 'raw' MS peak area in the technical control sample was subtracted from the corresponding 'raw' MS peak area in the biological sample to produce 'background subtracted' peak areas. Subsequently, to normalize each background subtracted peak area, as a fraction of the total area measured by MS for a given sample, each background subtracted peak area in a sample was divided by the total background subtracted area for that sample (normalized data is given in the **Supplementary Table 3b**). An arbitrary value of 0.000001 was then added to all normalized peak areas to avoid zero values to enable the calculation of a fold change between conditions that include presence *vs.* absence. To determine the relative changes in protein abundance between *C. elegans* L4+ENU and L4 mRBPomes, we calculated $\log_2$ fold changes for each of the three replicates by $\log_2$ (L4+ENU_normalised_area_replicate_x / L4_normalised_area_replicate_x), where x is the same replicate number due to the paired nature of the experimental design. We then determined the average $\log_2$ fold change between L4+ENU and L4 treated animals across all three replicates. To estimate an FDR for declaring differentially abundant proteins between conditions we generated 10,000 data sets by randomly shuffling the experimentally obtained fold changes of replicates. From this randomized data set we then computed average $\log_2$ fold changes to assess, for each possible $\log_2$ L4+ENU/L4 threshold of -10 to 10 in 0.5 increments, the fraction of the 10,000 randomly generated fold changes that were above that threshold. From these values it was possible to linearly interpolate FDRs for all experimentally obtained (non-random) fold changes. Thus, a suitable $\log_2$ L4+ENU/L4 threshold to determine differential protein abundance with a false discovery rate of less than 5% can be defined as having less than 500 of the 10,000 randomly generated $\log_2$ L4+ENU/L4 values greater than that threshold (fold changes and FDRs are given in **Supplementary Table 3b**).

Immunoblot analysis

0.1% and 0.05% of the yeast and worm input extract respectively (~5 μg of protein) and 10% of the eluates were resolved on 4-15% SDS polyacrylamide gels and transferred to polyvinylidene difluoride (PVDF) membranes (Thermo Scientific Pierce). Membranes were blocked in PBS-0.1% Tween-20 containing 5% low fat milk, probed with designated antibodies and horseradish peroxidase (HRP)-coupled secondary antibodies, and developed with the Immobilon Western Chemiluminescent HRP Substrate (Millipore). Blots were recorded with a FluorChem (Alpha Innotech). The following antibodies were used: rabbit anti-Scp16011 (1:10,000), mouse anti-Act1 (1:2,500; MP Biomedicals, 0869100), mouse anti-GLD-144 (1:50), mouse anti-CYC-1 (1:1,000; Invitrogen, 456100), mouse anti-GFP (1:2,000; Roche, 11814460001), HRP-conjugated donkey anti-rabbit IgG (1:5,000; Amersham, NA9340V), HRP-conjugated sheep anti-mouse IgG (1:5,000; Amersham, NXA931). The validation of all commercial primary antibodies is provided on the manufacturer's website.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

IP of proteins from UV-crosslinked samples and fluorescence mRNA-binding assay

The fluorescence RNA-binding assay is based on a protocol28 that we adapted to yeast. Yeast strains expressing endogenously GFP-tagged proteins45, were verified by immunoblot analysis (data not shown). Strain BY4741 was used for mock-control affinity isolations. Cells were grown in 75 ml of YPD at 30 °C with constant shaking at 220 r.p.m. and collected at mid-log phase by centrifugation and washed with PBS. The cells were resuspended in 4 ml of PBS and exposed to UV light for crosslinking as described above. Cells were collected and resuspended in 650 µl of lysis buffer (100 mM Tris-HCl, pH 7.5, 500 mM LiCl, 10 mM EDTA, 1% Triton X-100, 5 mM DTT, 20 U ml^{-1} DNase I (Promega), 100 U ml^{-1} RNasin (Promega), complete EDTA-free protease-inhibitor cocktail (Roche)) and extracts were prepared as described above.

To IP GFP-tagged proteins, 500 µl cell extract (2.5 mg protein) was incubated with 20 µl of pre-equilibrated GFP-Trap_A agarose beads (Chromotek, gta-20) for 2 h by shaking at 1,000 r.p.m. in a Thriller Thermoshaker (PeqLab) at 4 °C. Beads were washed twice with 750 µl of wash buffer C [10 mM Tris-HCl, pH 7.5, 500 mM LiCl, 1 mM EDTA, 0.1% Triton X-100, 0.05% SDS, 50 U ml^{-1} RNasin (Promega), complete EDTA-free protease-inhibitor cocktail (Roche)] and collected by centrifugation at 2,500 *g* for 2 min at 4 °C. The beads were then resuspended in 800 µl of wash buffer C and half of the sample (400 µl) was incubated with 100 U ml^{-1} of RNase ONE (Promega), and the other half was left untreated; and samples were incubated for 30 min at 4 °C. Beads were washed three times with 750 µl of wash buffer C and incubated with 500 µl of blocking buffer (10 mM Tris-HCl, pH 7.5, 500 mM LiCl, 1 mM EDTA, 0.01% Triton X-100, 100 µg/ml *E. coli* tRNA, 100 µg/ml BSA, 50 U ml^{-1} RNasin (Promega), EDTA-free protease-inhibitor cocktail (Roche)) for 15 min on the shaker at 4 °C. After blocking, the samples were incubated with 500 µl of hybridization buffer (10 mM Tris-HCl, pH 7.5, 500 mM LiCl, 1 mM EDTA, 0.01% Triton X-100, 0.05% LiDS, 5 mM DTT, 100 U ml^{-1} RNasin (Promega)) supplemented with 40 nM of oligo$[dT]_{25}$ labeled with Alexa 594 (IDT) for 1 h in the dark on the shaker at 4 °C. Finally, samples were washed three times with 500 µl of wash buffer D (10 mM Tris-HCl, pH 7.5, 500 mM LiCl, 1 mM EDTA, 0.01% Triton X-100, 0.01% LiDS. 5 mM DTT, 50 U ml^{-1} RNasin (Promega), complete EDTA-free protease-inhibitor cocktail (Roche)). Beads were resuspended in 100 µl of wash buffer D and transferred to an opaque 96 well plate. Fluorescence measurements were performed in a FLUOstar Omega microplate reader (BMG LABTECH), setting the gain to the positive control (Pab1) and with the following filters: GFP – Ex485-12, Em520; Alexa594 – Ex584, Em620-10. Background signal from the buffer was subtracted and the data was normalized to the average of the mock control sample (triplicates). Three independent experiments (biological replicates) were performed for each tagged strain, and fluorescence for each experiment was measured in triplicates.

TAP-tagged proteins46 were immunopurified like GFP-tagged proteins with the following modifications: 100 µl of Pan mouse IgG Dynabeads (Life Technologies, 11041) were used to capture tagged proteins; and the proteins were released from beads in 100 µl elution buffer (10 mM Tris-HCl, pH 7.5, 150 mM LiCl, 1 mM EDTA, 1 mM DTT) containing 80 U ml^{-1} of AcTEV protease (Life Technologies, 12575-023) for 2 h at 19 °C. IP was controlled by immunoblot analysis (data not shown).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Puromycin treatment and affinity isolation of tagged proteins

Isolation of TAP-tagged proteins from non-crosslinked cells and puromycin treatment of extracts is described in the Supplementary Note 1.

RT-PCR

To assess apoptosis induction in L4 worms, total RNA was isolated from 50 µl of extracts with ZR RNA MiniPrep Kit (Zymo Research) with in column DNA digestion. RT was performed with 500 ng of RNA combined with a mixture of with oligo[dT]$_{18}$ and random hexamer primers, and the Transcriptor High Fidelity cDNA Synthesis Kit according to the manufactures' instructions (Roche, 05091284001). PCR was performed with 1 µl of complementary DNA (cDNA) for 5 min at 94 °C, 35 cycles at 94 °C for 30 s, 57 °C for 30 s, 72 °C for 40 s, and 8 min at 72 °C with primers detailed in **Supplementary Table 6**. Mpk-1 was used as a housekeeping control.

To identify mRNA targets of glycolytic enzymes, total RNA was isolated from 50 µl of extract (= input) and 100 µl of GFP or TAP IPs. 500 ng of input total RNA or 9.4 µl of IPed RNA (~ 28% of IP) were combined with a mixture of oligo[dT]$_{18}$ and random hexamer primers for RT with the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). PCR was conducted with 1 µl of cDNA reaction with gene specific primer pairs (**Supplementary Table 6**). PCR was performed for 5 min at 94 °C, 33 cycles at 94 °C for 30 s, 57 °C for 30 s, 72 °C for 40 s, and 8 min at 72 °C. To amplify, *ENO1* and *SET1*, 30 cycles were used.

Bioinformatics and statistical analysis

Adetailed description of statistical tests and bioinformatics (GO, Wikipathway, domain and motif analysis) is given in Supplementary Note 1. TheR code script for performing the hypergeometric test and Benjamini and Hochberg (BH) FDR correction using the p.adjust function is given in the Supplementary Note 2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to K. Heesom (Proteomics Facility, University of Bristol) for performing the MS analysis; M. Hengartner and D. Subasic (Institute of Molecular Life Sciences, University of Zurich) for support and *C. elegans* strains; S. Leidel (Max-Planck-Institute for Molecular Biomedicine, Münster) for GFP and TAP-tagged *S. cerevisiae* strains; M. Seedorf (Center for Molecular Biology, University of Heidelberg) and R. Ciosk (Friedrich Miescher Institute for Biomedical Research, Basel) for anti-Scp160 and anti-GLD-1 antibodies, respectively; D. Perez-Mendoza and D. Subasic for reading of the manuscript, and members of the Gerber lab and the Sinergia project for discussions. This study was funded by a 'Sinergia' grant (CSRII3-141942 (A.P.G.)) from the Swiss National Science Foundation and (in part) by the Biotechnology and Biological Sciences Research Council (BB/K009303/1 (A.P.G.)).

References

1. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 2008; 582:1977–86. [PubMed: 18342629]

Europe PMC Funders Author Manuscripts

2. Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. Trends Genet. 2008; 24:416–25. [PubMed: 18597886]
3. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet. 2014; 15:829–45. [PubMed: 25365966]
4. Scherrer T, Mittal N, Janga SC, Gerber AP. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PLoS One. 2010; 5:e15499. [PubMed: 21124907]
5. Tsvetanova NG, Klass DM, Salzman J, Brown PO. Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. PLoS One. 2010; 5
6. Castello A, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell. 2012; 149:1393–406. [PubMed: 22658674]
7. Baltz AG, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell. 2012; 46:674–90. [PubMed: 22681889]
8. Mitchell SF, Jain S, She M, Parker R. Global analysis of yeast mRNPs. Nat Struct Mol Biol. 2013; 20:127–33. [PubMed: 23222640]
9. Kwon SC, et al. The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol. 2013; 20:1122–30. [PubMed: 23912277]
10. Zhang L, Zhang K, Prandl R, Schoffl F. Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. Biochem Biophys Res Commun. 2004; 322:705–11. [PubMed: 15336521]
11. Frey S, Pool M, Seedorf M. Scp160p, an RNA-binding, polysome-associated protein, localizes to the endoplasmic reticulum of Saccharomyces cerevisiae in a microtubule-dependent manner. J Biol Chem. 2001; 276:15905–12. [PubMed: 11278502]
12. Mittal N, Roy N, Babu MM, Janga SC. Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. Proc Natl Acad Sci U S A. 2009; 106:20300–5. [PubMed: 19918083]
13. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. Plos Biology. 2008; 6:2297–2313.
14. Arava Y, et al. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2003; 100:3889–94. [PubMed: 12660367]
15. Costanzo MC, et al. Saccharomyces genome database provides new regulation data. Nucleic Acids Res. 2014; 42:D717–25. [PubMed: 24265222]
16. Consortium, U. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014; 42:D191–8. [PubMed: 24253303]
17. Thomas MP, Lieberman J. Live or let die: posttranscriptional gene regulation in cell stress and cell death. Immunol Rev. 2013; 253:237–52. [PubMed: 23550650]
18. Gartner A, Milstein S, Ahmed S, Hodgkin J, Hengartner MO. A conserved checkpoint pathway mediates DNA damage--induced apoptosis and cell cycle arrest in C. elegans. Mol Cell. 2000; 5:435–43. [PubMed: 10882129]
19. Conradt B, Horvitz HR. The C. elegans protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. Cell. 1998; 93:519–29. [PubMed: 9604928]
20. Francis R, Maine E, Schedl T. Analysis of the multiple roles of gld-1 in germline development: interactions with the sex determination cascade and the glp-1 signaling pathway. Genetics. 1995; 139:607–30. [PubMed: 7713420]
21. Tamburino AM, Ryder SP, Walhout AJ. A compendium of Caenorhabditis elegans RNA binding proteins predicts extensive regulation at multiple levels. G3 (Bethesda). 2013; 3:297–304. [PubMed: 23390605]
22. Harris TW, et al. WormBase 2014: new views of curated biology. Nucleic Acids Res. 2014; 42:D789–93. [PubMed: 24194605]
23. Lettre G, et al. Genome-wide RNAi identifies p53-dependent and -independent regulators of germ cell apoptosis in C. elegans. Cell Death Differ. 2004; 11:1198–203. [PubMed: 15272318]
24. Huang CY, et al. C. elegans EIF-3.K promotes programmed cell death through CED-3 caspase. PLoS One. 2012; 7:e36584. [PubMed: 22590572]

Europe PMC Funders Author Manuscripts

25. Voutev R, Killian DJ, Ahn JH, Hubbard EJ. Alterations in ribosome biogenesis cause specific defects in C. elegans hermaphrodite gonadogenesis. Dev Biol. 2006; 298:45–58. [PubMed: 16876152]
26. Clery A, Blatter M, Allain FH. RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol. 2008; 18:290–8. [PubMed: 18515081]
27. Ostlund G, et al. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 2010; 38:D196–203. [PubMed: 19892828]
28. Strein C, Alleaume AM, Rothbauer U, Hentze MW, Castello A. A versatile assay for RNA-binding proteins in living cells. RNA. 2014; 20:721–31. [PubMed: 24664470]
29. Halbach A, et al. Cotranslational assembly of the yeast SET1C histone methyltransferase complex. EMBO J. 2009; 28:2959–70. [PubMed: 19713935]
30. Lee AS, Kranzusch PJ, Cate JH. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. Nature. 2015
31. Henics T, et al. Mammalian Hsp70 and Hsp110 proteins bind to RNA motifs involved in mRNA stability. Journal of Biological Chemistry. 1999; 274:17318–17324. [PubMed: 10358092]
32. Zimmer C, von Gabain A, Henics T. Analysis of sequence-specific binding of RNA to Hsp70 and its various homologs indicates the involvement of N- and C-terminal interactions. RNA. 2001; 7:1628–37. [PubMed: 11720291]
33. Liu ZR, Wilkie AM, Clemens MJ, Smith CW. Detection of double-stranded RNA-protein interactions by methylene blue-mediated photo-crosslinking. RNA. 1996; 2:611–21. [PubMed: 8718690]
34. Suh N, Jedamzik B, Eckmann CR, Wickens M, Kimble J. The GLD-2 poly(A) polymerase activates gld-1 mRNA in the Caenorhabditis elegans germ line. Proc Natl Acad Sci U S A. 2006; 103:15108–12. [PubMed: 17012378]
35. Baumeister W, Walz J, Zuhl F, Seemuller E. The proteasome: paradigm of a self-compartmentalizing protease. Cell. 1998; 92:367–80. [PubMed: 9476896]
36. Schmid HP, et al. The prosome: an ubiquitous morphologically distinct RNP particle associated with repressed mRNPs and containing specific ScRNA and a characteristic set of proteins. EMBO J. 1984; 3:29–34. [PubMed: 6200323]
37. Kulichkova VA, et al. 26S proteasome exhibits endoribonuclease activity controlled by extra-cellular stimuli. Cell Cycle. 2010; 9:840–9. [PubMed: 20139718]
38. Mittenberg A, et al. MassSpectrometric Analysis of Proteasome Subunits Exhibiting Endoribonuclease Activity. Cell and Tissue Biology. 2014; 8:423–440.
39. Makino DL, Halbach F, Conti E. The RNA exosome and proteasome: common principles of degradation control. Nat Rev Mol Cell Biol. 2013; 14:654–60. [PubMed: 23989960]
40. Ciesla J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? Acta Biochim Pol. 2006; 53:11–32. [PubMed: 16410835]
41. Hentze MW, Preiss T. The REM phase of gene regulation. Trends Biochem Sci. 2010; 35:423–6. [PubMed: 20554447]
42. Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet. 2007; 8:533–43. [PubMed: 17572691]
43. Vizcaino JA, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol. 2014; 32:223–6. [PubMed: 24727771]
44. Scheckel C, Gaidatzis D, Wright JE, Ciosk R. Genome-wide analysis of GLD-1-mediated mRNA regulation suggests a role in mRNA storage. PLoS Genet. 2012; 8:e1002742. [PubMed: 22693456]
45. Huh WK, et al. Global analysis of protein localization in budding yeast. Nature. 2003; 425:686–91. [PubMed: 14562095]
46. Ghaemmaghami S, et al. Global analysis of protein expression in yeast. Nature. 2003; 425:737–41. [PubMed: 14562106]

Europe PMC Funders Author Manuscripts

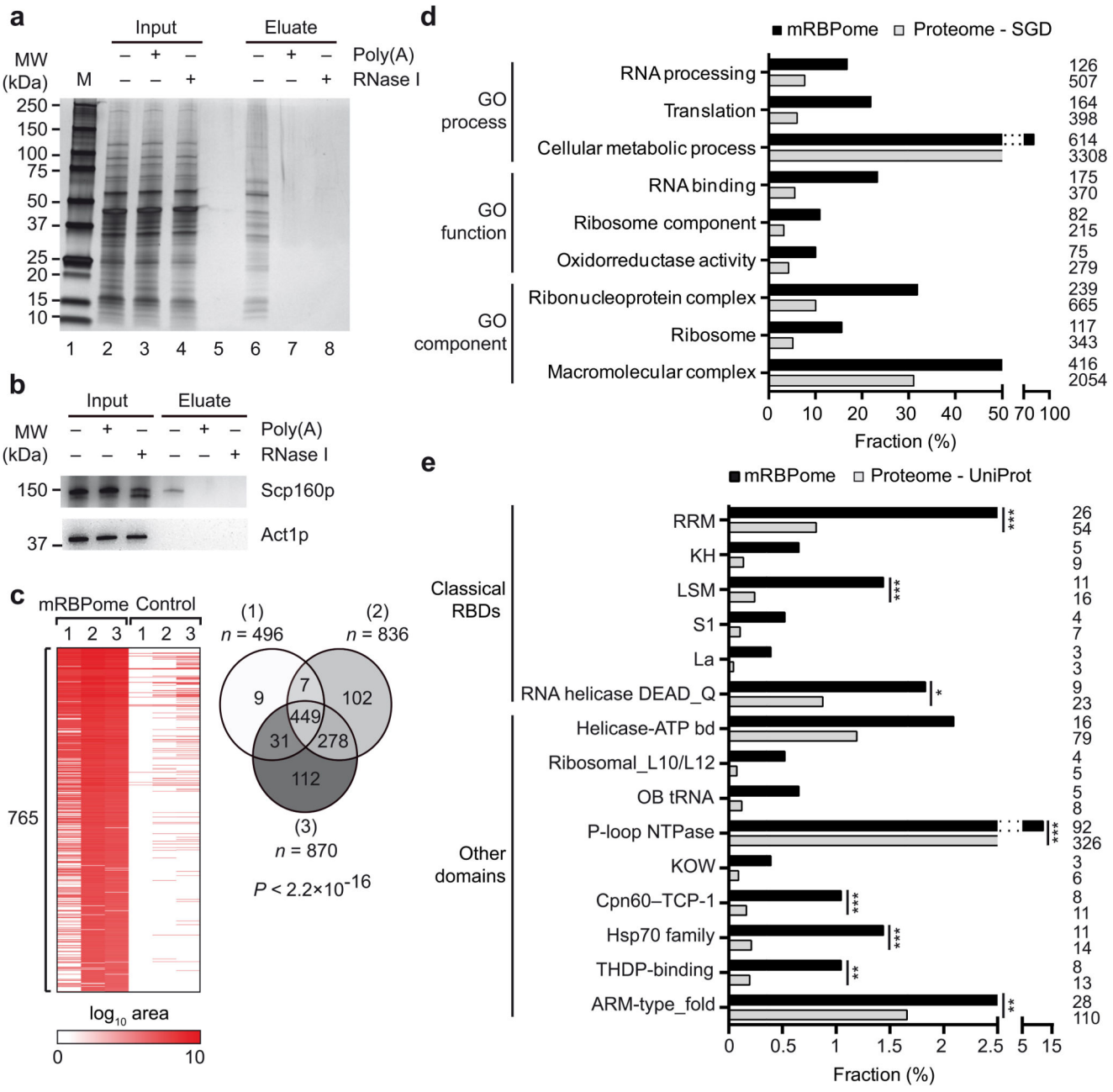

Figure 1.

Identification of mRBPs in *S. cerevisiae*. (**a**) Silver stained polyacrylamide gel. Lanes 2-4, input extracts and poly(A) and RNase treated control samples; lanes 6-8, eluates from poly(A) mRNA isolation. Lane 1 depicts the marker (M) with molecular weights (MW) indicated to the left. (**b**) Immunoblot analysis with Scp160p and Act1p antibodies. Original images are shown in the Supplementary Data Set 1. (**c**) Heatmap representation of the abundance of 765 proteins comprising the yeast mRBPome. Columns refer to three independent experiments and respective controls, rows represent individual proteins; for visualization purposes, the white-red color bar represents $\log_{10}$ transformed raw (non-

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

normalized) MS peak areas of respective proteins. The Venn diagram represents the overlap of identified proteins across the three experiments. The *P*-value (hypergeometric test) relates to the significance of overlap (see Methods). (**d**) Selective samples of significantly shared GO terms from SGD ($P < 0.01$, FDR < 5%) among proteins of the yeast mRBPome. Bars indicate the fraction of proteins annotated with respective GO term in the yeast mRBPome (765 proteins; black bars) and all GO annotated proteins in SGD (6,607 proteins; grey bars). (**e**) Selection of domains enriched in the mRBPome. Bars indicate fraction of annotated proteins bearing at least one of the indicated domains (InterPro) in the yeast mRBPome (765 proteins; black bars) and the reference proteome (6,621 proteins in UniProt; grey bars). The number of proteins within each fraction is shown to the right. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ at FDR < 5% (hypergeometric test).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

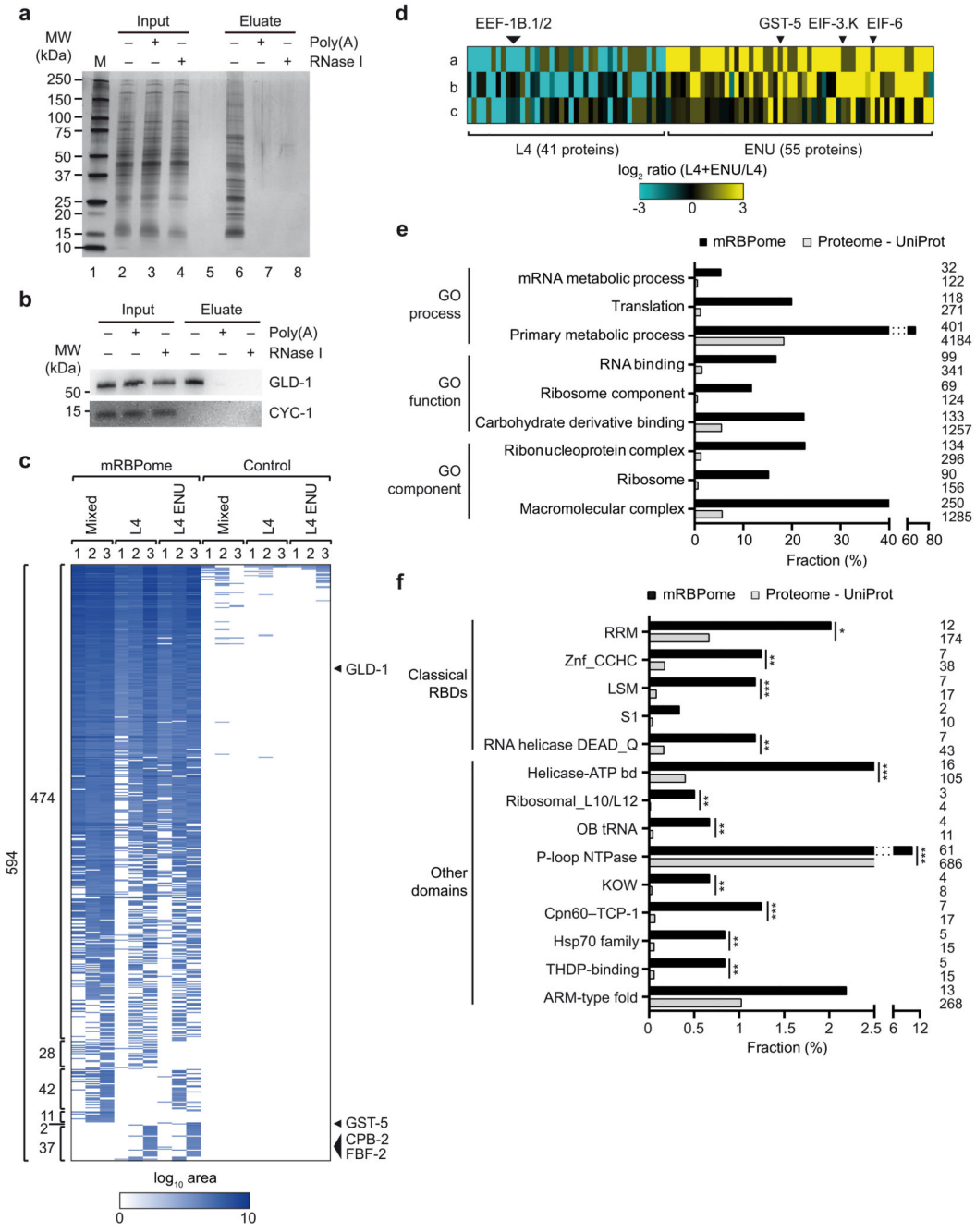

Figure 2.

Identification of mRBPs in *C. elegans*. (**a**) Silver stained polyacrylamide gel of UV-crosslinked samples from mixed-stages nematodes. (**b**) Immunoblot analysis with GLD-1 and CYC-1 antibodies. Original images of blots can be found in Supplementary Data Set 1. (**c**) Cluster heatmap representing the abundance of 594 proteins comprising the mRBPome within mixed-stage, L4-, and L4-staged animals treated with 5 mM ENU; and in the corresponding negative control samples. The white-blue scale shows raw (non-normalized) peak areas ($\log_{10}$ transformed) and number of proteins within groups is indicated to the left.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

(**d**) Relative changes of 96 proteins in the mRBPomes of synchronized L4-stage animals upon apoptosis (FDR < 5%). Rows indicate three pairwise comparisons within matched samples (a, b, c), columns refer to proteins. Fold changes are indicated with the blue-yellow color bar. (**e**) Significantly shared GO terms among proteins of the *C. elegans* mRBPome (FWER adj. $P < 0.01$). Bar diagrams indicate the fraction of proteins in the respective GO term among the *C. elegans* mRBPome (594 proteins; black bars) and the 22,817 GO annotated proteins in the UniProt reference proteome (grey bars). (**f**) Domains enriched in the mRBPome. Bars relate to the fraction of proteins bearing at least one of the domains (InterPro) in the *C. elegans* mRBPome (594 proteins; black) and 26,165 proteins comprising the UniProt reference proteome (grey). Numbers of proteins is shown to the right. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ at FDR < 5% (hypergeometric test).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

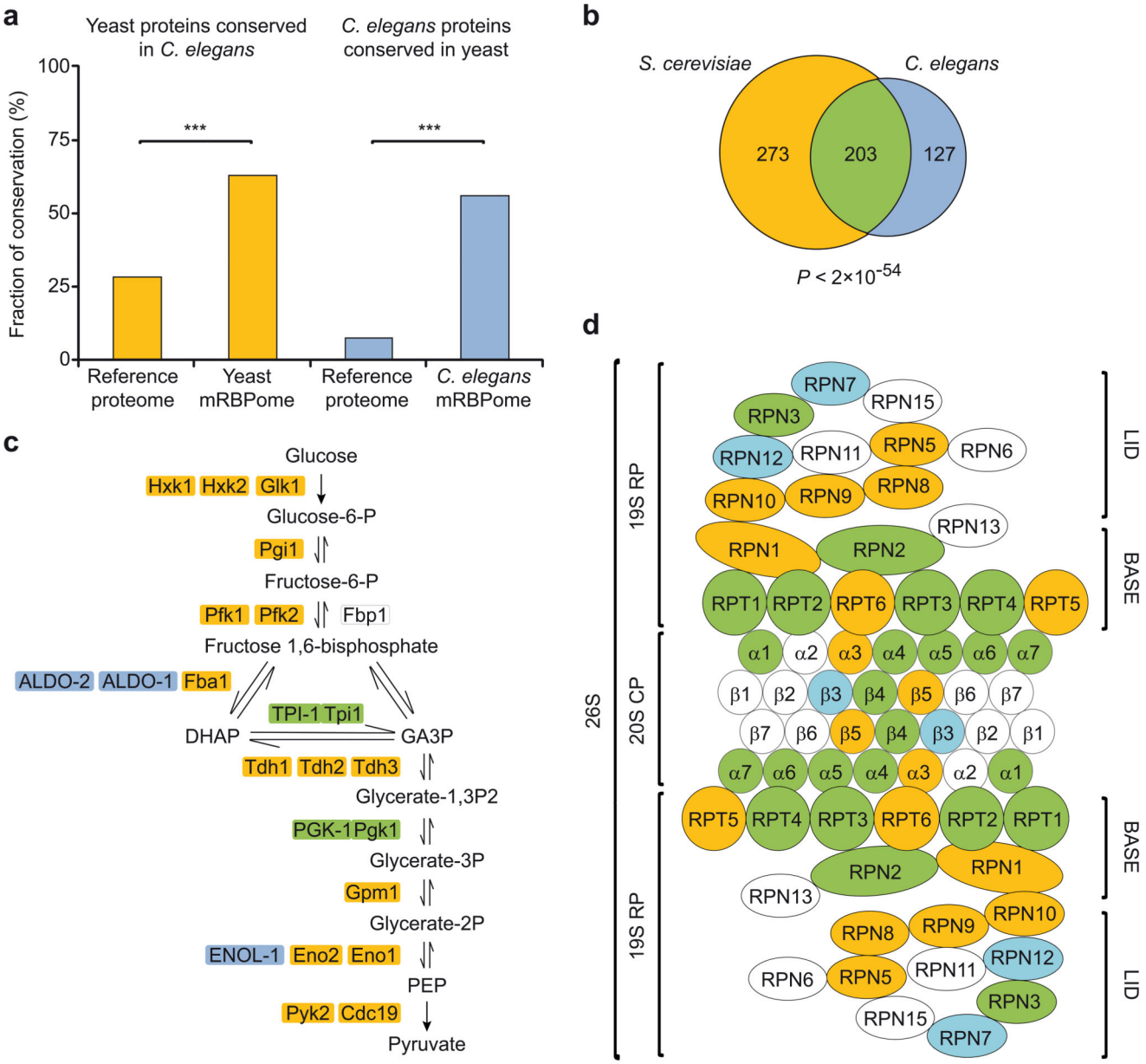

Figure 3.

Conservation of the mRBPome across species. (**a**) Conservation between *S. cerevisiae* and *C. elegans* proteins27. Yellow columns refer to the fraction of conserved *S. cerevisiae* proteins in *C. elegans* considering the reference proteome (6,621 proteins) and mRBPome (765 proteins). Blue columns refer to the fraction of conserved *C. elegans* proteins in *S. cerevisiae* considering the worm reference proteome (26,165 proteins) and mRBPome (594 proteins). Stars demarcate the significant difference between fractions ($P < 0.001$, hypergeometric test). (**b**) Venn diagram showing overlap of orthologous proteins of the mRBPome in *S. cerevisiae* (476 proteins) and *C. elegans* (330 proteins). The *P*-value relates to the significance of overlap (hypergeometric test). (**c**) Schematic view of the glycolytic pathway. Proteins highlighted in yellow are exclusively present in the yeast mRBPome,

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

proteins in blue are exclusively found in *C. elegans* mRBPome, and proteins shared in both mRBPomes are in green. (**d**) Schematic view of the proteasome comprising the core particle (20S CP) that can bind to one or two regulatory particles (19S RP).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

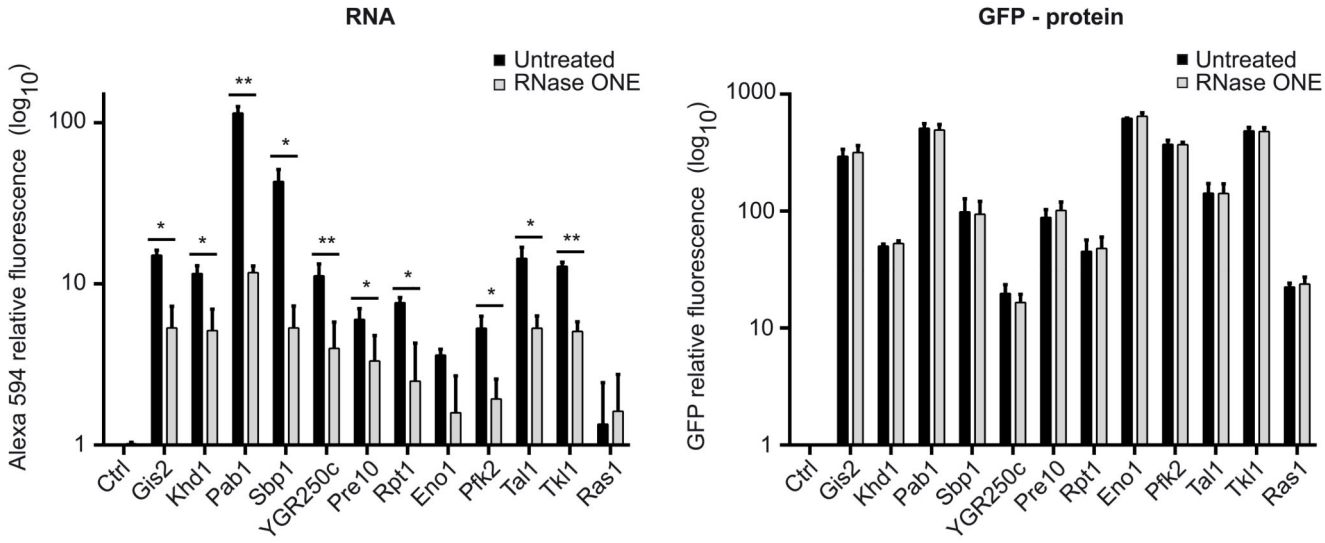

Figure 4.

Validation of mRNA-protein interactions by quantitative dual fluorescence-based mRNA detection assay. Normalized Alexa 594 fluorescence signal was used to monitor the presence of poly(A) mRNAs (left panel, black bars). As a control, the samples were treated with RNase ONE (grey bars). Data was normalized to untagged wild type cells (Ctrl). Error bars, s.e.m. ($n = 3$ independent immunopurifications). $*P \leq 0.05$; $**P \leq 0.01$ by paired, two-tailed, Student's t test. The right panel shows the normalized GFP fluorescence signal of immunopurified proteins. Data underlying the graphical representation can be found in Supplementary Data Set 2.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

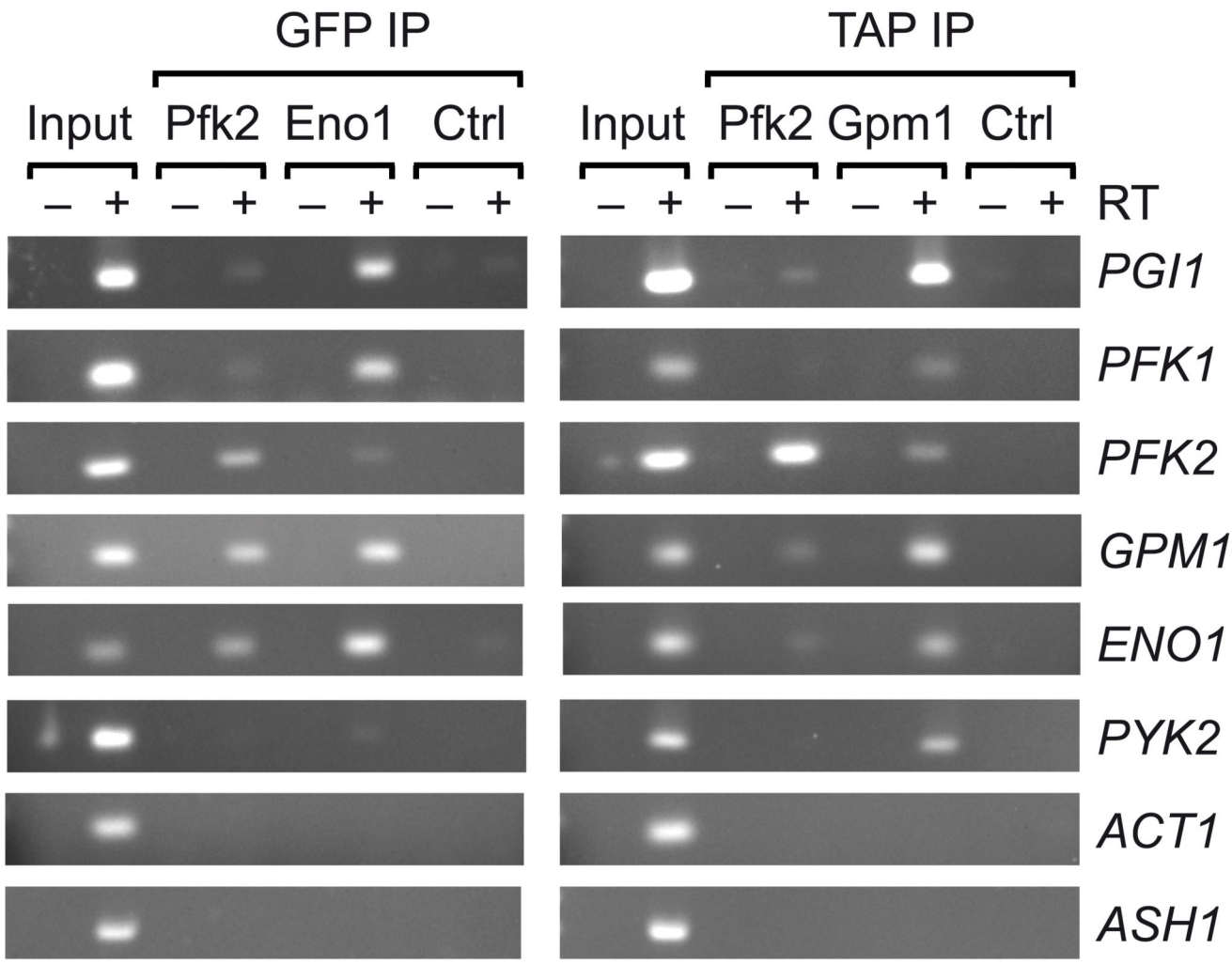

Figure 5.

Glycolytic enzymes selectively interact with glycolytic mRNAs. Agarose gel showing products from reverse transcription (RT)-PCR reactions with gene specific primers to detect glycolytic mRNAs (right) bound to immunopurified GFP or TAP-tagged proteins of indicated glycolytic proteins (top). Input refers to total RNA from crosslinked cells; Ctrl, untagged control cells. Full-sized images of gels can be found in Supplementary Data Set 1.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts