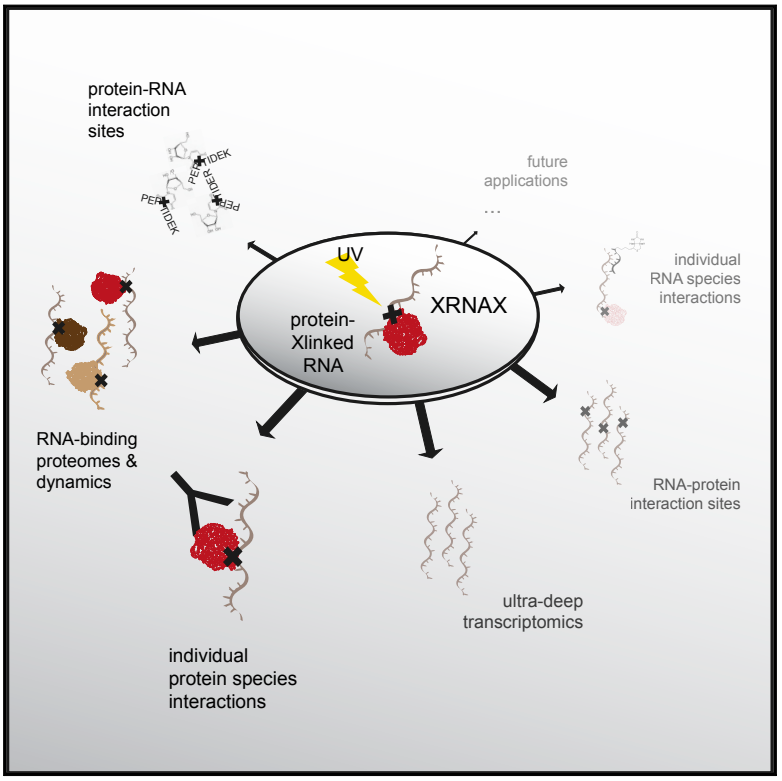
Resource

Cell

The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest

Graphical Abstract



Authors

Jakob Trendel, Thomas Schwarzl, Rastislav Horos, Ananth Prakash, Alex Bateman, Matthias W. Hentze, Jeroen Krijgsveld

Correspondence

j.krijgsveld@dkfz.de

In Brief

A general approach for characterizing cellular RNA-protein interactions allows examination of dynamic changes to the RNA-bound proteome.

Highlights

- XRNAX purifies protein-crosslinked RNA of all biotypes from UV-crosslinked cells
- Discovery of the WKF RNA-binding domain
- Discovery of more than 700 proteins interacting with non-polyadenylated RNA
- Profiling of stress-induced changes in RNA-binding proteomes

Trendel et al., 2019, Cell *176* , 391–403
January 10, 2019
https://doi.org/10.1016/j.cell.2018.11.004

CellPress

Resource

Cell

The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest

Jakob Trendel,[1,2,3] Thomas Schwarzl,[2] Rastislav Horos,[2] Ananth Prakash,[4] Alex Bateman,[4] Matthias W. Hentze,[2] and Jeroen Krijgsveld[1,5,6,*]

[1]German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, Heidelberg, Germany
[2]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg, Germany
[3]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences
[4]European Molecular Biology Laboratory, European Bioinformatics Institute (EBI), Wellcome Genome Campus, Hinxton, Cambridge, UK
[5]Heidelberg University, Medical Faculty, Im Neuenheimer Feld 672, Heidelberg, Germany
[6]Lead Contact
*Correspondence: j.krijgsveld@dkfz.de
https://doi.org/10.1016/j.cell.2018.11.004

SUMMARY

Proteins and RNA functionally and physically intersect in multiple biological processes, however, currently no universal method is available to purify protein-RNA complexes. Here, we introduce XRNAX, a method for the generic purification of protein-crosslinked RNA, and demonstrate its versatility to study the composition and dynamics of protein-RNA interactions by various transcriptomic and proteomic approaches. We show that XRNAX captures all RNA biotypes and use this to characterize the sub-proteomes that interact with coding and non-coding RNAs (ncRNAs) and to identify hundreds of protein-RNA interfaces. Exploiting the quantitative nature of XRNAX, we observe drastic remodeling of the RNA-bound proteome during arsenite-induced stress, distinct from autophagy-related changes in the total proteome. In addition, we combine XRNAX with crosslinking immunoprecipitation sequencing (CLIP-seq) to validate the interaction of ncRNA with lamin B1 and EXOSC2. Thus, XRNAX is a resourceful approach to study structural and compositional aspects of protein-RNA interactions to address fundamental questions in RNA-biology.

INTRODUCTION

Cellular proteins and RNA intimately interact in intricate ways to regulate a wide range of processes that are essential for cells to survive, replicate, and adapt to environmental changes. For instance, proteins accompany mRNA throughout its lifetime from transcription to decay, deploying specialized complexes for RNA splicing, capping, translation, and localization (for review, see Müller-McNicoll and Neugebauer, 2013). Indeed, in molecular machines like the ribosome proteins and various types of RNAs (i.e. rRNA, tRNA, and mRNA) converge for the genesis of new proteins. Despite the central role of such processes in the flow of genetic information, current methods to investigate protein-RNA interactions are limited to characterizing interactions of one particular protein species (as in crosslinking immunoprecipitation sequencing [CLIP-seq]) (for review, see Lee and Ule, 2018), of one particular RNA species (i.e., RAP [McHugh et al., 2015], CHIRP [Chu et al., 2015], or iDRIP [Minajigi et al., 2015]), or the bulk of all polyadenylated RNA (as in interactome capture) (for review, see Hentze et al., 2018). Because interactome capture relies on RNA polyadenylation, it ignores >95% of mammalian cellular RNA (per weight) that is non-polyadenylated (Hastie and Bishop, 1976). This includes highly abundant "housekeeping" RNAs such as tRNA, rRNA, small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs) and also a diverse group of other non-coding RNAs (ncRNAs) whose functionalities are only beginning to emerge (for review, see Rinn and Chang, 2012).

Inspired by the perspective that an unbiased, comprehensive characterization of protein-RNA interactions will greatly aid in understanding the functionality of both the RNA and protein involved, we developed a method for the extraction and purification of protein-crosslinked RNA irrespective of RNA biotype. We named this method XRNAX, for "protein-crosslinked RNA extraction." Crucially, XRNAX purifies protein-crosslinked RNA as a physical entity that can serve as a universal starting point for various downstream applications, both using sequencing and mass spectrometry (MS). Here, we combined XRNAX with proteomics to derive an integrated draft of the human RNA-binding proteome and to locate protein-RNA interfaces through ribonucleotide-crosslinked peptides. Subsequently, we quantified the dynamics in protein-RNA interactions following arsenite treatment and report massive remodeling of the translation-associated and RNA-interacting proteome in human cells. Finally, we used XRNAX prior to CLIP-seq, validating LMNB1 as a novel RNA-binder and following RNA-binding kinetics of the RNA exosome component EXOSC2. Thus, we demonstrate that XRNAX is a powerful, new tool for RNA biology, which we envision to foster our understanding of post-transcriptional regulatory processes in many cell types and organisms.

RESULTS

Organic Phase Separation for the Extraction of Protein-Crosslinked RNA

Because acid guanidinium thiocyanate-phenol-chloroform extraction (in the following called TRIZOL) (Chomczynski and Sacchi,

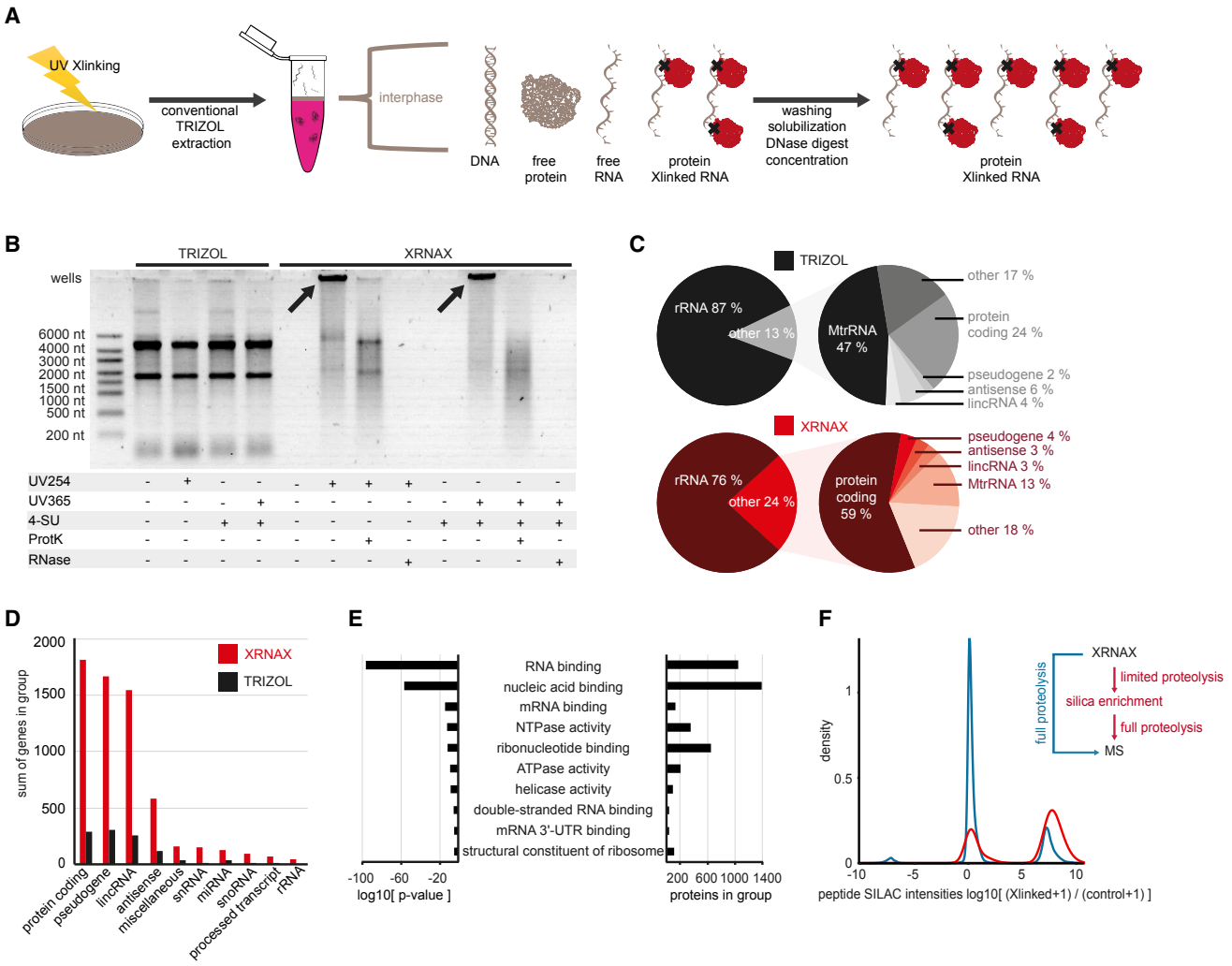


Figure 1. XRNAX Extracts Protein-Crosslinked RNA from UV-Crosslinked Cells

(A) Experimental scheme of XRNAX.

(B) Comparison of classic TRIZOL and XRNAX by agarose gel electrophoresis. MCF7 cells were subjected to crosslinking with UV at 254 nm (UV254) or incubated with 100 μM 4-thiouridine (4-SU) for 16 hr and crosslinked at 365 nm (UV365). RNA was extracted using either conventional TRIZOL or XRNAX and subsequently treated as indicated. ProtK, digestion with proteinase K; RNase, digestion with RNase A and I.

(C) Pie diagrams comparing RNA composition of TRIZOL extracts (top) in comparison to XRNAX extracts (bottom). RNA from XRNAX extracts or TRIZOL extracts was sequenced before (left) and after (right) depletion of ribosomal RNA. Reads for each GENCODE biotype were normalized to the total number of reads in one library.

(D) Transcripts exclusively identified after sequencing of either TRIZOL or XRNAX-extracted RNA from MCF7 cells that were grown in the presence of 4-SU. Sum of genes are summarized within the top-10 GENCODE biotypes.

(E) GO-enrichment analysis (molecular function) for proteins in XRNAX extracts from MCF7 cells. Displayed are the ten most-enriched terms.

(F) Density plot showing SILAC ratios of peptides from MCF7 XRNAX-extracts before (blue) and after silica purification (red). Pseudo-counts were added to display peptides exclusively identified in either SILAC channel.

See also Figure S1.

2006) is a classical method to separate RNA (in the aqueous phase) from protein (in the organic phase), we hypothesized that protein-crosslinked RNA might end up in the insoluble TRIZOL interphase (Figure 1A). We therefore UV-crosslinked MCF7 cells and collected the TRIZOL interphase, washed it to remove free protein and RNA, and DNase-digested to eliminate DNA, yielding an extract containing highly concentrated RNA (>1,000 ng/μL) and protein (>0.7 mg/mL) (see STAR Methods for details).

Interestingly, when run on an agarose gel, this extract showed a band with a distinct upward shift compared to conventional TRIZOL-extracted RNA (Figure 1B, arrows), which disappeared upon RNase as well as proteinase K digestion, indicating that it consisted of protein-crosslinked RNA. Identical results were achieved when applying UV-crosslinking at 365 nm using cells grown in the presence of the photoactivatable nucleotide 4-thiouridine (4-SU, Figure 1B).

RNase-digested XRNAX extracts showed DNA contamination <0.1% (weight DNA/weight RNA) using fluorescence-based, DNA-specific Qubit quantification, reflecting earlier findings that UV-crosslinking of DNA to protein requires energy doses 10–100 times higher than used here for protein-RNA crosslinking (Steube et al., 2017).

In conclusion, this indicated that the procedure, consisting of UV-crosslinking followed by TRIZOL extraction and subsequent processing steps of the interphase, purified protein-crosslinked RNA. We termed this method protein-crosslinked RNA eXtraction, or XRNAX (Figure 1A).

XRNAX-Extracts Contain RNA of All Major Biotypes

To obtain a view on the type and relative distribution of RNA biotypes, we digested an XRNAX extract obtained from MCF7 cells with proteinase K and sequenced its RNA. For comparison, we sequenced total RNA extracted from non-crosslinked MCF7 cells using the conventional TRIZOL protocol. In addition, RNA was sequenced before and after depletion of rRNA to display the entire relative composition of each preparation. This showed that XRNAX extracts contained all major RNA biotypes also detected in classical TRIZOL extracted RNA, albeit in different relative amounts (Figure 1C). Very similar results were obtained when using 4-SU labeling and crosslinking at 365 nm. Interestingly, we found that, at virtually identical sequencing depth, libraries prepared via XRNAX had much better coverage for medium and low abundant transcripts (Figures S1B and S1C). Effectively, this led to the detection of 6,306 transcripts exclusively observed in XRNAX-derived libraries, including ~1,500 protein-coding and ~1,500 large intergenic noncoding RNAs (lincRNAs) (Figure 1D).

XRNAX Extracts Are Enriched in RNA-Binding Proteins Crosslinked to RNA

Next, we used MS to profile the protein composition of XRNAX extracts and to validate that these contain RNA-crosslinked protein. We therefore compared the protein composition of an XRNAX extract to a total lysate of MCF7 cells. Gene ontology (GO) enrichment analysis (Figure 1E) and ranking protein abundances using intensity-based absolute quantification (iBAQ) (Schwanhäusser et al., 2011) (Figure S1A) confirmed that XRNAX extracts were strongly enriched in RNA-binding proteins.

To control for unspecific or indirect protein-protein rather than direct protein-RNA interactions, we applied stable isotope labeling in cell culture (SILAC) in three validation experiments. First, heavy SILAC MCF7 cells were UV-crosslinked and combined with non-crosslinked light SILAC cells, subjected to XRNAX and the protein content of the extract analyzed by MS. This revealed many peptides that only showed intensity in the heavy SILAC channel, indicating enrichment of RNA-crosslinked peptides (Figure 1F). However, we noted that many non-enriched peptides originated from bona fide RNA-binding proteins, suggesting that they were trapped in the TRIZOL interphase even without crosslinking.

To solve this, we designed a denaturing silica-based cleanup procedure downstream of XRNAX. Because silica columns retain RNA, but not protein-crosslinked RNA under standard conditions, we subjected XRNAX extracts to limited tryptic digestion resulting in efficient retention of RNA crosslinked to protein fragments. Subsequently, the majority of all peptides had an intensity ratio >1,000 of crosslinked over non-crosslinked, in the following referred to as "super-enriched peptides" (Figure 1F). By integration of SILAC intensities, we estimated that before silica enrichment, XRNAX extracts contained on average 69% (65% median) RNA-crosslinked protein, whereas after silica enrichment this was 89% (100% median) (Figures S1D and S1E).

In a second validation, we verified that silica enrichment depended on the presence of RNA. Therefore, we degraded RNA in an XRNAX extract from heavy SILAC cells with NaOH/Mg^{2+} and combined it with an untreated XRNAX extract from light SILAC cells. As expected, MS analysis identified many thousands of super-enriched peptides exclusively originating from the sample with intact RNA (Figure S1F).

In a third validation we established that super-enriched peptides could be used to quantify changes in protein-RNA interactions. To this end, we mixed decreasing amounts of UV-crosslinked MCF7 cells with non-crosslinked cells both grown with heavy SILAC label and combined this with a fixed amount of UV-crosslinked cells with light SILAC label. Thereby, these samples contained discrete ratios of RNA-crosslinked over non-crosslinked proteins ranging from 1:4 to 1:256 (Figure S1G). Indeed, after processing samples via XRNAX and silica enrichment, most peptides were accurately quantified according to their dilution (Figure S1G). Interestingly, the large majority of these peptides had previously been classified as super-enriched (Figure 1F), indicating that filtering for super-enrichment (see STAR Methods) efficiently preselects peptides to precisely quantify their changes in RNA binding.

Direct Detection of Nucleotide-Crosslinked Peptides from XRNAX Extracts

Having established XRNAX, we applied it to identify protein-RNA interaction sites through RNA-crosslinked peptides. Therefore, we digested an MCF7 XRNAX extract with excess trypsin, purified RNA-bound peptides via silica columns, digested RNA, and analyzed RNA-peptide hybrids by MS (Figure 2A, left). Importantly, for the unbiased identification of any remaining nucleotide adducts, we used the mass-tolerant search engine MSFragger (Kong et al., 2017). This identified 324 and 306 Dalton as the most dominant mass shifts (Figure 2B), corresponding to peptide modifications by uridine monophosphate (U) and cyclic uridine monophosphate (cyclic U), respectively. These mononucleotide-peptide adducts of U were in accordance with an earlier study using interactome capture (Kramer et al., 2014), however, in addition, we identified oligonucleotide-adducts with permutations of di- and trinucleotide sequences carrying at least one U (Figure 2B). Based on this, we suggest that U acts as the main crosslinking base, although we cannot exclude that crosslinking with other ribonucleotides escaped tandem mass spectrometry (MS/MS) detection. Analysis of amino acid frequencies in cyclic-U-crosslinked peptides showed a clear enrichment for phenylalanine (F), lysine (K), and glycine (G) compared to peptides from MCF7 total lysates (Figure 2E). Moreover, we could locate F-containing peptides in the cryo-electron microscopy (cryo-EM) structure of the human 80S ribosome (Khatter

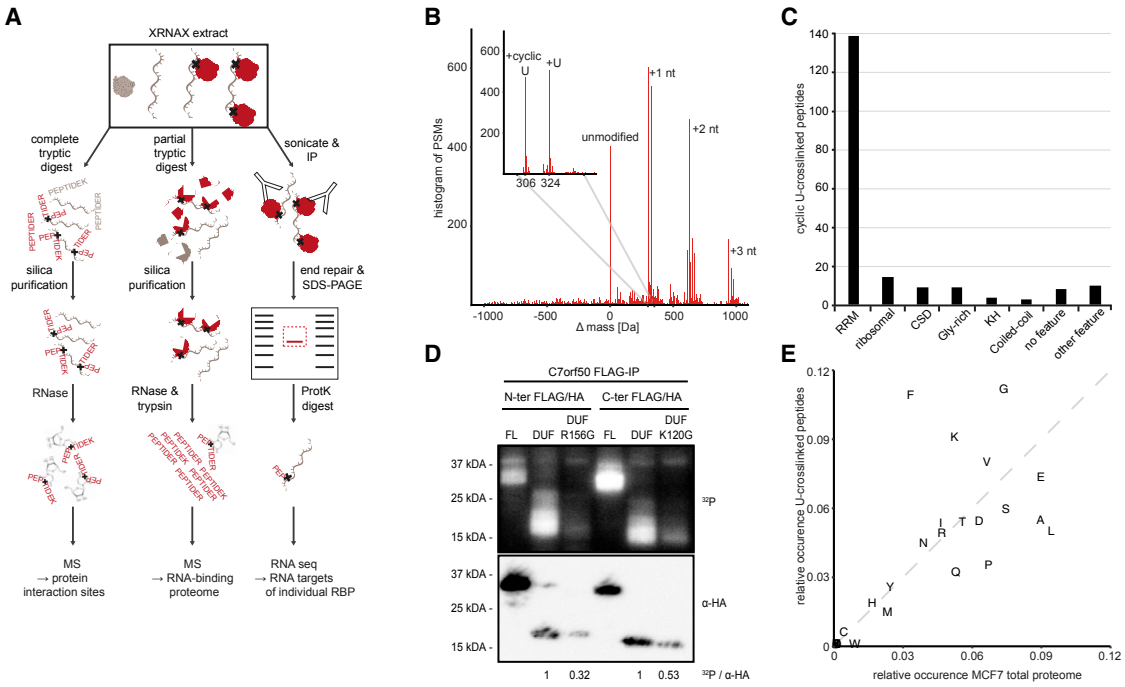


Figure 2. Identification of RNA Interaction Sites and Domains

(A) Proteomic and transcriptomic applications downstream of XRNAX.

(B) Identification of adduct masses on XRNAX-purified peptides by MSFragger.

(C) Uniprot feature-annotation of cyclic U-crosslinked peptides. Bar graph displays number of cyclic-U crosslinked peptides that were found in stretches of their cognate protein with indicated feature. RRM, RNA recognition motif; KH, K homology domain; CSD, cold shock domain; Gly-rich, glycine-rich amino acid sequence; no feature, no feature deposited in Uniprot; other feature, feature other than the ones mentioned in the other categories.

(D) PNK assay for C7orf50 and its DUF2373. MCF7 cells were transiently transfected with the full-length protein (FL), DUF2373 of C7orf50 (DUF), or two mutant DUFs (R156G, K120G). Top: representative phosphoimage showing radioactive labeling of covalently bound RNA in RNP complexes. Bottom: anti-HA western blot of same membrane showing efficiency of anti-FLAG IP. Values at the bottom indicate relative RNA-binding of DUF2373 to the respective DUF2373 mutant using relative intensities from the images above.

(E) Comparison of amino acid frequencies in peptides crosslinked to cyclic U and peptides found in the MCF7 total proteome.

See also Figure S2 and Table S1.

et al., 2015), where frequently F was in the close vicinity of a uracil-base (Figure S2E).

Collectively, this led to the detection of 197 cyclic U-crosslinked peptides from 93 proteins (Table S1), including the ones that had been observed before (Kramer et al., 2014). Importantly, these findings provided direct evidence that XRNAX extracts contained RNA-crosslinked proteins.

Nucleotide-Crosslinked Peptides Locate Known and Unknown Protein-RNA Interfaces and Identify DUF2373 as Novel RNA-Binding Domain

Of the 93 proteins that were directly identified by nucleotide-crosslinked peptides, 90% were annotated as RNA-binding (Table S1). More than 85% of nucleotide-crosslinked peptides mapped to bona fide RNA-binding domains, confirming that crosslinks denote protein-RNA interfaces. Most of the peptides localized in RNA-recognition motif (RRM), K-homology (KH), or cold-shock domains (CSD) (Figure 2C; Table S1). In addition, we frequently identified glycine-rich regions, in accordance with earlier RBDmap and RBR-ID studies showing that low-complexity regions are abundantly involved in RNA-binding (Castello et al., 2016; He et al., 2016). Proteins identified with the most crosslinked peptides were HNRNPA2B1 (13), NCL (10), and HNRNPAB (8), all of which located to RRMs or glycine-rich regions (Figure S2A).

Interestingly, we identified a nucleotide-crosslinked peptide in the domain of unknown function DUF2373 of the uncharacterized protein C7orf50 (Table S1). We previously had identified C7orf50 as RNA-binding protein in HeLa cells (Castello et al., 2012), however, it did not carry any known RNA-binding domains nor did it appear in RBDmap or RBR-ID studies (Castello et al., 2016; He et al., 2016). Analysis in STRING implied C7orf50 in ribosome biogenesis and pseudouridylation (Figure S2C), supported by nucleolar localization reported in the Human Protein Atlas (Bailey et al., 2009) (Figure S2D). We purified full-length human C70orf50 and its DUF2373 from *E. coli*. Native agarose gel electrophoretic mobility shift assays (EMSA) showed that both were able to bind MCF7 total RNA *in vitro* (Figure S2F). Next, we transiently expressed FLAG-HA-tagged versions of C7orf50 and its DUF2373 in MCF7 cells and were able to confirm their ability to bind RNA *in vivo* in a polynucleotide kinase (PNK) assay (Figure 2D). Mutagenesis at computationally predicted

polynucleotide interaction sites in DUF2373-reduced RNA binding. Furthermore, we performed XRNAX followed by complete RNase digestion, anti-FLAG immunoprecipitation, and MS. Here, MSfragger reproduced detection of the nucleotide-cross-linked peptide in immunoprecipitates from wild-type C7orf50 and DUF2373 but not from DUF2373 mutants (data not shown).

We propose renaming DUF2373 as the WKF domain after the three most conserved amino acids in the sequence alignment (Figure S2B). This naming scheme has been adopted by the Pfam database, where the WKF domain is now annotated as RNA-binding domain. Interestingly, in the human genome, the WKF domain only occurs once—in C7orf50. However, the WKF domain is found widely in animal and fungal species suggesting it has an important yet unknown function. In summary, these data significantly expanded the number of proteins with direct MS evidence for RNA interaction. Furthermore, we demonstrated that these interaction sites can give valuable clues to the function of unknown sequence features within RNA-binding proteins.

An Integrated Draft of the Human RNA-Binding Proteome from Three Cell Lines

To explore the complete RNA-bound proteome, we applied our enrichment scheme (XRNAX and partial digestion followed by silica enrichment) (Figure 2A, center) to three commonly used human cells lines, MCF7, HeLa, and HEK293, which had been subject of poly(A)-interactome capture studies before (Baltz et al., 2012; Castello et al., 2012; Milek et al., 2017). To classify proteins as RNA-binding, we only considered proteins that were detected with at least two super-enriched peptides to reflect that they were exclusively found in crosslinked cells (Figure 1F) in a reproducible manner (Figure 3A). This resulted in high-confidence RNA-binding proteomes of 1,207 proteins for MCF7, 1,239 proteins for HeLa, and 1,357 proteins for HEK293 cells, of which 858 were shared by all three cell lines (Figure 3B; Table S2).

The XRNAX-derived HeLa RNA-binding proteome rediscovered 647 (i.e. 80%) of the proteins found in our previous poly(A)-interactome capture (Castello et al., 2012) and added almost 600 more (Figure 3C). Although we applied a more stringent enrichment cut-off for calling RNA-binding proteins (presence-absence instead of false discovery-controlled statistics), we were able to confirm some of the proteins that were earlier reported as "candidate mRNA-binder" (Figure 3C) (Castello et al., 2012). Proteins only discovered through interactome capture did not show significant differences in physicochemical features, abundance in the total proteome, GO annotation, or occurrence of specific domains when compared to proteins discovered by both XRNAX and interactome capture. However, they were reported with much lower confidence (Wilcox rank-sum test $p < 10E-16$, Figure S3A). These observations showed that the discovery of RNA-binding proteins through XRNAX added both sensitivity and specificity compared to interactome capture.

Similarly, XRNAX-derived RNA-binding proteomes in MCF7 and HEK293 cells showed large (>75%) overlap with their published poly(A) interactomes (Figure 3D, top). Furthermore, in each cell line, we identified more than 600 additional proteins, which here we refer to as the non-poly(A) interactome. Notably, we cannot exclude that any previously discovered poly(A)-binding proteins do not also bind non-poly(A) RNA. However, we can make this distinction for non-poly(A) interactors—a novel and distinct group of RNA interactors, which the following analysis will focus on.

We were interested to see how much of the non-poly(A) interactomes could be explained by current GO annotations. More than 70% of proteins in the non-poly(A) interactomes could be assigned to GO terms in five RNA-related categories (Figure 3D, bottom). Thus, taking the XRNAX-derived poly(A) and non-poly(A) interactomes together, more than 80% of proteins had a prior annotation related to the interaction with RNA or ribonucleotides.

Combining our findings from all three cell lines resulted in a collection of 1,753 proteins that we named the integrated human RNA-binding proteome or ihRBP, containing 978 proteins (70%) of previous poly(A) interactomes and 775 proteins constituting the novel non-poly(A) interactome (Figure 3E).

GO analysis of the ihRBP revealed strong enrichment for RNA-related terms in the poly(A) interactome (Figure S3B), using deep total proteome data as a background (Geiger et al., 2012). For the non-poly(A) interactome, we removed proteins of the two initial poly(A) interactome studies by Baltz et al. (2012) and Castello et al. (2012) from our background dataset and subsequently found strong enrichment for GO terms relating to non-coding RNA such as rRNA, 7S RNA, tRNA, snRNA, or snoRNA and the ribosome (Figure S3B).

Indeed, the non-poly(A) interactome of the ihRBP contained numerous proteins known to interact with non-coding RNAs, such as ribosomal proteins, numerous ribosome biogenesis factors (e.g., LTV1, MDN1, RIOK1-3, and RIOX2), 17 out of 23 aminoacyl tRNA synthetases, translation initiation factors (e.g., ABCE1, EIF2S3, EIF3B, and EIF3J), RNA exosome components (e.g., SKIV2L, EXOSC2, EXOSC3, and DIS3), splicing factors (e.g., SF3B3, LSM8, ESRP1, or CWC22), proteins involved in micro RNA biogenesis (e.g., DICER1, TSN, and TARBP2), many transcription-associated proteins (e.g., known to interact with 7SK RNA) (e.g., HEXIM1, CCNT1, and CDK9), and interactors of telomerase RNA (NOP10 and PINX1).

Thus, XRNAX in combination with silica enrichment allowed us to identify highly reproducible, high-confidence RNA-binding proteomes. These included the large majority of the known poly(A)-interactomes and added more than 600 proteins, many of which are known interactors of ncRNA.

Sequence-Encoded Information Distinguishes Poly(A) from Non-poly(A) RNA Interactors

We next asked if any sequence-encoded properties could be identified that distinguished poly(A) from non-poly(A)-binding proteins in the ihRBP. While amino acid frequencies were virtually identical between the two groups (Figure S3C), distinct differences became apparent when comparing di- and tripeptide frequencies within proteins (Wilcox rank-sum test $p = 0.02$). Interestingly, proteins of the poly(A) interactome showed the strongest enrichment for tripeptides carrying combinations of amino acids known to contribute to intrinsically disordered regions (IDRs) (Figure S3C) (Brangwynne et al., 2015). A focused analysis toward these amino acids specifically highlighted hexapeptides

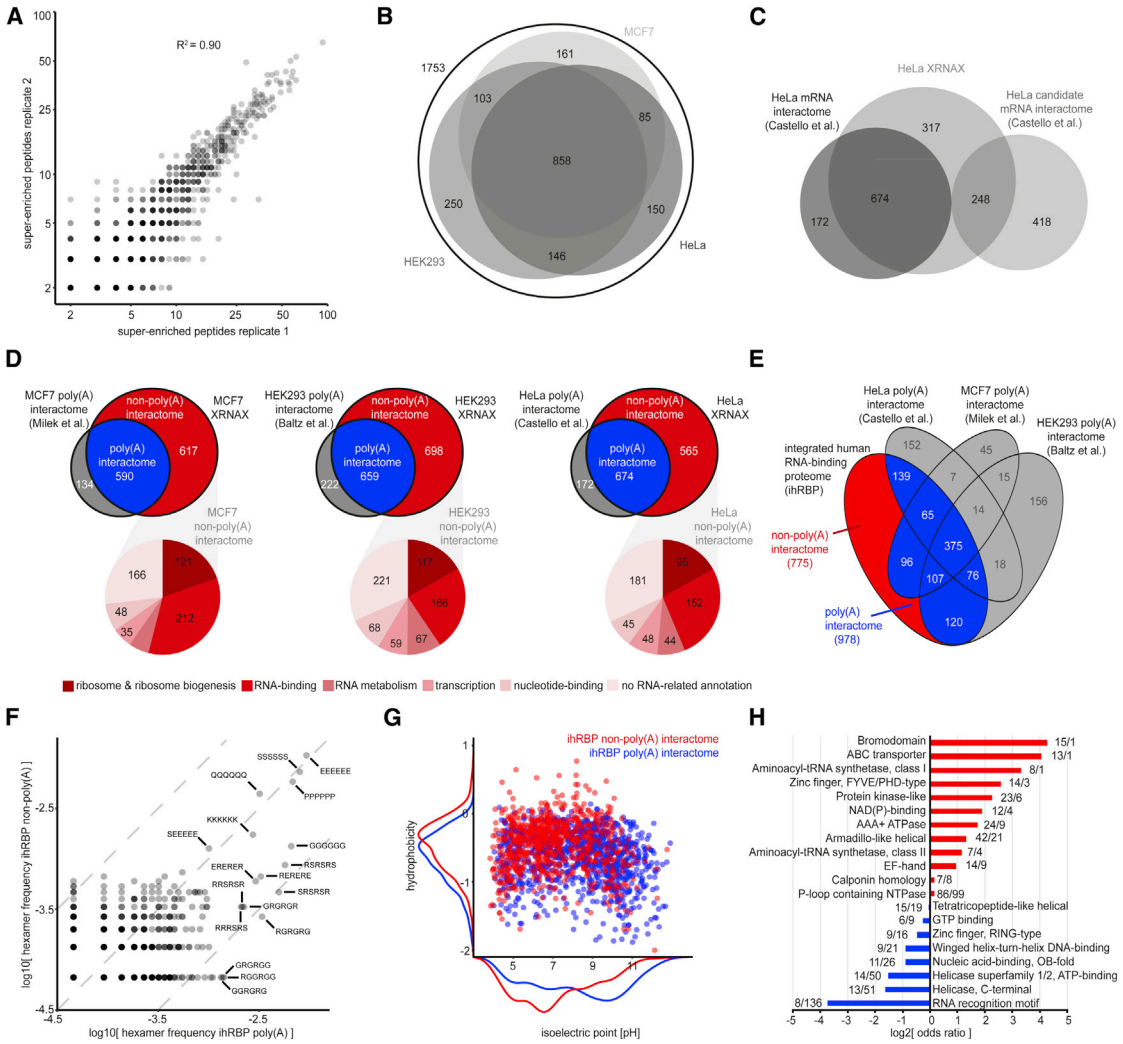


Figure 3. The Integrated Human RNA-Binding Proteome Derived from Three Cell Lines

(A) Scatterplot showing reproducibility of protein identifications by super-enriched peptides. Each point represents one protein and indicates how many super-enriched peptides were found per replicate.

(B) Venn diagram for XRNAX-derived RNA-binding proteomes. Numbers indicate the proteins identified in each of the three cell lines MCF7, HeLa, and HEK293.

(C) Venn diagram comparing the XRNAX-derived HeLa RNA-binding proteome to the published poly(A) RNA interactome by Castello et al. (2012).

(D) Composition of XRNAX-derived RNA-binding proteomes. Top: Venn diagrams comparing XRNAX-derived RNA-binding proteomes to published poly(A) interactomes. Non-poly(A) interactomes (red) were derived by subtraction of poly(A) interactomes (blue). Bottom: pie charts for the functional annotation of proteins in the non-poly(A) interactome into five RNA-related categories. Categories were applied hierarchically in the listed order and proteins assigned to the first category they shared a GO (molecular function) annotation with.

(E) Venn diagram comparing the integrated human RNA-binding proteome (ihRBP) to published poly(A) interactomes. A non-poly(A) RNA interactome (red) was derived by subtraction of the combined known poly(A) interactomes (blue).

(F) Scatterplot comparing hexamer frequencies in poly(A) and non-poly(A) interactomes. All possible hexameric permutations of the amino acids G, S, N, Q, P, E, K, and R were counted in proteins of each group and normalized to the total number of counts. Dashed lines indicate fold-changes of 1 and 10.

(G) Scatterplot comparing isoelectric points and hydrophobicity of proteins in poly(A) and non-poly(A) interactomes. Density plots outside the axes illustrate the distribution for each feature.

(H) Odds ratios of interpro domain occurrences in poly(A) and non-poly(A) interactomes. The ten most frequent domains in either group are compared.

See also Figure S3 and Table S2.

consisting of alternating R and E, R and S, R and G, as well as hexa(G) in the poly(A) interactome (Figures 3F and S3D). Increased frequency of low-complexity motifs in poly(A) binding proteins has been reported before (Calabretta and Richard, 2015; Castello et al., 2016). However, our analysis revealed for RNA-binding proteins in general that, depending on the RNA biotype they bind to (i.e., poly(A) versus non-poly(A) RNA) they contained specific low-complexity motifs more frequently.

Furthermore, poly(A) binders were generally more hydrophobic and had a more alkaline isoelectric point (Figure 3G), whereas

non-poly(A) binders were larger and often carried more negative charges (Figure S3E).

Analysis of Protein Domains Highlights Chromatin Remodelers as Novel RNA-Binders

A comparison of domain occurrence revealed notable commonalities and differences between the poly(A) and the non-poly(A) interacting proteomes (Figure 3H). Interestingly, RRM-containing proteins were almost exclusively found in the poly(A) interactome. As expected, we found both classes of aminoacyl-tRNA synthetase domains strongly enriched in the non-poly(A) interactome, which also contained an overrepresentation of the ABC-transporter-like and the AAA^+ ATPase domains. Surprisingly, the domain with the strongest enrichment in the non-poly(A) interactome was the bromodomain, which very recently was reported to bind enhancer RNA (eRNA) and mediate transcriptional activity (Rahnamoun et al., 2018), explaining its absence in poly(A)-interactome studies and presence in RBR-ID (He et al., 2016) and our XRNAX datasets.

When looking more systematically for chromatin-modifying proteins in the non-poly(A) interactome, we identified TP53 and the DNA-damage regulators TP53BP1, BAX, FANCI, RPA1, DDB1, RIF1, MDC1, as well as BRCA1. TP53BP1 (Francia et al., 2012) and BRCA1 (Ganesan et al., 2002) has been reported to interact with non-coding RNA, whereas RNA-binding by TP53 has been controversial (Riley and Maher, 2007). Notably, we detected 25 super-enriched peptides for TP53BP1, ranking it among the most confidently classified non-poly(A) interactors.

As non-coding RNAs emerge as important players in genome regulation (for review, see Rinn and Chang, 2012) future studies may investigate if the function of chromatin components identified here could be guided by RNA.

Arsenite Induces Translational Arrest and Autophagic Remodeling of the Translation Machinery

We next applied XRNAX to investigate changes in the protein-RNA interactome of arsenite-treated MCF7 cells. Arsenite stress has been thoroughly studied in the context of protein-RNA interactions during translational arrest and the formation protein-RNA complexes known as stress granules (for review, see Buchan and Parker, 2009). While the composition of arsenite-induced stress granules (Jain et al., 2016; Khong et al., 2017) and the effect of arsenite-induced translational arrest on the transcriptome (Andreev et al., 2015) are well understood, very little is known about its effect on the proteome. We used SILAC and azidohomoalanine (AHA) labeling to isolate nascent proteins by click-chemistry and quantify them by MS (Eichelbaum et al., 2012) as direct measure for the effect of arsenite on translation. Indeed, arsenite-induced decrease in translation was already apparent after 10 min and plateaued after 30 min (Figure 4A).

Overall protein expression levels did not change for most proteins, however, a distinct subset of proteins showed a gradual decrease in abundance (Figure 4B; Table S3). GO analysis revealed that downregulated proteins were strongly enriched for translation-related terms emerging already after 5 min of arsenite treatment (Figure 4C). Closer examination of ribosomal proteins, as the most prominently affected group, revealed the specific decrease of cytosolic but not mitochondrial ribosomes (Figure 4D). This was concomitant with the collective downregulation of eukaryotic translation initiation factors (EIFs) with similar kinetics, effectively reducing their expression level to 50% within 30 min (Figure S4A). In addition, other RNA-binding proteins decreased in abundance, most of them with functionalities in protein biosynthesis (Figure S4B; Table S3).

Arsenite-induced protein degradation was markedly decreased upon pre-treatment with the autophagy inhibitor spautin-1 (Figure 4F), while this was not the case after pre-treatment with the proteasome inhibitor bortezomib (Figure S4C). Western blotting against the autophagy marker LC3 confirmed increased autophagic flux over the course of arsenite stress (Figure 4E). Thus, arsenite stress led to strong translational arrest within 30 min and triggered profound, rapid, and selective remodeling of the proteome through autophagy.

Exploring the Dynamics of Protein-RNA Interactions during Arsenite Stress

We next investigated how arsenite stress induced changes in the RNA-bound proteome. MCF7 cells of one SILAC label were challenged with arsenite and compared to untreated cells of the complementary label. Using XRNAX followed by silica enrichment and MS, we recorded data for five time points in a 30-min window. After filtering for the super-enriched peptides identified previously, we quantified the association of 765 proteins with RNA over all time points (Figures 5A and 5B; Table S4). While most proteins did not change their RNA-binding under arsenite stress, several proteins showed significantly decreased association with RNA, whereas the only protein increasing more than 2-fold was TP53BP1 (Figure 5B). The kinetic profile of these proteins showed a steady incline or decline, with the exception of the RNA exosome component EXOSC2, which already showed increased RNA-binding after 5 min of arsenite stress, and then stayed constant.

In order to control for total protein abundances, we intersected this dataset with our data for the total proteomes, which resulted in 619 proteins quantified over all time points (Figures 5C and 5D). Figure 5C illustrates that ~25% of the quantified RNA-binding proteome was affected by the autophagic degradation process, which we had characterized earlier (Figure 4B). Direct comparison of changes in RNA-binding to total protein abundances (Figures 5E and S5A) revealed several remarkable patterns. First, the increase in RNA-binding of the proteins TP53BP1 and EXOSC2 could be entirely attributed to their association with RNA, because their absolute abundance stayed constant over time (Figure 5E, green). Second, the proteins showing the strongest reduction in RNA-binding also decreased in protein abundance (Figure 5E, magenta). However, this decrease was small in comparison to their decreased RNA binding, primarily indicating a release from RNA. Interestingly, this included the ribosomal proteins RPS28, RPS14, and RPS3, which are all positioned in the cleft of the 80S ribosome that directly interacts with mRNA (Figure 5F).

A third striking observation was the 40%–50% decrease in protein abundance of nearly all detected EIFs, without a significant change in RNA binding (Figure 5E, yellow, see also Figure S5B). A similar observation could be made for cytosolic

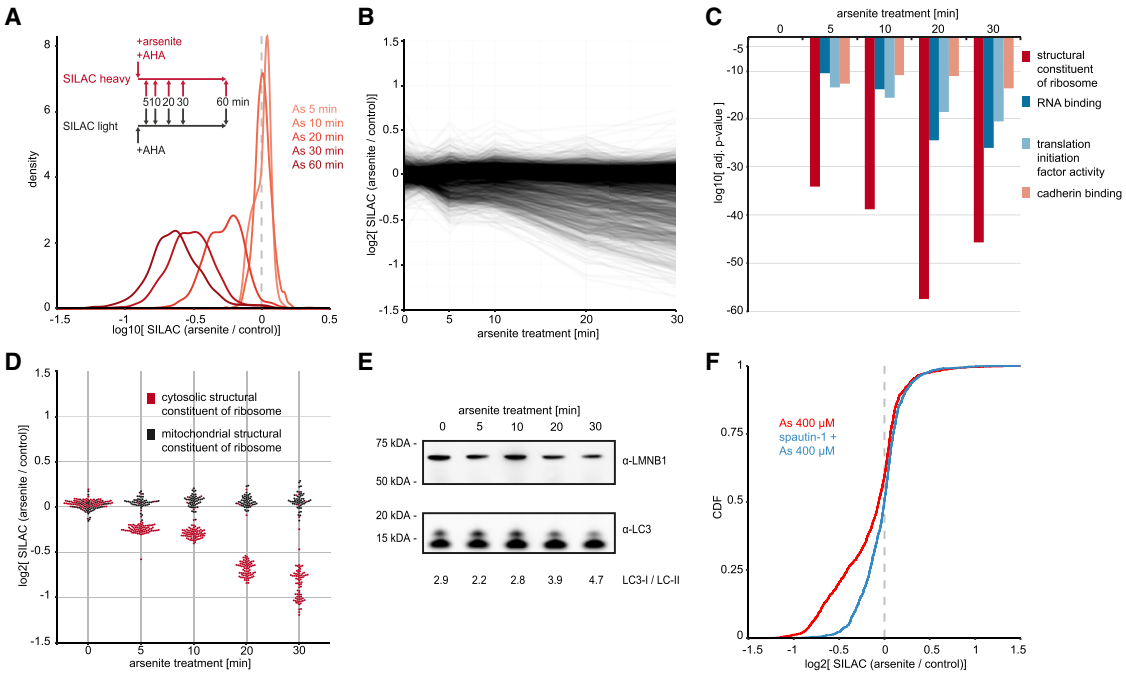


Figure 4. Impact of Arsenite Stress on the Proteome

(A) Density plot showing fold-changes of nascent proteins produced under normal conditions or arsenite stress. Displayed are means of duplicate experiments with label-swap.

(B) Time course of total proteome changes during arsenite stress in MCF7 cells. Each line represents one protein. Values are means of duplicate experiments with label-swap filtered for a variance of <15%.

(C) GO-enrichment analysis (molecular function) of proteins that change in expression during arsenite stress. Shown are the top-4 GO terms with highest significances after 30 min.

(D) Dotplot displaying changes in total proteome upon arsenite stress for proteins under the GO term "structural constituent of ribosome." Each dot represents one protein. Values are means of duplicate experiments with label swap filtered for a variance of <15%.

(E) Western blot for LC3 to monitor autophagic flux in MCF7 cells over 30 min of arsenite treatment. LMNB1 was used as a loading control. Numbers under images quantify the relative intensity of LC3-I and LC3-II bands.

(F) Cumulative distribution of changes in total proteome of MCF7 cells upon arsenite treatment, with (blue) and without (red) inhibition of autophagy by spautin-1 (see also Figure S4G).

See also Tables S3 and S4.

ribosomal proteins other than the ones mentioned above (Table S4), suggesting that RNA-binding might protect from protein degradation.

Two exceptions here were EIF4A3 and EIF2S2. EIF4A3 is one of the RNA-binding components of the exon junction complex and thereby only tangentially involved in translation initiation, which might explain why it did not share the same behavior as the other EIFs. EIF2S2, also known as eIF2-α and core component of the EIF2 complex involved in 43S preinitiation complex formation (for review, see Jackson et al., 2010), increased RNA-binding steadily over all time points, although its protein abundance decreased significantly (Figure S5B). In fact, by normalizing RNA-binding to protein abundance, its effective 2.8-fold increase in RNA-binding was the strongest of all quantified proteins (Table S4). This might indicate that 43S preinitiation complexes assembled on RNA, while translational arrest occurred downstream of this process.

These observations were not affected by changes in the integrity or amount of RNA, because total RNA from arsenite-treated MCF7 cells was neither subject of degradation (Figures S4D and S4E) nor altered turnover (Figure S4F), in line with previous transcriptomic data (Andreev et al., 2015).

In summary, quantification of RNA-binding using XRNAX recapitulated RNA-ribosome dissociation known to occur during translational arrest and added molecular detail on the association of EIFs with RNA.

Linking XRNAX with CLIP-Seq to Identify Protein-Binding RNAs

We next aimed to combine XRNAX with CLIP-seq to validate our proteomic data. Conceptually, XRNAX as a sample preparation step prior to CLIP-seq is advantageous for a number of reasons: (1) xontaminants like DNA, which could physically obstruct immunoprecipitation or mask target protein in chromatin complexes, are eliminated; (2) sample volumes are reduced from milliliters to microliters, thereby allowing for higher antibody concentrations; and (3) RNA fragmentation can be supplemented by high-intensity sonication, thereby circumventing cumbersome optimization and potential biases of RNase treatment (Haberman et al., 2017).

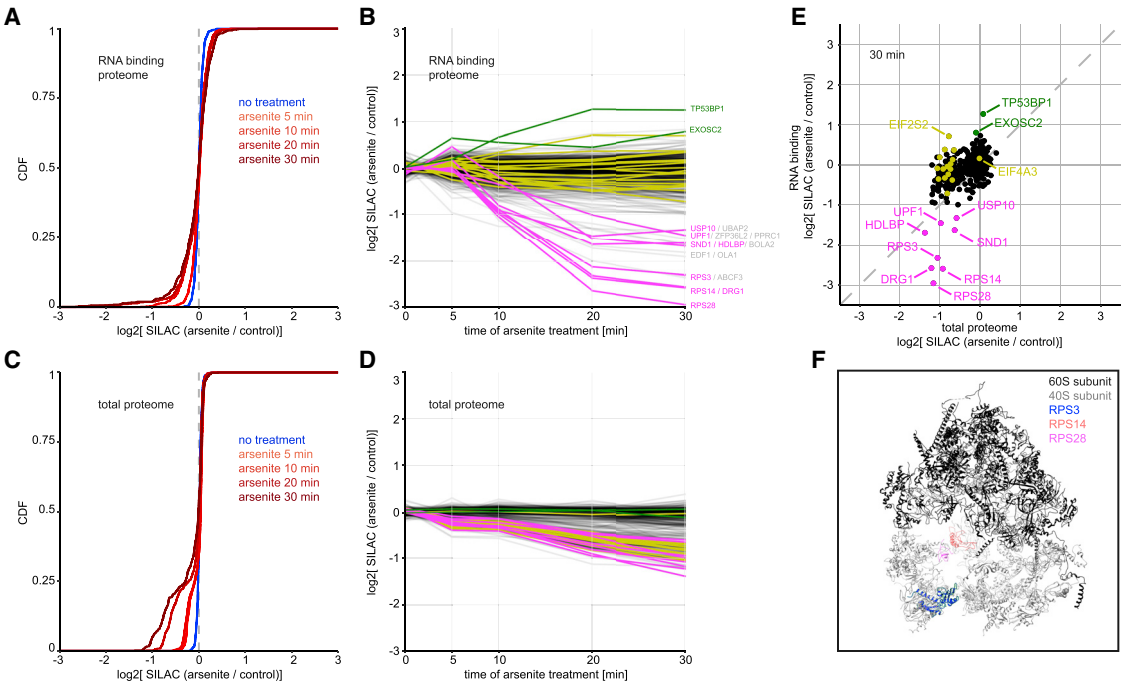


Figure 5. Changes in the Total and RNA-Interacting Proteome Induced by Arsenite-Mediated Translational Arrest

(A) Cumulative changes in the RNA-binding proteome, assessed by XRNAX in MCF7 cells that were treated with arsenite for the indicated times.

(B) As in (A), showing temporal data for individual proteins. Each line represents one protein. Values displayed are means of duplicate experiments with label swap filtered for a variance of <15%.

(C) Cumulative distribution of changes in the total MCF7 proteome upon arsenite stress for proteins displayed in (A).

(D) Timeline of changes in the total proteome upon arsenite stress for proteins displayed in (B).

(E) Scatterplots comparing changes in the total proteome to changes in RNA-binding after 30 min of arsenite stress. For other time points, see Figure S5. Color-coding in (B), (D), and (E) refers only to proteins quantified in both the RNA-binding and total proteome. The proteins TP53BP1 and EXOSC2 are displayed in green, proteins with >50% change in RNA-binding after 30 min are displayed in magenta, and EIFs are displayed in yellow.

(F) Cryo-EM structure of the human ribosome highlighting the location of RPS28, RPS14, and RPS3. For visibility, nucleic acids are not displayed.

See also Table S4.

We selected lamin B1 (LMNB1) as a CLIP target, to validate it as a novel RNA binder, which we had identified among the proteins with the highest number of super-enriched peptides in the non-poly(A) interactome of MCF7 cells (Table S2). We fragmented RNA in an XRNAX extract using ultrasonication and immunoprecipitated LMNB1 using a variation of the eCLIP protocol (Van Nostrand et al., 2016) (Figure S6A; for details, see STAR Methods). RNA sequencing identified various ncRNAs that were significantly enriched over the size-matched input control, primarily snoRNAs but also other small nuclear RNAs (Figure 6A). These experiments showed that XRNAX extracts could serve as direct input for CLIP-seq experiments, demonstrating that XRNAX can be applied both for the discovery and validation of novel RNA-binding proteins by MS and RNA-sequencing, respectively, from the same sample.

Differential RNA-Binding of EXOSC2 upon Arsenite Stress

Our differential quantification of RNA-binding had shown that EXOSC2 rapidly increased its interaction with RNA upon arsenite stress (Figure 5B; Table S4). We performed immunohistochemistry and confocal microscopy to locate EXOSC2 upon arsenite stress in MCF7 cells (Figure 6B). Under normal conditions, EXOCS2 was located in the cytosol as well as the nucleus, changing to primarily nuclear localization already 5 min into arsenite treatment. Notably, our total proteome data had shown that the overall abundance of EXOSC2 (or any other exosomal protein) was not affected by arsenite stress (Table S3). Because the change in RNA-binding (Figure 5B) coincided with nuclear relocalization (Figure 6B), we asked if EXOSC2 had changed RNA interaction partners in the process. Therefore, we performed XRNAX CLIP-seq for EXOSC2 in untreated cells, and cells treated with arsenite for 5 or 30 min. Indeed, we identified particularly nuclear transcripts that increased their association with EXOSC2 upon arsenite stress (Table S5). In addition, on average 65% ($\sigma = 13\%$) of unique reads mapped to ribosomal transcripts, in line with a previous study in yeast (Schneider et al., 2012). Because the exosome participates in processing of the 45S pre-rRNA by trimming it into 18S, 5.8S, and 28S rRNA (for review, see Henras et al., 2015), we analyzed how read coverage of the 45S pre-rRNA by EXOSC2 changed upon arsenite stress. We found very significant enrichment for specific regions (Figure 6C), most notably in segments 3′ of the mature 5.8S transcript, which is the main region known to be degraded by the exosome

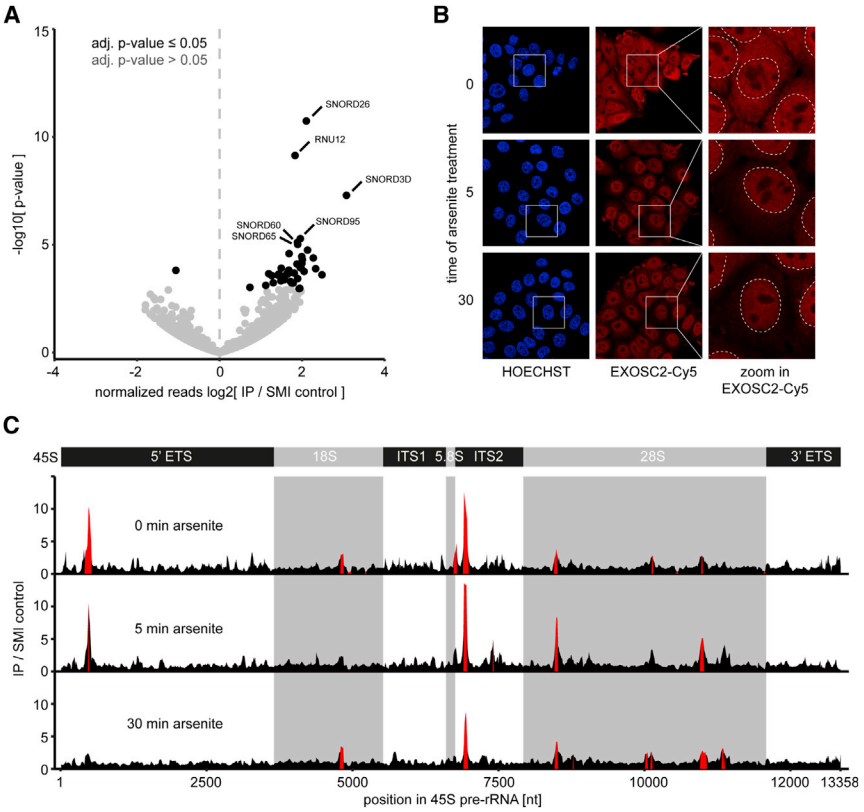


Figure 6. Combination of XRNAX with CLIP-Seq to Identify RNA Bound to Novel RNA-Binding Proteins

(A) Volcano plot showing the enrichment of non-coding nuclear transcripts in XRNAX CLIP-seq for LMNB1 in MCF7 cells.

(B) Immunofluorescence detection of EXOSC2 with confocal microscopy during translational arrest. White squares in the left and middle images mark the magnified areas displayed on the right. White, dashed lines in images on the right indicate the outline of nuclei.

(C) Gene track for the 45S pre-RNA displaying coverage of XRNAX-CLIP-seq for EXOSC2, normalized to the SMI control. IP coverage enriched over the SMI control with an adjusted p value <0.001 is highlighted in red (for details, see STAR Methods). Top scheme indicates location of transcripts that get processed from the 45S pre-mature transcript. Grey shading refer to the mature 18S, 5.8S and 28S transcripts. 5′ ETS, 5′ external transcribed spacer; 3′ ETS, 3′ external transcribed spacer; ITS1, internal transcribed spacer 1; ITS2, internal transcribed spacer 2.

See also Figure S6 and Table S5.

during canonical 5.8S rRNA maturation. Interestingly, two peaks demarcating intermediates known as 7S and 6S pre-rRNA (Tafforeau et al., 2013), and a peak 5′ of the mature 18 S transcript, which is known to be cooperatively degraded by the exonuclease XRN2 and the exosome (Sloan et al., 2014), were highly prominent in untreated cells but decreased after 30 min of arsenite stress (Figure 6C). Thus, our data suggested that EXOSC2 was imported into the nucleus upon arsenite stress in order to promote rRNA maturation.

DISCUSSION

We have developed XRNAX as a generic method for the biochemical purification of protein-crosslinked RNA, making it accessible for transcriptomic and proteomic applications to chart the composition, dynamics, and interaction sites of protein-RNA complexes. Importantly, this should provide a tool to understand how protein-RNA interactions underlie cellular processes and phenotypes.

Characterization of global RNA-bound proteomes confirmed the large majority of proteins previously shown to interact with poly(A) RNA (Figure 3D; Table S2). Beyond this, we identified many hundreds of proteins that we designated as non-poly(A) interactors, many of which had never before been reported to interact with RNA. This included a group of bromodomain proteins, reported to bind eRNAs (Rahnamoun et al., 2018), expanding the collection of chromatin-modifying complexes known to interact with ncRNA (e.g., PRC2, CoREST, or SMCX) (Khalil et al., 2009), and reinforcing the notion that ncRNAs may play a more wide-spread role in regulating gene expression than previously thought. In addition, we found that low-complexity motifs were enriched in poly(A)-binding proteins, providing a potential link to molecular self-organization through liquid droplet formation. In particular, many of the poly(A)-binding proteins harboring IDRs, such as TIA, hnRNPA2, or FUS, have been implicated in the nucleation of mRNA in macroscopic complexes such as stress granules (Markmiller et al., 2018), which have been shown to include mRNA and exclude ncRNA (Khong et al., 2017).

When investigating arsenite-induced translational arrest, we observed rapid and specific degradation of the translational machinery (Figure 4C and D), highly reminiscent of a process coined as ribophagy in yeast (Kraft et al., 2008). While a recent report observed protein degradation and induction of autophagy upon arsenite stress in yeast (Guerra-Moreno et al., 2015), these processes were never causally linked or observed in a mammalian system. Notably, our report of massive, arsenite-induced protein degradation on a timescale of minutes is phenomenologically different to reports of ribophagy upon nutrient starvation, which were demonstrated to occur in human cells on a timescale of hours to days (Gretzmeier et al., 2017).

Ribophagy in yeast depends on Ubp3 and Bre5 (Kraft et al., 2008), the orthologs of the human stress granule markers USP10 and G3BP1. This is interesting because we identified USP10 among the proteins with the most prominent decrease in RNA-binding upon arsenite stress (Figure 5E), suggesting that this may be the missing link to explain the previously observed role of USP10 in counteracting arsenite-induced oxidative stress (Takahashi et al., 2013). Notably, USP10 is one of the two known

targets of the autophagy inhibitor spautin-1 (Liu et al., 2011) thus closely tying together USP10 as an RNA-binding protein, a mediator of autophagy, and a constituent of stress granules.

Beyond the shown applications of XRNAX, we anticipate that many more can be explored in the future. First, XRNAX is not limited to human cells. UV-crosslinking (Darnell, 2010) and TRIZOL extraction have been successfully applied to bacteria, viruses, yeast, plant, and animal tissue (Chomczynski and Sacchi, 2006), making them readily accessible for XRNAX. We envision that especially infection biology, involving non-adenylated bacterial and viral RNA, will greatly benefit from our methodology. Second, XRNAX can be interfaced with the wide variety of CLIP-seq methodologies (Lee and Ule, 2018) for detailed characterization of particular protein-RNA interaction sites. Third, the purity of XRNAX extracts allows for the intersection with methods in chemical biology (e.g., to selectively isolate RNA or protein carrying post-transcriptional or post-translational modifications, respectively). This includes enzymatic reactions, where as a proof-of-concept we were able to biotinylate RNA in XRNAX extracts using poly(U) polymerase (data not shown). Fourth, XRNAX offers an efficient and scalable procedure for the extraction of RNA-peptide hybrids. This may spark MS-based strategies for simultaneous sequencing of peptides and RNA in crosslinked hybrids to ultimately describe a global RNA-protein interaction network in one experiment. Fifth, complementary sampling of the full protein-bound transcriptome (by RNA-seq and XRNAX-RNA-seq, respectively) may allow for much better sequencing depth, to identify transcripts that change their association with proteins as a consequence of cellular perturbations. Finally, the purity of XRNAX extracts should be an excellent starting point to identify proteins that interact with individual RNA species, which should be appealing especially for many ncRNAs whose function remain to be established.

Collectively, XRNAX is a versatile tool to illuminate many uncharted areas of the transcriptome and the RNA-binding proteome.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, five tables, and one methods file and can be found with this article online at https://doi.org/10.1016/j.cell.2018.11.004.

ACKNOWLEDGMENTS

We thank Stefan Wilkening (DKFZ Heidelberg) for access to a Covaris ultrasonicator. We thank Vladimir Benes, Bettina Hase and Nayara Azevedo (Genecore EMBL Heidelberg) for RNA sequencing as well as advice and discussion. We thank Kim Remans and Jacob Scheurich (EMBL PepCore) for the production of C7orf50 and its DUF2373 in *E. coli*. We thank Markus Landthaler and Miha Milek for comprehensive access to their MCF7 interactome capture dataset. We thank Christian Frese, Gianluca Sigismondo, and Gertjan Kramer for continuous support and discussion. We acknowledge financial support by the Excellence Cluster CellNetworks (to J.K.).

AUTHOR CONTRIBUTIONS

Conceptualization, J.T. and J.K.; Methodology, J.T. and J.K.; Formal Analysis, J.T.; Formal Analysis of DUF2373, A.P. and A.B.; PNK assay, R.H.; Formal Analysis of XRNAX-CLIP-seq, T.S.; Investigation, J.T.; Writing – Original Draft,

J.T. and J.K.; Writing – Review & Editing, J.T., M.H., and J.K.; Visualization, J.T.; Supervision, J.K.; Funding Acquisition, J.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 4, 2018
Revised: September 21, 2018
Accepted: October 31, 2018
Published: December 6, 2018

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andreev, D.E., O'Connor, P.B., Fahey, C., Kenny, E.M., Terenin, I.M., Dmitriev, S.E., Cormican, P., Morris, D.W., Shatsky, I.N., and Baranov, P.V. (2015). Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. eLife *4*, e03971.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202-8.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA *6*, 11.

Brangwynne, C.P., Tompa, P., and Pappu, R.V. (2015). Polymer physics of intracellular phase transitions. Nat. Phys. *11*, 899–904.

Buchan, J.R., and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. Mol. Cell *36*, 932–941.

Calabretta, S., and Richard, S. (2015). Emerging roles of disordered sequences in RNA-binding proteins. Trends Biochem. Sci. *40*, 662–672.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive identification of RNA-binding domains in human cells. Mol. Cell *63*, 696–710.

Chomczynski, P., and Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nat. Protoc. *1*, 581–585.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip. Rev. RNA *1*, 266–286.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Eichelbaum, K., Winter, M., Berriel Diaz, M., Herzig, S., and Krijgsveld, J. (2012). Selective enrichment of newly synthesized proteins for quantitative secretome analysis. Nat. Biotechnol. *30*, 984–990.

Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., and d'Adda di Fagagna, F. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature *488*, 231–235.

Ganesan, S., Silver, D.P., Greenberg, R.A., Avni, D., Drapkin, R., Miron, A., Mok, S.C., Randrianarison, V., Brodie, S., Salstrom, J., et al. (2002). BRCA1 supports XIST RNA concentration on the inactive X chromosome. Cell *111*, 393–405.

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol. Cell. Proteomics *11*, M111.014050.

Girardot, C., Scholtalbers, J., Sauer, S., Su, S.Y., and Furlong, E.E.M. (2016). Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. BMC Bioinformatics *17*, 419.

Gretzmeier, C., Eiselein, S., Johnson, G.R., Engelke, R., Nowag, H., Zarei, M., Küttner, V., Becker, A.C., Rigbolt, K.T.G., Høyer-Hansen, M., et al. (2017). Degradation of protein translation machinery by amino acid starvation-induced macroautophagy. Autophagy *13*, 1064–1075.

Guerra-Moreno, A., Isasa, M., Bhanu, M.K., Waterman, D.P., Eapen, V.V., Gygi, S.P., and Hanna, J. (2015). Proteomic analysis identifies ribosome reduction as an effective proteotoxic stress response. J. Biol. Chem. *290*, 29695–29706.

Haberman, N., Huppertz, I., Attig, J., König, J., Wang, Z., Hauer, C., Hentze, M.W., Kulozik, A.E., Le Hir, H., Curk, T., et al. (2017). Insights into the design and interpretation of iCLIP experiments. Genome Biol. *18*, 7.

Hastie, N.D., and Bishop, J.O. (1976). The expression of three abundance classes of messenger RNA in mouse tissues. Cell *9*, 761–774.

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. Mol. Cell *64*, 416–430.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.F., Chakraborty, A., and Gleizes, P.E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA *6*, 225–242.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nat. Rev. Mol. Cell Biol. *19*, 327–341.

Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., and Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. Mol. Syst. Biol. *10*, 757.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-modulated stress granules contain a diverse proteome and substructure. Cell *164*, 487–498.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA *106*, 11667–11672.

Khatter, H., Myasnikov, A.G., Natchiar, S.K., and Klaholz, B.P. (2015). Structure of the human 80S ribosome. Nature *520*, 640–645.

Khong, A., Matheny, T., Jain, S., Mitchell, S.F., Wheeler, J.R., and Parker, R. (2017). The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. Mol. Cell *68*, 808–820.

Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nat. Methods *14*, 513–520.

Kraft, C., Deplazes, A., Sohrmann, M., and Peter, M. (2008). Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. Nat. Cell Biol. *10*, 602–610.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and

high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, F.C.Y., and Ule, J. (2018). Advances in CLIP technologies for studies of protein-RNA interactions. Mol. Cell *69*, 354–369.

Liu, J., Xia, H., Kim, M., Xu, L., Li, Y., Zhang, L., Cai, Y., Norberg, H.V., Zhang, T., Furuya, T., et al. (2011). Beclin1 controls the levels of p53 by regulating the deubiquitination activity of USP10 and USP13. Cell *147*, 223–234.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lun, A.T.L., and Smyth, G.K. (2015). Csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. *44*, e45.

Markmiller, S., Soltanieh, S., Server, K.L., Mak, R., Jin, W., Fang, M.Y., Luo, E.C., Krach, F., Yang, D., Sen, A., et al. (2018). Context-dependent and disease-specific diversity in protein interactions within stress granules. Cell *172*, 590–604.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Milek, M., Imami, K., Mukherjee, N., De Bortoli, F., Zinnall, U., Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., et al. (2017). DDX54 regulates transcriptome dynamics during DNA damage response. Genome Res. *27*, 1344–1359.

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*, aab2276.

Müller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat. Rev. Genet. *14*, 275–287.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Rahnamoun, H., Lee, J., Sun, Z., Lu, H., Ramsey, K.M., Komives, E.A., and Lauberth, S.M. (2018). RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. Nat. Struct. Mol. Biol. *25*, 687–697.

R Development Core Team (2011). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

Riley, K.J.-L., and Maher, L.J., 3rd. (2007). p53 RNA interactions: new clues in an old mystery. RNA *13*, 1825–1833.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., and Tollervey, D. (2012). Transcriptome-wide analysis of exosome targets. Mol. Cell *48*, 422–433.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Sloan, K.E., Bohnsack, M.T., Schneider, C., and Watkins, N.J. (2014). The roles of SSU processome components and surveillance factors in the initial processing of human ribosomal RNA. RNA *20*, 540–550.

Steube, A., Schenk, T., Tretyakov, A., and Saluz, H.P. (2017). High-intensity UV laser ChIP-seq for the study of protein-DNA interactions in living cells. Nat. Commun. *8*, 1303.

Tafforeau, L., Zorbas, C., Langhendries, J.L., Mullineux, S.T., Stamatopoulou, V., Mullier, R., Wacheul, L., and Lafontaine, D.L.J. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. Mol. Cell *51*, 539–551.

Takahashi, M., Higuchi, M., Matsuki, H., Yoshita, M., Ohsawa, T., Oie, M., and Fujii, M. (2013). Stress granules inhibit apoptosis by reducing reactive oxygen species production. Mol. Cell. Biol. *33*, 815–829.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Lamin B1 rabbit polyclonal antibody	proteintech	12987-1-AP, RRID:AB_2136290
EXOSC2 rabbit polyclonal antibody	proteintech	14805-1-AP, RRID:AB_2101837
EXOSC2 mouse monoclonal antibody	proteintech	66099-1-Ig
Goat anti-mouse Cy5	abcam	ab6563, RRID:AB_955068
Anti-FLAG M2 Magnetic Beads	Sigma	M8823, RRID:AB_2637089
HA-tag antibody	proteintech	51064-2-AP, RRID:AB_11042321
Chemicals, Peptides, and Recombinant Proteins		
Dialysed FBS	GIBCO	26400-044
Pen-Strep	GIBCO	15140-122
DMEM for SILAC	Silantes	280001300
SILAC heavy L-lysine ($^{13}C_6$,$^{15}N_2$-L-Lysine HCl)	Silantes	211604102
SILAC heavy L-arginine ($^{13}C_6$,$^{15}N_4$-L-Arginine HCl)	Silantes	201604102
GlutaMAX	GIBCO	35050061
TRI reagent	Sigma	T9424
GlycoBlue	Ambion	AM9515
4-thiouridine (4SU)	biomol	Cay-16373
SYBRSafe	invitrogen	S33102
EDTA-free protease inhibitor	Sigma (Roche)	11873580001
Benzonase	Novagen	70664
SP3 beads	GE	44152105050250
Trypsin/LysC	Promega	V5073
NEB DNase buffer 10 x	NEB	B0303S
NEB DNase	NEB	M0303L
RNASin Plus RNase inhibitor	Promega	N2615
RNase A	Thermo	EN0531
RNase I	Ambion	AM2295
RNase T1	Thermo	EN0541
Sodium arsenite (50 mM solution in water)	Santa Cruz	sc-301816
Spautin-1	Sigma	SML0440
Rapamycin (2.5 mg/ml in DMSO)	Sigma	R8781
Ethynyl-uridine (EU)	Jena Biosciences	CLK-N002
RNA polymerase I inhibitor CX5461	Millipore	509265
sulfo-Cy5-azide	Jena Biosciences	CLK-AZ118
HOECHST33342 (20 mM in water)	Thermo	62249
ProlongGold antifade mountant	invitrogen	P36934
Proteinase K	Thermo	EO0491
FastAP Thermosensitive Alkaline Phosphatase	Thermo	EF0651
T4 Polynucleotide Kinase	Thermo	EK0032
Protein G Magnetic Beads	Pierce	88847
Lipofectamine 3000	invitrogen	L3000008
Turbo DNase	Ambion	AM2238
Critical Commercial Assays		
Quiagen RNeasy Midi Kit	Quiagen	75144
NEXTflex Small RNA Sequencing Kit	Bioo Scientific	NOVA-5132

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
RNA sequencing data	This paper, EMBL-EBI ENA	PRJEB26441
XRNAX CLIP-seq data	This paper, EMBL-EBI ENA	PRJEB26442
Proteomics data	This paper, ProteomeXchange	PXD010520
Human proteome (search term: 'reviewed:yes AND proteome:up000005640', 20216 entries, retrieved 11 September 2017)	Uniprot	UP000005640
Human reference genome hg19	GENCODE	Release 19 (GRCh37.p13)
Human reference genome hg38	GENCODE	Release 29 (GRCh38.p12)
HeLa poly(A)-binding proteome	Castello et al., 2012	NA
HEK293 poly(A)-binding proteome	Baltz et al., 2012	NA
MCF7 poly(A)-binding proteome	Milek et al., 2017	NA
Deep MCF7, HeLa, HEK293 proteomes	Geiger et al., 2012	NA
Experimental Models: Cell Lines		
Human (female): MCF7 cells	ATCC	RRID:CVCL_0031
Human (female): HEK293 cells	Laboratory of Rolf Sprengel (MPI Heidelberg)	RRID:CVCL_0045
Human (female): HeLa cells	ATCC	RRID:CVCL_0030
Recombinant DNA		
pcDNA5 FRT TO c7orf50 full length FLAG-HA c-ter	This laboratory	N/A
pcDNA5 FRT TO DUF2373 FLAG-HA c-ter	This laboratory	N/A
pcDNA5 FRT TO DUF2373_K120G FLAG-HA c-ter	This laboratory	N/A
pcDNA5 FRT TO c7orf50 full length FLAG-HA n-ter	This laboratory	N/A
pcDNA5 FRT TO DUF2373 FLAG-HA n-ter	This laboratory	N/A
pcDNA5 FRT TO DUF2373_R156G FLAG-HA n-ter	This laboratory	N/A
Software and Algorithms		
GOrilla	Eden et al., 2009	http://cbl-gorilla.cs.technion.ac.il/
BBMap (37.68)	JGI	https://sourceforge.net/projects/bbmap
Je (version 1.2)	Girardot et al., 2016	https://git.embl.de/grp-gbcs/Je
STAR (version 2.5.0a)	Dobin et al., 2013	https://github.com/alexdobin/STAR, RRID:SCR_015899
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
HTSeq	Anders et al., 2015	https://htseq.readthedocs.io/en/release_0.9.1/, RRID:SCR_005514
UCSF Chimera (1.12)	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/, RRID:SCR_004097
MaxQuant (1.5.1.2)	Cox and Mann, 2008	http://www.biochem.mpg.de/5111795/maxquant, RRID:SCR_014485
MSFragger	Kong et al., 2017	http://www.nesvilab.org/software.html
R	R Development Core Team, 2011	https://www.r-project.org/, RRID:SCR_001905
RStudio (0.99.903, RRID:SCR_000432)	RStudio: Integrated Development for R. RStudio, Boston, MA	https://www.rstudio.com/, RRID:SCR_000432
csaw (R)	Lun and Smyth, 2015	http://bioconductor.org/packages/release/bioc/html/csaw.html
DESeq2 (R)	Love et al., 2014	http://bioconductor.org/packages/release/bioc/html/DESeq2.html, RRID:SCR_015687
Additional Resources		
Advanced online documentation for XRNAX protocols	This paper	https://www.xrnax.com/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jeroen Krijgsveld (j.krijgsveld@dkfz.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mammalian Cell Culture and Stable Cell Lines

The cell lines MCF7, HEK293 and HeLa (all human, female) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) for SILAC supplemented with 10% dialysed FBS and Pen-Strep (100 U / ml penicillin, 100 mg / ml streptomycin) at 37°C, 5% CO_2. DMEM for SILAC was supplemented with 1 mM L-lysine and 0.5 mM L-arginine of the individual SILAC labels as well as 1.7 mM light L-proline and 1 x GlutaMAX. The heavy SILAC label was introduced during six passages in heavy DMEM for SILAC.

All experiments were performed on MCF7 cells, except when deriving the integrated human RNA-binding proteome (ihRBP), where MCF7, HEK293 and HeLa cells were used as indicated in the text.

METHOD DETAILS

Guanidinium Thiocyanate–Phenol–Chloroform (TRIZOL) Extraction

Up to 10 million MCF7 cells were lysed in 1 mL TRI reagent by pipetting up and down. For phase-separation, 200 µl chloroform was added and samples mixed by turning tubes upside down several times. After 5 minutes incubation at room temperature, samples were spun down with 12000 g for 10 minutes at 4°C. Approx. 400 µl of the aqueous phase was transferred to a fresh tube, NaCl was added to a final concentration of 300 mM along with 1 µl GlycoBlue. Samples were combined with 500 µl isopropanol, mixed by inversion and RNA precipitated by centrifugation with 18000 g for 15 minutes at −10°C.

The supernatant was removed and the RNA pellet washed with 1 mL of 70% ethanol before resuspension in the desired volume of nuclease-free water.

UV-Crosslinking of Cells

All cells were grown in 245 mm x 245 mm dishes to the desired confluence. For the incorporation of 4-thiouridine (4SU) into RNA, cells were incubated with 100 µM 4SU for 16 hours prior to UV-crosslinking. Media was decanted and cells washed with 50 mL ice-cold PBS. In order to remove as much liquid as possible dishes were propped up straight and residual PBS drained onto a paper towel through gravity. UV-crosslinking occurred on ice with 200 mJ / cm^2 at 254 nm wavelength with a BIO-LINK UV-crosslinker (Vilber). Cells that had incorporated 4SU were UV-crosslinked at 365 nm wavelength. Subsequently, cells were harvested into ice-cold PBS, pelleted and either directly subjected to XRNAX or stored at −80°C for up to 14 days.

Protein-Crosslinked RNA Extraction (XRNAX)

Up to 100 million cells (typically one confluent 245 mm x 245 mm dish of UV-crosslinked MCF7, HEK293 or HeLa cells combined with one confluent 245 mm x 245 mm dish of non-crosslinked cells) were lysed in 8 mL TRI reagent by pipetting up and down. Cell clumps were disintegrated by flushing the lysate repeatedly against the wall of the tube. Lysis was further facilitated by incubation on a rotating wheel for 5 minutes at room temperature. Lysates were combined with 1.6 mL chloroform, mixed by inversion and incubated for 5 minutes at room temperature. Tubes were spun down with 7000 g for 10 minutes at 4°C.

The aqueous phase was removed and the interphase transferred to a 2 mL tube. The interphase was gently washed twice with 1 mL low SDS buffer (tris-Cl 50 mM, EDTA 1 mM, SDS 0.1%), flushing protein off the walls of the tube while retaining the integrity of the interphase flakes. Flakes were spun down with 5000 g for 2 minutes at room temperature and the supernatant discarded. After the washing, flakes were disintegrated by pipetting into 1 mL of low SDS buffer. The disintergrated interphase was spun down with 5000 g for 2 minutes at room temperature and the supernatant saved as interphase eluate 1. Disintegration of the interphase was repeated with another 1 mL of low SDS buffer, then twice with 1 mL of high SDS buffer (tris-Cl 50 mM, EDTA 1 mM, SDS 0.5%) each time yielding approx. 1 mL of interphase eluates.

NaCl was added to a final concentration of 300 mM to each of the 4 interphase eluates, along with 1 µl GlycoBlue and 1 mL isopropanol before mixing by inversion. Samples were spun down for 15 minutes with 18000 g at −10°C. The supernatants were discarded and pellets from all four elutes were combined in 2 mL of 70% ethanol. The combined sample was again centrifuged for 1 minute with 18000 g at room temperature, supernatant discarded and all residual ethanol removed. The pellet was taken up in 1.8 mL of nuclease-free water and detached from the wall of the tube with a pipette tip. The pellet was allowed to swell for 1 hour on ice with occasional mixing by inversion and eventually dissolved by pipetting.

200 µl NEB DNase I buffer 10 x was added along with 2 µl RNasin Plus, 100 µl NEB DNase and incubated for 60 minutes at 37°C and 700 rpm shaking. Subsequently, the sample was isopropanol precipitated as described above without further addition of GlycoBlue. Pellets were taken up in 1000 µl nuclease-free water and dissolved by pipetting. RNA concentration was estimated by UV-spectroscopy on a NanoDrop One UV photospectrometer (Thermo Scientific), neglecting adsorbtion by protein. Purification of protein-free RNA from XRNAX extracts after proteinase K digestion showed that this estimation was within 15% of the actual RNA content. All

amounts of XRNAX extracts mentioned in the following are given in μg of RNA referring to this estimation and do not take protein content into account.

A detailed, photo-documented version of the XRNAX protocol is included as Methods S1. For updates on XRNAX and its applications visit https://www.xrnax.com.

Comparison of RNA Extracts Using Agarose Gel-Electrophoresis

To verify the integrity of RNA extracted by TRIZOL or XRNAX, agarose-gel electrophoresis was performed using 1% agarose in TBE and SYBRSafe staining. Specifically, for Figure 1B 0.05% of the total yield extracted from 10 million MCF7 cells using the indicated method was subjected to the indicated treatment. Samples were denatured in RNA gel loading dye containing formamide for 2 minutes at 85°C and run for 40 minutes with 3 W.

RNA Sequencing for Quantifying the Relative Composition of RNA Extracts

For RNA sequencing, 10 μg RNA (as determined by NanoDrop UV-spectroscopy neglecting the protein content of samples) were digested for 30 minutes at 55°C in proteinase K buffer (50 mM tris-Cl, EDTA 5 mM, NaCl 150 mM, SDS 1%) using 10 μl proteinase K. Note that both TRIZOL and XRNAX extracted samples were treated identically. Subsequently, RNA was cleaned-up using the RNeasy mini kit (QIAGEN) and was ready for RNA-Seq library preparation.

Specifically, for TRIZOL and XRNAX extracted RNA derived from MCF7 cells that were crosslinked at 254 nm wavelength (Figure 1C) RNA library preparation occurred with the TruSeq RNA Library Prep Kit v2 (Illumina, not stranded) after conditional depletion of ribosomal RNA (rRNA) using the Ribo-Zero rRNA Removal Kit (Illumina). Biological duplicates extracted with TRIZOL or XRNAX (4 samples in total) were barcoded to be sequenced in one lane on a HiSeq2000.

For TRIZOL and XRNAX-extracted RNA from 4SU-labeled MCF7 cells, library preparation occurred with the TruSeq Stranded Total RNA kit (Illumina, stranded) after depletion of rRNA using the Ribo-Zero rRNA Removal Kit. Biological duplicates extracted with TRIZOL or XRNAX (4 samples in total) were barcoded to be sequenced in one lane on a HiSeq2000.

Proteomic Sample Preparation

For mass spectrometry (MS) sample preparation, a modification of the SP3 protocol described by Hughes et al. (2014) was used. For total proteome analysis approx. 1 million cells were lysed and reduced in 1 mL lysis buffer (tris-Cl 50 mM, DTT 10 mM, SDS 0.05%) at 95°C, 700 rpm shaking for 30 minutes. For samples other than cells, e.g XRNAX extracts, samples were brought to a total volume of 100 μl with MilliQ water and combined with 900 μl lysis buffer before reduction at 95°C, 700 rpm shaking for 30 minutes. Magnesium chloride (final concentration of 5 mM), CAA (20 mM), and EDTA-free protease inhibitor were added and mixed before addition of 1 μl of benzonase. Subsequently, digestion of nucleic acids and alkylation occurred for 2 hours at 37°C, 700 rpm shaking. 400 μl SP3 beads were preconditioned by washing with MilliQ water 3 times, before reconstitution in 1 mL MilliQ water. EDTA was added to 10 mM final concentration along with 1% SDS and 20 μl SP3 beads. Samples were vortexed vigorously and subsequently combined with 1 mL acetonitrile. Samples were mixed again and incubated for 15 minutes at room temperature for protein binding to occur. The beads were collected on a magnetic stand for 2 minutes and supernatants decanted. While on the magnetic stand beads were then washed 3 times with 2 mL ethanol 70%, which was added for 1 minute and subsequently decanted. Residual ethanol was removed and beads were taken up in the desired digestion volume of TEAB 20 mM and adequate amounts of trypsin/LysC added to the solution (for total proteomes from 1 million cells 1 μg trypsin/LysC in 100 μl TEAB). Samples were digested at 37°C, 700 rpm shaking overnight. For single run analysis formic acid was added to a final concentration of 1% and samples spun down for 5 minutes with 20000 g. Supernatants were transferred to fresh tubes without disturbing the pellet and analyzed by high-pressure, liquid chromatography (HPLC).

High pH reversed-phase fractionation occurred under standard settings described below. Of the 40 collected fractions the initial 8 fractions up to approx. 18% B were discarded, the following 32 fractions were combined to 8 using the scheme 1+9+17+25/…/8+16+24+32. The combined fractions were dried by SpeedVac and taken up in 1% formic acid before analysis by HPLC-MS.

High pH Reversed-Phase Fractionation of Proteomic Samples

Fractionation at high pH occurred on an Agilent Infinity 1260 LC system (Agilent) using a Phenomenex Gemini 3 μM C18, 100 × 1 mm column (Phenomonex). Buffer A was NH_4COOH 20 mM, buffer B was 100% acetonitrile. The following gradient was used for all applications described in this manuscript: 0-2 minutes 0% B, 2-60 minutes linear gradient to 65% B, 61-62 minutes linear gradient to 85% B, 62-67 minutes 85% B, 67-85 minutes 0% B. Eluates were collected in 40 fractions and combined as described in the individual paragraphs.

HPLC-MS for the Analysis of Nucleotide-Crosslinked Peptides, Discovery of RNA-Binding Proteins or the Differential Quantification of RNA-Binding

Separation by HPLC prior to MS occurred on an Easy-nLC1200 system (Thermo Scientific) using an Acclaim PepMap RSCL 2 μM C18, 75 μm x 50 cm column (Thermo Scientific) heated to 45°C with a MonoSLEEVE column oven (Analytical Sales and Services). Buffer A was 0.1% formic acid, buffer B was 0.1% formic acid in 80% acetonitrile. The following gradient was used for all applications described in this manuscript: 0 minutes 3% B, 0-4 minutes linear gradient to 8% B, 4-6 minutes linear gradient to 10% B, 6-74 minutes

linear gradient to 32% B, 74-86 minutes linear gradient to 50% B, 86-87 minutes linear gradient to 100% B, 87-94 minutes 100% B, 94-95 linear gradient to 3% B, 95-105 minutes 3% B.

Single-run, total proteome analysis was performed on a Fusion Orbitrap mass spectrometer (Thermo Scientific). MS1 detection occurred in orbitrap mode at 60000 resolution, AGC target 1E6, maximal injection time 50 ms and a scan range of 375-1500 DA. MS2 detection occurred with an HCD collision energy of 33 in ion trap top20 mode with an isolation window of 1.6 Da, AGC target 1E4 and maximal injection time of 50 ms.

Detection of XRNAX-derived nucleotide-crosslinked peptides, XRNAX-derived RNA-binding proteomes and XRNAX-derived differential analysis of RNA-binding upon arsenite stress, as well as all analysis of fractionated total proteome samples was performed on a QExactive HF mass spectrometer (Thermo Scientific). MS1 detection occurred at 120000 resolution, AGC target 3E6, maximal injection time 32 ms and a scan range of 350-1500 DA. MS2 occurred with stepped NCE 26 and detection in top20 mode with an isolation window of 2 Da, AGC target 1E5 and maximal injection time of 50 ms.

Isolation of Nucleotide-Crosslinked Peptides from XRNAX Extracts

For the isolation of nucleotide-crosslinked peptides, 1000 μg of XRNAX extract were produced from MCF7 cells using the extraction method described above (from 2 confluent 245 mm x 245 mm dishes). Two aliquots of 500 μg of XRNAX extract were brought to 950 μl final volume containing 50 mM tris-Cl, 0.1% SDS and 10 mM DTT. 10 μg trypsin/LysC was added to each aliquot to a final volume of 1 mL and digestion occurred for 1 hour at 37°C, 700 rpm shaking. CAA was added to a final concentration of 20 mM and digestion continued for another hour. Purification of peptide-crosslinked RNA from the digests occurred by silica column purification using the QIAGEN RNeasy Midi Kit with modified protocol (refer to kit manual for buffer descriptors). 1 mL digest was combined with 3.5 mL buffer RLT in a 15 mL falcon tube, mixed by inversion and heated to 60°C for 15 min. The sample was allowed to reach room temperature. 2.5 mL of 100% ethanol was added, the sample mixed by inversion and applied to an RNeasy Midi column by centrifugation with 3000 g for 5 minutes. Washing occurred twice with 2.5 mL buffer RPE, buffer RW1 was not used. Elution occurred twice with 250 μl nuclease-free water. All eluates combined to approx. 900 μl, which were transferred to a fresh tube. NaCl was added to a final concentration of 300 mM along with 1 μl glycoblue, 1 mL isopropanol, the sample mixed by inversion and incubated for 1 hour at −20°C. Precipitation occurred by centrifugation with 18000 g at −10°C for 60 minutes. The supernatant was discarded and the pellet washed with 70% ethanol. All residual ethanol was removed and the pellet taken up in 60 μl tris-Cl 10 mM. The sample was heated to 85°C for 5 minutes and cooled on ice before addition of 1.5 μl of RNase A, RNase I and RNase T1. RNA digestion occurred for 12 hours at 37°C, 700 rpm shaking before the sample was heated to 85°C again for 5 minutes and cooled on ice. Another 1.5 μl of RNase A, RNase I and RNase T1 was added and the sample digested for another 12 hours.

High pH reversed-phase fractionation occurred under standard settings described above. The initial peak with high adsorption up to approx. 18% B containing RNA contaminations was discarded, the following fractions combined, completely dried by SpeedVac, taken up in 1% formic acid and analyzed by HPLC-MS.

A detailed version of this protocol is included as Methods S1. For updates on downstream applications of XRNAX visit https://www.xrnax.com/applications.

Expression and Purification of C7orf50 and DUF2373 from *E. coli*

Synthetic genes encoding full-length C7orf50 and the C7orf50 domain-of-unknown-function (DUF2373, residues 94-194) with a C-terminal HA-tag were ordered from GenScript. Both genes were codon optimized for expression in *E. coli* and subcloned into the pETM11-Sumo3 vector (EMBL). The expression plasmids were transformed into *E. coli* BL21 (DE3) cells (Novagen). Cells were grown in LB supplemented with 30 μg / ml kanamycin at 25°C until OD600 ~0.6. The temperature was then lowered to 18°C and expression was induced by the addition of 0.5 mM IPTG. The cells were grown further overnight at 18°C and harvested by centrifugation. The cell pellets were resuspended in running buffer (tris-Cl 50 mM pH = 8.0, NaCl 800 mM, imidazole 20 mM and glycerol 10%) supplemented with 1 x protease inhibitors, benzonase and 10 μg / ml lysozyme (Sigma). The cells were lysed via sonication and the cleared lysates were loaded onto a 5 mL Protino Ni-NTA column (Macherey-Nagel). The His6-Sumo3-c7orf50 and His6-Sumo3-c7orf50_94/194 fusion proteins were eluted in running buffer containing 300 mM imidazole. To remove the N-terminal fusion tag, His6-tagged SenP2 protease was added to the elution fractions in a 1:100 ratio. The samples were digested overnight at 4°C while being dialysed to a buffer containing tris-Cl pH = 8.0, NaCl 250 mM, imidazole 20 mM and glycerol 10%. The dialyzed samples were loaded again onto a 5 mL Protino Ni-NTA column and the C7orf50 and C7orf50_94/194 proteins without the N-terminal fusion tag were collected in the flow through. These flow through fractions were then loaded onto a 5 mL HiTrap Heparin HP column (GE Healthcare) equilibrated with tris-Cl 50 mM pH = 8.0, NaCl 250 mM and glycerol 10% in order to remove RNA. Proteins were eluted from the heparin column in a gradient to tris-Cl 50 mM pH = 8.0, NaCl 1.5 M and glycerol 10%. The full length C7orf50 eluted at approximately 800 mM NaCl in the gradient, while C7orf50_94/194 eluted around 500 mM NaCl. After SDS-PAGE analysis, the elution fractions containing full-length C7orf50 or C7orf50_94/194 were pooled and adjusted to 0.5 mg / ml concentration using tris-Cl 50 mM pH = 8.0, glycerol 10% and NaCl 800 or 500 mM, respectively. Proteins were stored at −80°C.

In Vitro Validation of RNA-Binding by Native Agarose Gel Electrophoretic Mobility Shift Assay

Total RNA was extracted from MCF7 cells using the conventional TRIZOL procedure (see above). RNA was diluted to 1000 ng / μl in nuclease-free water, heated to 85°C for two minutes and immediately transferred onto ice before use. An assay dilution of

4 μg RNA / 70 μl assay buffer (tris-Cl 50 mM, NaCl 100 mM) was prepared and stored on ice until use. Dilutions of C7orf50 and DUF2373 (C7orf50_94/194) expressed in *E. coli* (see above) were prepared in 10 μl of their respective storage buffer (tris-Cl 50 mM pH = 8.0, glycerol 10% and NaCl 800 or 500 mM, respectively). Protein dilutions were combined with 70 μl of the RNA assay dilution for a total volume of 80 μl and mixed by pipetting. RNA-binding was allowed to occur for 5 minutes at 37°C, 300 rpm shaking before samples were transferred onto ice again. 20 μl of the samples were combined with 4 μl purple gel loading dye (SDS-free, NEB) and immeadiately run on a 1% agarose gel (TBE) with SYBRSafe staining at 4 W for 45 minutes.

In Vivo Validation for RNA-Binding by PNK Assay or IP-MS

Per construct approximately ten million MCF7 cells were transfected for 24 hours with 15 μg plasmid DNA using Lipofectamine 3000 according to the manufacturer's instructions.

For the T4 polynucleotide kinase (PNK) assay cells were crosslinked as described above and harvested into 2 mL immunoprecipitation (IP) buffer (tris-Cl 50 mM, NaCl 100 mM, $MgCl_2$ 1 mM, $CaCl_2$ 0.1 mM, NP40 1%, SDS 0.1%, sodium deoxycholate 0.5%, 1 x EDTA-free protease inhibitor) by scraping. Of the lysates 1 mL was sonicated with a Sonifier (Branson) and treated with 3 units Turbo DNase and 5 μg RNase A for 15 minutes at 37°C. 15 μl of anti-FLAG M2 beads slurry was used for one hour of IP at 4°C on a rotating wheel, followed by three washes with IP buffer and two washes with PNK buffer (tris-Cl 50 mM, NaCl 50 mM, $MgCl_2$ 10 mM, NP40 0.5%). The PNK labeling reaction was then performed on-bead for 15 minutes at 37°C by addition of 30 μl hot PNK mix (27 μl PNK buffer supplemented with DTT 5 mM, 3 μl T4 PNK, 0.3 μl γ-ATP (0.03 μCi)), followed by four washes with PNK buffer. Complexes were eluted 5 minutes at room temperature into 50 μl elution buffer (glycine 0.1 M, pH = 3). The pH of eluates was neutralized by addition of 7.5 μl tris-Cl 1 M, samples combined with 12.5 μl NuPAGE loading dye supplemented with DTT 200 mM and the sample run on an SDS-PAGE (NuPAGE Bis-Tris, MOPS buffer) at 180 V for 45 minutes. Protein-RNA complexes were blotted onto a nitrocellulose membrane for one hour at 1 mA / cm^2 and exposed a phosphorimaging screen over night before imaging on a Typhoon scanner (GE healthcare). Subsequently, anti-HA Western-blot was performed on the same membranes.

For analysis by MS cells were crosslinked and extracted through XRNAX as described above. The entire yield of the XRNAX extract was dissolved in 200 μl of 50 mM tris-Cl and digested over night at 37°C, 700 rpm shaking, by addition of 2.5 μl RNase I and 2.5 μl RNase A. The digest was combined with 250 μl IP buffer 2 x (tris-Cl 100 mM, NP40 1%, LiCl 300 mM, LiDS 0.2%). 50 μl anti-FLAG M2 beads were washed twice with 1 mL IP buffer 1 x (tris-Cl 50 mM pH = 7.5, NP40 0.5%, LiCl 150 mM, LiDS 0.1%) and reconstituted in 50 μl IP buffer 1 x before addition to the XRNAX digest in 2 mL tubes. IP occurred for 4 hours at room temperature on a rotating wheel. Beads were washed three times with 500 μl IP buffer 1 x, each time resuspending 5 minutes on a rotating wheel. All residual IP buffer was removed and elution occurred twice for 30 minutes at 37°C, 700 rpm shaking into μl 200 elution buffer (Tris-Cl 50 mM, SDS 10%). Eluates were combined and subjected to SP3 protein cleanup as described above with following alterations. No alkylation or reduction was performed, 10 μl SP3 beads were added to the eluates along with 500 μl ACN 100%. Samples were digested in 20 μl TEAB 20 mM and 100 ng trypsin/LysC over night at 37°C, 700 rpm shaking. Beads were captured on a magnet and digests transferred to a fresh tube before addition of formic acid to a final concentration of 1%. Samples were spun down for 5 minutes at 20000 g, supernatants transferred to a fresh tube and analyzed on a QExactive HF MS as described above.

A detailed version of this protocol is included as Methods S1. For updates on downstream applications of XRNAX visit https://www.XRNAX.com/applications.

SILAC-Controlled Discovery of RNA-Binding Proteins From XRNAX Extracts

To maximize coverage of the RNA-bound proteome, we produced XRNAX extracts from half-confluent and confluent cells (~40 million cells per condition), each of which were subjected to silica purification after 15 or 30 minutes of partial tryptic digestion. Cells of one SILAC label were crosslinked with UV-light of 254 nm wavelength as described above, while control cells of the complementary label stayed non-crosslinked. Crosslinked and non-crosslinked cells were combined and extracted by XRNAX as described above.

Per replicate, 930 μg of XRNAX extract (in 930 μl) was further processed. Therefore tris-Cl was added to a final concentration of 50 mM, DTT to 10 mM and SDS to 0.1% before 20 minutes of incubation at 60°C, 700 rpm shaking. CAA was added to a concentration of 20 mM and samples incubated for 20 minutes at room temperature. For predigestion, 100 ng trypsin/lysC was added and samples were pre-digested at 37°C, 700 rpm shaking for 15 or 30 minutes, respectively.

Purification of protein-crosslinked RNA from the digests occurred by silica column purification using the QIAGEN RNeasy Midi Kit with modified protocol (refer to kit manual for buffer descriptors). Predigestion was stopped by combining the sample (approx. 1 ml) with 3.5 mL buffer RLT. The sample was mixed by inversion and heated to 60°C for 15 minutes. 2.5 mL of 100% ethanol was added, the sample mixed by inversion and applied to an RNeasy Midi column by centrifugation with 3000 g for 5 minutes. The flow-through was saved for additional rounds of purification. Washing occurred twice with 2.5 mL buffer RPE. Elution occurred with with 250 μl nuclease-free water. The purification was repeated 3 times, each time using the saved flow through and reusing the same RNeasy Midi column for the individual sample. To the combined eluates (approximately 900 μl total volume) NaCl was added to a final concentration of 300 mM along with 1 μl glycoblue and 1 mL isopropanol. The sample was mixed by inversion and incubated for 1 hour at −20°C. Precipitation occurred by centrifugation with 18000 g at −10°C for 30 minutes. The supernatant was discarded and the pellet washed with 70% ethanol and taken up in 65 μl tris-Cl 50 mM. The sample was heated to 85°C for 5 minutes and cooled

on ice before 2.5 μl of RNase A, RNase I and RNase T1 was added. RNA digestion occurred over night at 37°C, 700 rpm shaking. 500 ng trypsin/LysC were added for 16 hours digestion at 37°C, 700 rpm shaking.

High pH reversed-phase fractionation occurred under standard settings described above. The initial fractions up to approx. 18% B containing RNA contaminations and few peptides were discarded, the following eluate was collected in six consecutive fractions, which were subsequently dried by SpeedVac and taken up in 1% formic acid before analysis by HPLC-MS.

A detailed version of this protocol is included as Methods S1. For updates on downstream applications of XRNAX visit https://www.xrnax.com/applications.

Quantification of Nascent Protein Upon Arsenite Stress Using Azidohomoalanine-Labeling

MCF7 cells with heavy and light SILAC labels were expanded for three days on 15 cm dishes until 80% confluent. Cells of one SILAC label were treated with 400 μM sodium arsenite for 5, 10, 20, 30 or 60 minutes, while cells of the complementary label were left untreated. Azidohomoalanine-labeling (AHA-labeling) and protein purification using click-chemistry for MS quantification was performed as described before (Eichelbaum et al., 2012). In brief, cells were deprived of methionine using methionine-free media for 30 minutes. Labeling started with the addition of AHA-containing media, which occurred simultaneously with the addition of arsenite. After labeling, cells were immediately transferred onto ice, washed with ice-cold PBS and harvested by scraping. Protein enrichment was performed using the Click-iT Protein Enrichment Kit (Thermo Scientific) using the manufacturer's instructions. Protein captured on agarose beads was subjected to tryptic digestion using 500 ng trypsin/LysC in 200 μl TEAB 50 mM and peptides were cleaned up using an Oasis PRiME HKB μElution Plate (Waters). HPLC-MS detection occurred on an Orbitrap Fusion MS using the parameters described above.

Differential Quantification of RNA-Binding Upon Arsenite Stress

For the differential quantification of RNA-binding upon arsenite stress, MCF7 cells were expanded for three days to approx. 70% confluence. 30 million cells of one SILAC-label were exposed to 100 μM sodium arsenite for 0, 5, 10, 20, or 30 minutes, while control cells of the complementary label remained untreated. Duplicate experiments were performed for each time point, which included SILAC-label swap. Both treated and control cells were UV-crosslinked, combined and subjected to XRNAX and silica-enrichment as described above with the only difference that the predigestion time was kept constant at 30 minutes for all samples. Samples were high pH reversed-phase fractionated into 8 fractions as described above.

Total proteomes where analyzed from cells treated with arsenite at the identical time points, in duplicates and with SILAC label-swap. After treatment the media was discarded and cells were immediately put on ice and washed with ice-cold PBS. Cells were harvested by scraping and subjected to the standard proteomic workflow described above before fractionation into 8 fractions at high pH. Importantly, for the quantification only peptides were used, which previously had been discovered as super-enriched during the generation of the ihRBP. Refer to the 'QUANTIFICATION AND STATISTICAL ANALYSIS' section below for further elaboration.

A detailed version of this protocol is included as Methods S1. For updates on downstream applications of XRNAX visit https://www.xrnax.com/applications.

Total Proteome Analysis of Arsenite-Induced Protein Degradation

For total proteome analysis of MCF7 under controlled cell culture conditions 0.5×10^6 cells were seeded in 10 cm dishes and cultured for 3 days. Inhibition of autophagy through 10 μM spautin-1 was induced 24 hours prior to arsenite stress. Sodium arsenite was applied at 400 μM concentration. Cells were harvested and subjected to the standard proteomic workflow described above.

We note here that the degree to which autophagic degradation was induced heavily depended on the growth state that the cells were in: During the initial lag-phase of cell culture one day after seeding, arsenite-induced protein degradation was minor compared to the effect observed in the following days of culture (data not shown). For cells seeded at a density to reach confluence after 5 days of culture the most pronounced effect was observed after 3 days.

Ethynyl-Uridine Incorporation and Confocal Microscopy for Monitoring RNA Turnover Upon Arsenite Stress

For the visualization of nascent transcripts using ethynyl-uridine (EU), MCF7 cells were grown on glass coverslips for 3 days. EU was applied at 1 mM, sodium arsenite at 400 μM and the RNA polymerase I inhibitor CX5461 at 10 μM concentration. Treatment occurred for 30 or 60 minutes, the media was discarded and cells washed once with PBS. Fixation occurred with 3% paraformaldehyde in PBS at room temperature for 10 minutes subsequent to washing with PBS. Cells were permeabilized using 0.5% Triton X-100 in PBS for 15 minutes at room temperature and washed again with PBS. The copper-catalyzed click reaction occurred in 100 mM HEPES pH = 8, 150 mM NaCl, 5 mM sodium ascorbate, 100 μM $CuSO_4$, 500 μM THPTA and 20 μM sulfo-Cy5-azide for 30 minutes at room temperature. This and all following steps occurred under protection from light. The reaction solution was discarded and slides washed once with TBST (50 mM tris-Cl, 150 mM NaCl, 0.1% Tween-20). HOECHST33342 was applied at a concentration of 10 μM in TBSP for 10 minutes, slides again washed twice with TBST and mounted using ProlongGold antifade mountant.

Imaging was performed on a Leica SP5 (Leica) using a 63x oil emersion objective. Detection of sulfo-Cy5-EU occurred using the default Leica Cy5 filter settings with excitation at 633 nm and detection at 650-750 nm wavelength.

Immunofluorescence Staining and Confocal Microscopy for Monitoring EXOSC2 Localization Upon Arsenite Stress

For the visualization of EXOSC2, MCF7 cells were grown on glass coverslips for 3 days. Sodium arsenite was applied at 400 μM for 0, 5 or 30 minutes, the media was discarded and cells washed once with PBS. Fixation occurred with 3% paraformaldehyde in PBS at room temperature for 10 minutes subsequent to washing with PBS. Cells were permeabilized using 0.5% Triton X-100 in PBS for 15 minutes at room temperature and washed again with PBS. The antibody was diluted 1:100 in TBST (50 mM tris-Cl, 150 mM NaCl, 0.1% Tween-20) and binding allowed to occur over night at 4°C. Slides were washed twice with TBST before the secondary antibody (goat anti-mouse Cy5) was applied at a 1:500 dilution in TBST for 2 hours. Slides were stained with HOECHST33342 and imaged as described above.

XRNAX CLIP-seq for the Detection of EXOSC2 RNA Targets

For crosslinking and immunoprecipitation followed by sequencing (CLIP-seq) from XRNAX extracts we first validated antibodies for normal IP by MS using the IP buffer 1 x (tris-Cl 50 mM pH = 7.5, 0.5% NP40, 150 mM LiCl, 0.1% LiDS) later used for the CLIP-seq experiment.

For each sample XRNAX extracts were prepared from MCF7 cells as described above. For RNA fragmentation 10 mM tris-Cl and 5 mM EDTA were added to 100 μg XRNAX extract, which were sonicated in microTUBEs with AFA fiber (Covaris, 520045) using a S220 focused-ultrasonicator (Covaris) with the settings: 900 s, peak power 175, duty factor 50, cycles / burst 200 and average power 87.5.

For the size-matched input control (SMI-control) 5 μg of the sonicated XRNAX extract (approx. 2 μl) were mixed with 33 μL MilliQ water, 5 μL FastAP buffer 10 x and 10 μL FastAP. Dephosphorylation occurred for 15 minutes at 37°C, then FastAP was inactivated for 5 minutes at 80°C and the sample transferred to ice. 5 μL PNK buffer 10 x, 10 μL ATP 10 mM, 25 μL MilliQ and 10 PNK were added and incubated another 15 minutes at 37°C. 15 μL of the SMI-control (approx. 200 ng RNA) was combined with 5 μL SDS-loading dye 5 x (tris-Cl pH = 6.8 250 mM, SDS 10%, 0.02% bromphenol blue, glycerol 30%) and 5 μL DTT 1 M. Samples were heated to 70°C for 15 minutes before they were run on an SDS-PAGE along with the IP samples.

For the IP 100 μg sonicated XRNAX extract in approx. 125 μl was combined with 125 μl IP buffer 2 x (tris-Cl 100 mM pH = 7.5, 1% NP40, 300 mM LiCl, 0.2% LiDS). 1 μg antibody was added for 4 hours at 4°C on a rotating wheel before antibody capture with 100 μl protein G beads overnight. The beads were collected on a magnetic stand and the supernatant discarded. The beads were washed three times with 1 mL IP buffer, each time carefully turning the tube upside down until the beads were completely resuspended. Subsequently, beads were washed twice with 1 mL TBST while on the magnet. For end-repair the beads were resuspended in 100 μL dephosphorylation mix (80 μL MilliQ, 10 μL FastAP buffer 10 x, 8 μL FastAP, 2 μL RNASin) and incubated for 15 minutes at 37°C, 1000 rpm shaking. Beads were collected on a magnetic stand, the supernatant discarded and the beads washed twice with 1 mL TBST while on the magnet. Subsequently, the beads were resuspended in 100 μL PNK mix (70 μL MilliQ water, 10 μL ATP 10 mM, 10 μL PNK buffer A 10 x, 8 μL PNK, 2 μL RNASin) and incubated for another 15 minutes at 37°C, 1000 rpm shaking. Beads were collected on a magnetic stand and the supernatant discarded. Protein-RNA complexes were eluted into 5 μL SDS loading dye 5 x (NuPAGE), 5 μL DTT 1 M and 15 μL MilliQ for 15 minutes at 70°C. Beads were collected on a magnetic stand and the IP sample transferred to a fresh tube.

IP and SMI control were run alongside on a 4%–12% SDS-PAGE (NuPAGE BisTris, MES buffer) and blotted onto nitrocellulose with 500 mA for one hour at 4°C. The area corresponding to the molecular weight of the protein of interest plus 75 kDA were excised, cut into pieces and transferred to a fresh tube. RNA was released by digestion with 50 μl proteinase K in 200 μl proteinase K buffer (tris-Cl 50 mM, EDTA 10 mM, NaCl 150 mM, SDS 1%) at 55°C for 30 minutes. 250 μl PCI for RNA was added, the sample mixed by inversion, incubated 10 minutes on ice and spun down 10 minutes with 12000 g at 4°C. 200 μl of the aqueous phase were transferred to a fresh tube, NaCl added to a final concentration of 300 mM, combined with 1 μl GlycoBlue and 200 μl isopropanol. Samples were mixed and precipitated for 2 hours at −20°C before centrifugation with 18000 g at −10°C for 1 hour. Pellets were washed with 80% ethanol and resuspended in nuclease-free water.

RNA produced by this protocol was approx. 30-80 nt in size, carried a 5′ phosphate and a 3′ hydroxyl. For generation of sequencing libraries, we used the NextFlex Small RNA 3.0 kit and gel-based size-selection of RNA fragments from 30-50 nt according to the manufacturer's instructions. Samples were barcoded so that twelve samples could be run in one lane on a HiSeq2000 (Illumina).

A detailed version of this protocol is included as Methods S1. For updates on downstream applications of XRNAX visit https://www.xrnax.com/applications.

QUANTIFICATION AND STATISTICAL ANALYSIS

MS Database Search

All MS raw files were searched using MaxQuant, except for data of nucleotide-crosslinked peptides. The database searched was the reviewed UniProt human proteome (search term: 'reviewed:yes AND proteome:up000005640', 20216 entries, retrieved 11 September 2017) and the default Andromeda list of contaminants. All settings were used at their default value, except for specifying SILAC configurations and indicating the appropriate number of fractions per sample. For the differential quantification of RNA-binding during arsenite stress the match-between-runs option was activated, for all other searches this was explicitly not the case.

MS data of nucleotide-crosslinked peptides was searched with MSFragger using the same UniProt database as mentioned above. Precursor mass tolerance was set to 1000 Da and the export format set to tsv, otherwise all settings were used at their default value.

Processing and Analysis of MS Data

Cyclic-U crosslinked peptides were selected from peptide-spectrum matches (PSMs) carrying a modification within 306-307 Da, as detected by MSFragger.

For the analysis of RNA-binding proteins from XRNAX extracts, the MaxQuant peptides.txt table was filtered to remove entries in 'potential contaminants' and 'reverse'. Furthermore only peptides that matched the category 'Unique Groups' were used. To derive RNA-binding proteins for the individual cell lines, peptides from the four replicates were combined and filtered with the condition (SILAC intensity crosslinked +1) / (SILAC intensity non-crosslinked +1) > 1000. Pseudo-counts were added to include peptides where the non-crosslinked channel had zero intensity. Proteins ('Leading razor protein') identified with two or more unique peptides were included in the RNA-binding proteome of the individual cell line.

K-mer frequencies in proteins of the ihRBP were determined using the R package biostrings. Global protein sequence features were computed using the R package 'peptides' with the scales 'Kyte-Doolittle' for hydrophobicity, 'EMBOSS' for isoelectric point and 'EMBOSS' for charge.

For differential quantification of RNA-binding upon arsenite stress, the MaxQuant peptides.txt table was filtered to remove entries in 'potential contaminants' and 'reverse'. Furthermore, only peptides that matched the category 'Unique Groups' and which occurred in the list of ihRBP super-enriched peptides were used. The latter was essential for proper quantification, as demonstrated in Figure S1G. Note that a protein can contribute super-enriched and background proteins, so that combining the SILAC ratios of all peptides, the ones with the highest intensity, or similar approaches in a differential XRNAX experiment, will lead to wrong quantification. As an example, consider a protein with one super-enriched peptide and one background peptide in a differential XRNAX experiment with identical cell numbers for each of the two SILAC channels. If the SILAC ratio of the super-enriched peptide is 0.1 but the background peptide 0.9, the median of the two will be 0.5. As the super-enriched peptide reflects the true change, its integration with a background peptide will falsely adjust the outcome by 500% (i.e., $0.1 \times 5 = 0.5$). A list of super-enriched peptides collated from the cell lines MCF7, HEK293 and HeLa during generation of the ihRBP can be downloaded from https://www.xrnax.com/applications. Subsequently to filtering, the median of normalized SILAC Ratios was computed in order to integrate peptide ratios to individual proteins ('Leading razor protein'). Here, a minimum of two super-enriched peptides quantified over all time points was required.

Control total proteome data for the experiment was analyzed identically, except for filtering for ihRBP super-enriched peptides, which was omitted. The combined data presented in this manuscript was the mean of biological replicates for each time point filtered for a variance of 15% or smaller.

For the analysis of the arsenite-induced degradation normalized ratios from the MaxQuant proteinGroups.txt table were used. Dose-dependence was analyzed using the R package 'drc'.

Gene-Ontology (GO) Enrichment Analysis

GO enrichment analysis was performed using the GOrilla web interface and uniprot identifier as input. Type (ranked or background controlled) of the enrichment and background control (if applicable) are indicated in the text.

Visualization of the Human 80S Ribosome

Nucleotide-crosslinked peptides from ribosomal proteins and ribosomal proteins affected in their RNA-binding upon arsenite stress were located in the cryo-EM structure of the human ribosome (Khatter et al., 2015) (PDB 4UG0) using UCSF Chimera.

Processing and Analysis of RNA Sequencing Data

For the estimation of the rRNA content of libraries, which had not been RiboZero depleted, reads were aligned to a collection of human ribosomal sequences of the hg19 assembly retrieved using the UCSC table browser. The table used was 'rmsk' and filtering was applied so that 'repClass does match rRNA'. All reads were aligned to those sequences using bowtie2 and reads that failed to align were written to a new file. Reads in this file were aligned to the complete hg19 assembly. Percentages of the rRNA content were estimated by comparing the number of reads aligning to hg19 rRNA sequences and residual reads aligning to the complete hg19 assembly.

For estimating the content of RNA biotypes in RiboZero depleted libraries, reads were aligned to the hg19 assembly using bowtie2. Subsequently, counting was performed with HTSeq-count using the geneset annotated by GENCODE 19 (release 12.2013) and using the GTF feature 'gene' for counting.

Processing and Analysis of XRNAX-CLIP-seq Data

Unique molecular identifiers (UMIs) were extracted using Je. Exact PCR duplicates were removed using BBMap. Adapters were trimmed using cutadapt and processed reads sequentially aligned to the 45S pre-ribosomal RNA (NR_046235.3), Repbase (Bao et al., 2015) and the hg38 genome with STAR, reusing reads which escaped previous alignment efforts. Non-exact PCR duplicates were removed with Je. The csaw library was used to calculate 20 nt coverage windows for the 45S rRNA for each individual time

point. Subsequently, the DESeq2 library was used to calculate fold changes from this coverage between duplicates of the IP and duplicates of the SMI control. P values were corrected for multiple testing with Benjamini-Hochberg.

Statistical Analysis and Data Visualization

All data handling apart from what is mentioned above was performed in RStudio and visualized using the ggplot2 library. Microscopy, western blot and phosphor images were processed in Photoshop (CC, Adobe 2015) and figures arranged in Illustrator (CC, Adobe 2015).

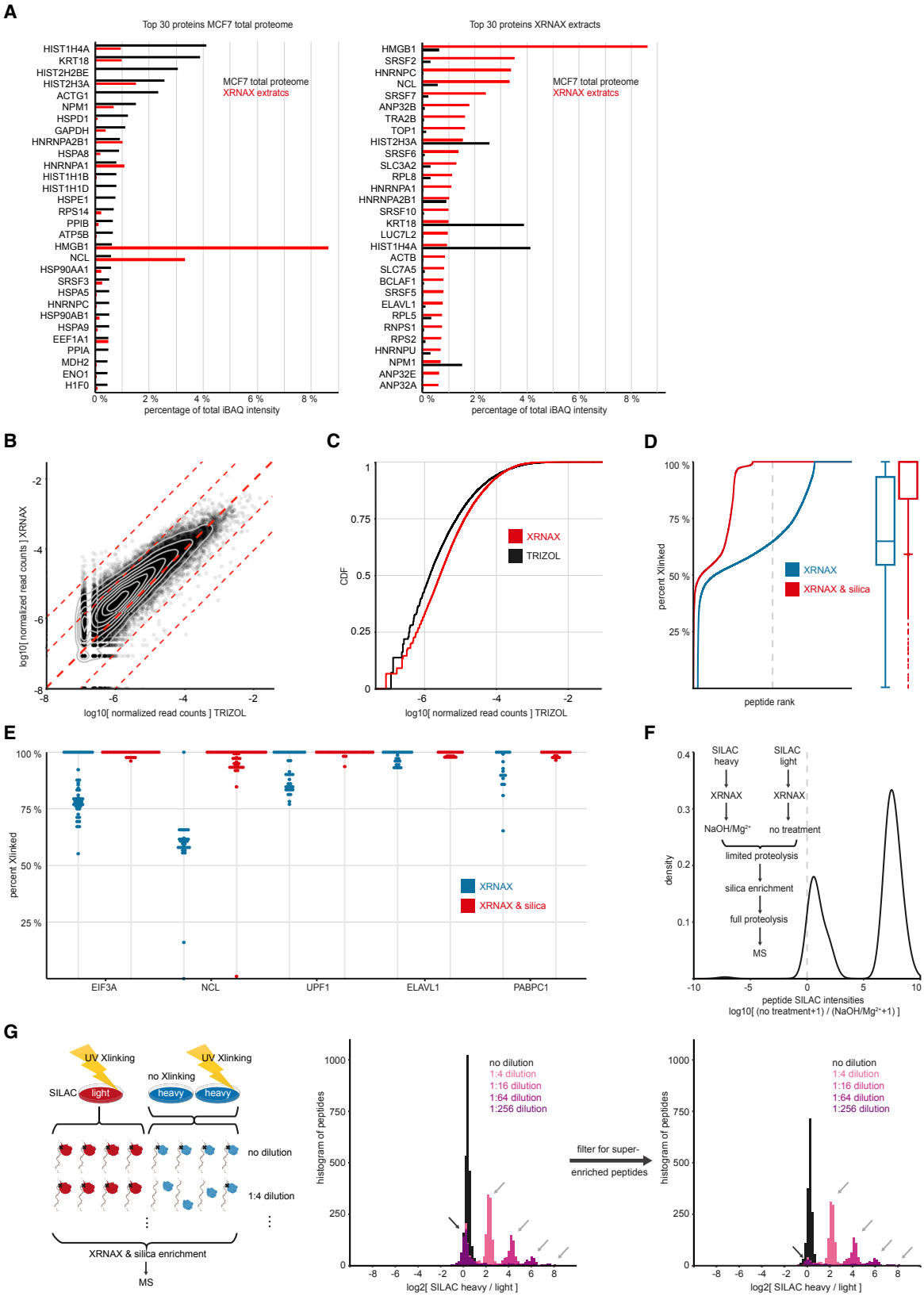
DATA AND SOFTWARE AVAILABILITY

Mass spectrometry data has been deposited to the ProteomeXchange Consortium under the PRIDE identifier PXD010520. RNA sequencing data has been deposited to the EMBL-EBI ENA server under the accessions PRJEB26441 and PRJEB26442.

ADDITIONAL RESOURCES

Detailed, photo-documented protocols for XRNAX and its downstream applications are available under https://www.xrnax.com.

Supplemental Figures



(legend on next page)

Figure S1. Proteomic and Transcriptomic Features of XRNAX-Extracts, Related to Figure 1

(A) Relative abundance of proteins in the total proteome and XRNAX-extracts of MCF7 cells, estimated from iBAQ values. Bar graphs display contribution of the top-30 proteins to the combined iBAQ intensity of all detected proteins.

(B) Scatterplot comparing normalized read counts for all GENCODE-annotated transcripts in RNA obtained by XRNAX and TRIZOL. MCF7 cells were exposed to 4SU for 16 hours before UV-crosslinking at 365 nm and processing via XRNAX, or without crosslinking and conventional TRIZOL extraction. Reads were counted per gene and normalized to the total number of counts. Each point represents one gene and displays the mean of two replicates. Contour lines indicate highest density of the points in the plot. Dashed lines indicate fold-changes of 1, 10 and 100. Sequencing for all replicates was performed in one lane and read count for all libraries was within 10% deviation from the average read-count.

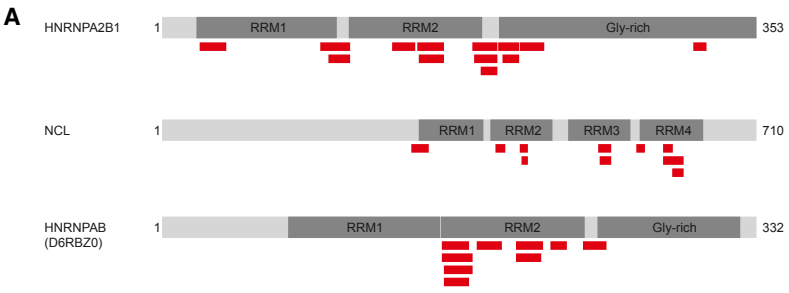
(C) Comparison of normalized read counts for all GENCODE genes between XRNAX and TRIZOL-extracted RNA. Same data as in B shown as cumulative distribution.

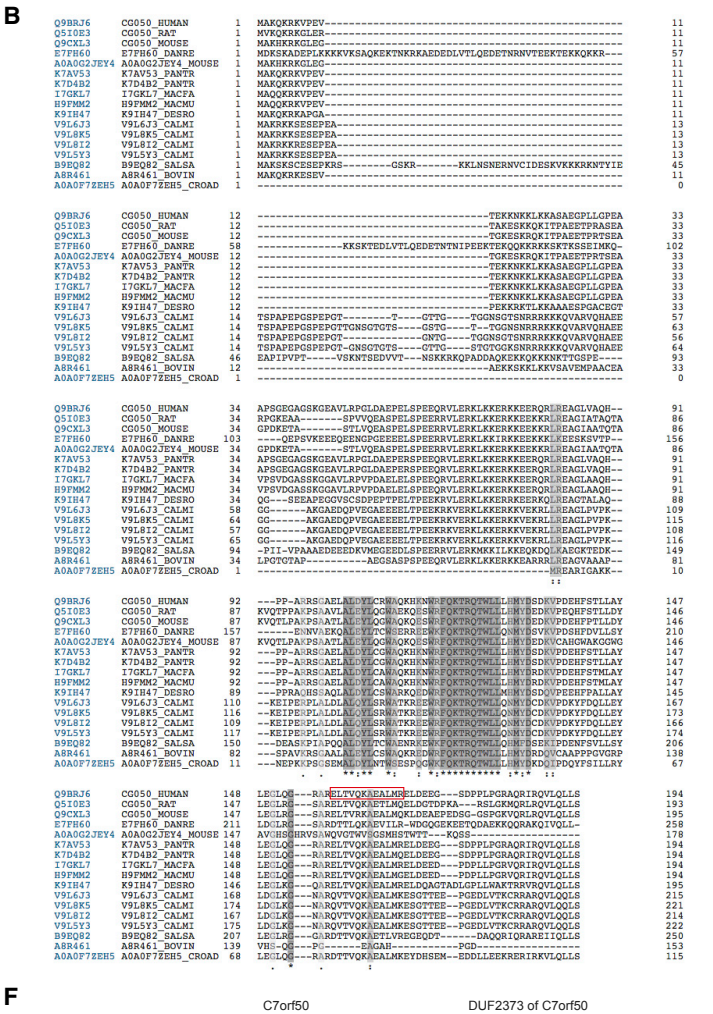
(D) Ranked scatterplot (left) and boxplot (right) for RNA-crosslinked protein in XRNAX extracts before and after silica enrichment. Percentages of crosslinking were computed using peptide SILAC intensities and the formula percent Xlinked = (Xlink/control) / (Xlink/control +1).

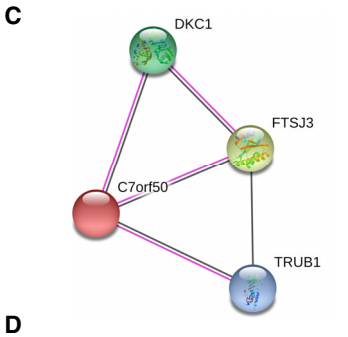
(E) Dotplot showing the enrichment of five well-studied RNA-binding proteins through XRNAX. Each dot represents one peptide of the indicated protein and how much of it occurs in the crosslinked SILAC channel (using the formula from D). Blue coloring shows this percentage after XRNAX only, whereas red coloring after XRNAX and subsequent silica enrichment.

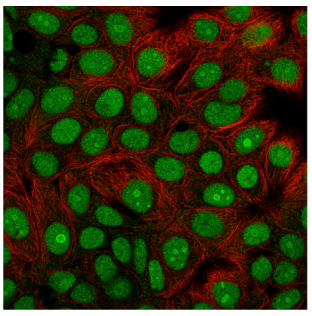
(F) Density plot showing the enrichment of peptides from XRNAX-extracts with intact RNA over XRNAX-extracts where RNA was degraded. For details see text.

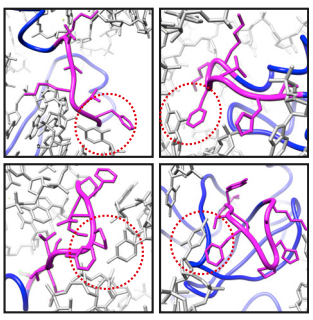
(G) Proof-of-concept for the differential quantification of RNA-binding using XRNAX and silica enrichment. Heavy SILAC-labeled MCF7 cells were UV-crosslinked and mixed with non-crosslinked heavy MCF7 cells in 5 defined ratios. These mixtures of heavy cells were combined with the identical amount of UV-crosslinked, light cells and subjected to XRNAX followed by silica enrichment and MS quantification. Histogram displays SILAC ratios without normalization. Peptides that were found super-enriched in previous experiments using a non-UV-crosslinked control (Figure 1F) showed discrete fold-changes corresponding to mixing ratios (gray arrows), whereas peptides that were not super-enriched before showed a 1:1 ratio (black arrow).

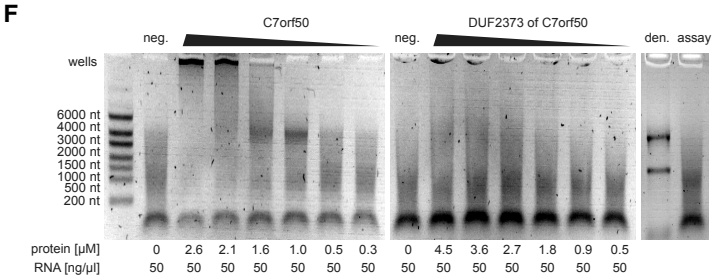












(legend on next page)

Figure S2. Sequence Analysis of C7orf50 and the Putative RNA-Binding Domain DUF2373, Related to Figure 2

(A) Proteins with the largest number of identified cyclic U-crosslinked peptides, indicating their annotated domain structure (dark gray) and the position of the detected cyclic U-containing peptides.

(B) Sequence alignment for 17 high-confidence hits from a HMMER search for human C7orf50. The identified cyclic U-crosslinked peptide, ELTVQKAEALMR, is highlighted in red.

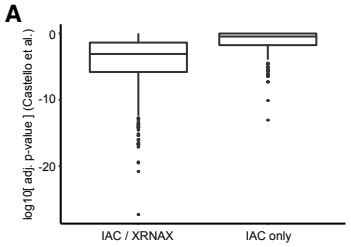
(C) STRING interaction network for C7orf50. DKC1 and TRUB1 are involved in pseudourylation of rRNA, FTSJ3 in methylation of rRNA.

(D) Immunofluorescence image of C7orf50 in MCF7 cells. Red: microtubules; green: C7orf50. Image courtesy: http://www.proteinatlas.org.

(E) Protein-RNA interactions in crystal structure of the human ribosome for four exemplary proteins. The amino-acid backbone is indicated in blue, peptides found crosslinked to cyclic U are indicated in magenta and RNA is indicated in gray. Red circles highlight phenylalanine in the vicinity of a uracil base. Upper left: GFVKVVK in RPL5, upper right: IHGVGFKK in RPL31, lower left: MKFNPFVTSDR in RPL26L1, lower right: VAYVSFGPHAGK in RPL14.

(F) Native agarose gel electrophoretic mobility shift assay (EMSA) for C7orf50 and its DUF2373. Proteins were expressed and purified from *E. coli*. Increasing amounts of protein were incubated with constant amounts of MCF7 total RNA and run on an agarose gel without prior denaturation. Left: Addition of denaturing running buffer and heating (den.) demonstrates that RNA under assay conditions (assay) is intact.

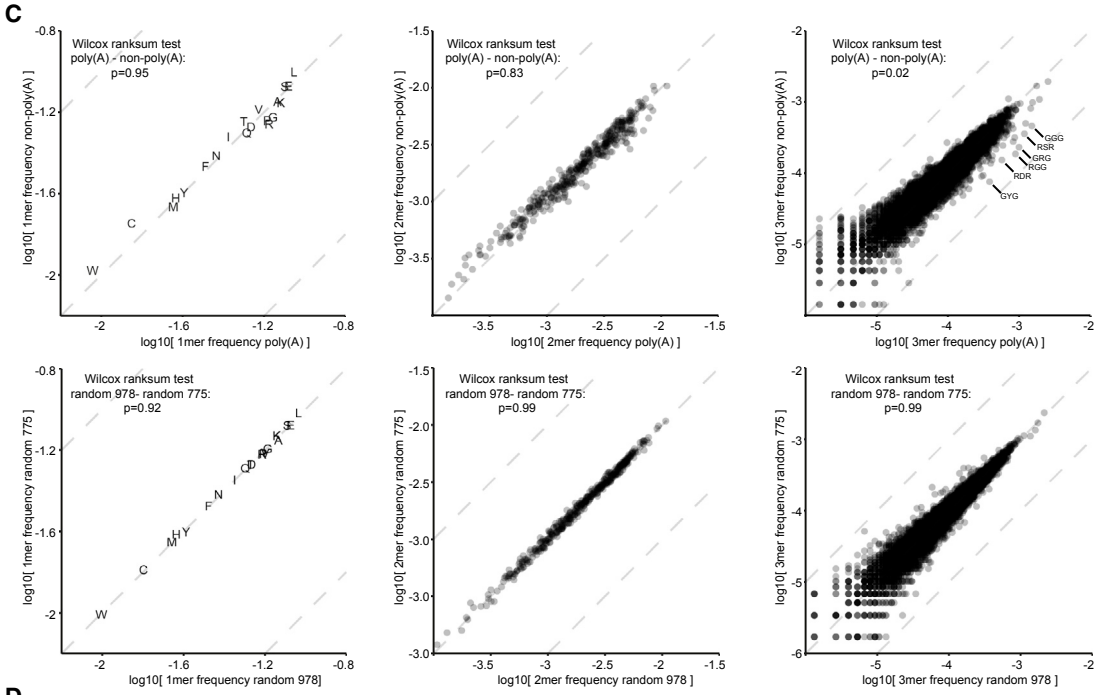
A



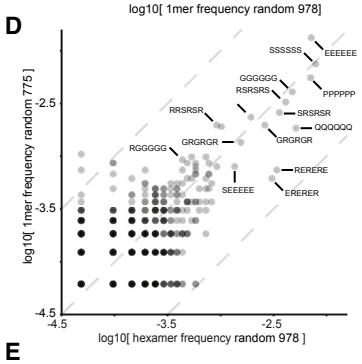
B

GO term	poly(A) interactome			non-poly(A) interactome		
	MCF7	HEK293	HeLa	MCF7	HEK293	HeLa
RNA binding	1E-Inf	1E-Inf	1E-Inf	1.22E-135	4.94E-96	1.06E-88
structural constituent of ribosome	4.55E-49	3.03E-48	3.92E-39	3.63E-27	8.55E-26	2.14E-27
ribosome binding	1.20E-05	3.84E-06	5.19E-06	4.98E-09	1.59E-10	2.29E-06
translation factor activity, RNA binding	2.35E-10	2.15E-11	2.54E-09	1.58E-08	1.05E-06	8.14E-05
snoRNA binding	9.70E-05	1.74E-07	1.31E-11	1.80E-08	5.11E-07	NS
7S RNA binding	NS	NS	2.44E-04	1.37E-06	1.73E-05	NS
ribonuclease activity	NS	NS	NS	5.22E-06	NS	8.56E-06
snRNA binding	1.07E-10	3.16E-10	4.65E-06	1.19E-05	3.81E-05	9.37E-09
aminoacyl-tRNA ligase activity	NS	NS	NS	1.63E-05	6.87E-06	3.03E-08
mRNA binding	8.89E-49	3.22E-58	7.42E-53	2.13E-05	2.33E-07	6.06E-08

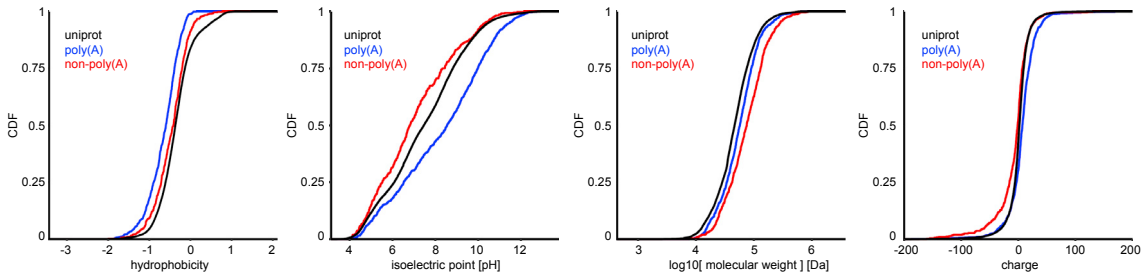
C



D



E



Wilcox ranksum test:
uniprot - poly(A): p<2.2E-16
uniprot - non-poly(A): p=1.2e-9
poly(A) - non-poly(A): p<2.2E-16

Wilcox ranksum test:
uniprot - poly(A): p<2.2E-16
uniprot - non-poly(A): p=1.3e-9
poly(A) - non-poly(A): p<2.2E-16

Wilcox ranksum test:
uniprot - poly(A): p=5.4e-11
uniprot - non-poly(A): p<2.2e-16
poly(A) - non-poly(A): p=3.9e-9

Wilcox ranksum test:
uniprot - poly(A): p< 2.2e-16
uniprot - non-poly(A): p=4.7e-14
poly(A) - non-poly(A): p< 2.2e-16

(legend on next page)

Figure S3. Properties of XRNAX-Derived RNA-Binding Proteomes, Related to Figure 3

(A) Boxplot comparing adjusted p values as reported by Castello et al. (2012) for proteins discovered through both interactome capture (IAC) and XRNAX, or interactome capture only.

(B) GO-enrichment analysis (molecular function) for XRNAX-derived RNA-binding proteomes from three cell lines. Displayed are adjusted p values for ten RNA-related terms with especially strong enrichment in either group. Note that enrichment analysis was performed against two different background sets. The poly(A) interactome was analyzed using the combined proteome of MCF7, HeLa and HEK293 cells reported by Geiger et al. as background control. A similar analysis was problematic for the non-poly(A) interactome because of the high prevalence of RNA-binding proteins in the total proteome of the three cell lines, annotated as such in reference to the two initial poly(A) interactome studies by Baltz et al. (2012) and Castello et al. (2012). Since for this particular analysis we were exclusively interested in non-poly(A) interactors, we removed the published poly(A) interactome from the background control. Inf: infinity, NS: not significant with a p value < 10E3.

(C) Scatterplots comparing amino acid, dipeptide and tripeptide (k-mer) frequenies between poly(A) and non-poly(A) interactomes in the ihRBP (see Figure 3E). Top: All possible permutations of the 20 amino acids for each k-mer were counted in proteins of each group and normalized to the total number of counts. Bottom: Control analysis for two groups of the same size but containing randomly selected proteins from the ihRBP.

(D) Control analysis referring to Figure 3F. Scatterplot comparing hexamer frequencies for control groups with randomly selected proteins as described in panel C.

(E) Cumulative distributions of physicochemical properties in the poly(A) and non-poly(A) interactomes (see Figure 3E) and the entire UniProt human proteome (uniprot).

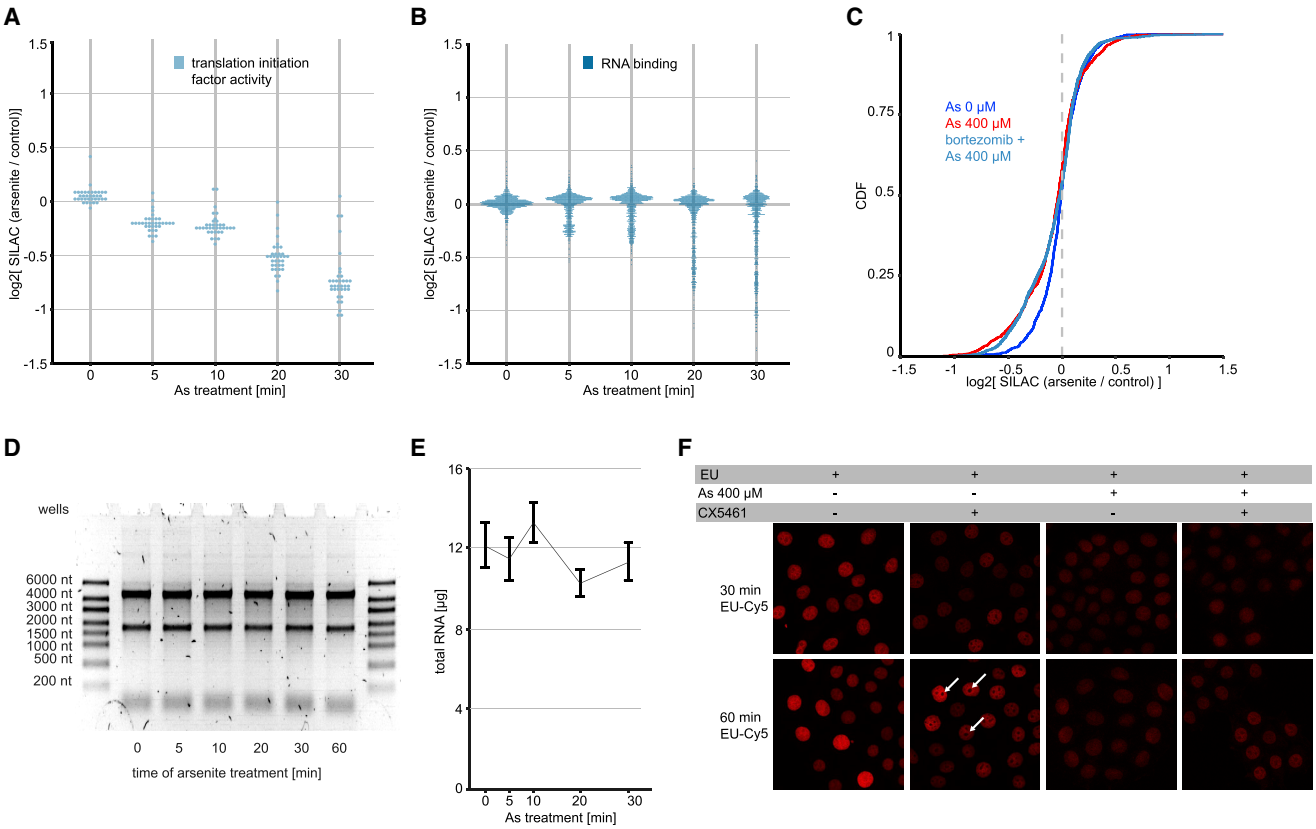


Figure S4. Changes in the Total Proteome upon Arsenite Stress, Related to Figure 4

(A) Dotplots displaying changes in total proteome of MCF7 cells upon arsenite stress for proteins with the GO term 'Translation initiation factor activity'. Each dot represents one protein. Values displayed are means of duplicate experiments with label swap filtered for a variance of 15% or smaller.

(B) Same as panel A for proteins with the GO term 'RNA binding'.

(C) Cumulative distribution of changes in total proteome of MCF7 cells upon arsenite treatment and proteasome inhibition by bortezomib. Cells were treated for 1 hour with 500 nM bortezomib (green) or left untreated (red) before heavy SILAC labeled cells were treated with 400 µM arsenite for 30 minutes.

(D) Agarose gel electrophoresis of total RNA extracted from MCF7 cells upon 400 µM arsenite stress for indicated time.

(E) Timeline showing yield of total RNA extracted from MCF7 cells after arsenite stress. Identical number of cells were treated with 400 µM arsenite for indicated time, total RNA was extracted and quantified using UV-spectrosopy. N = 6, error bars indicate standard error of the mean (SEM).

(F) Confocal microscopy of MCF7 cells incorporating EU upon arsenite stress. To exclude the possibility that RNA turnover may be increased upon arsenite stress, ethenyl-uridine (EU) labeling was applied to visualize newly synthesized RNA using click-chemistry. After applying the RNA-polymerase I inhibitor CX5461 discrete areas (white arrows), presumably representing nucleoli, were not stained anymore demonstrating specific labeling of nascent transcripts. Under arsenite stress incorporation of EU was lower, indicating reduced transcription. Exchange of cytosolic RNA with nascent RNA was not apparent during the chosen time window.

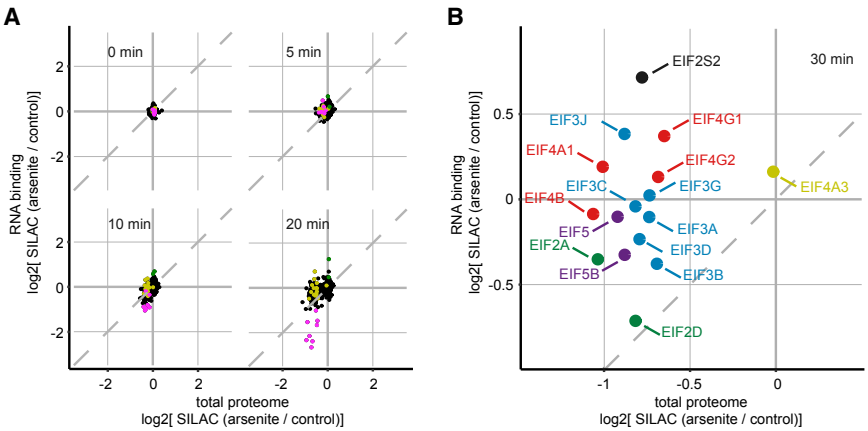


Figure S5. Quantification of Changes in RNA-Binding Proteins, Related to Figure 5

(A) Scatterplots comparing changes in the total proteome to changes in RNA-binding after 0, 5, 10 and 20 minutes of arsenite stress. See also Figure 5.

(B) Scatterplot comparing RNA-binding versus total proteome changes after 30 minutes of arsenite treatment. Magnification of data in Figure 5E, only showing eukaryotic translation initiation factors (EIFs). Color coding refers to complexes these proteins are known to be part of. Black: EIF2 complex, blue: EIF3 complex, red: EIF4 complex, magenta: EIF5 complex, yellow: exon junction complex, green: auxiliary EIFs without membership in any core complex.

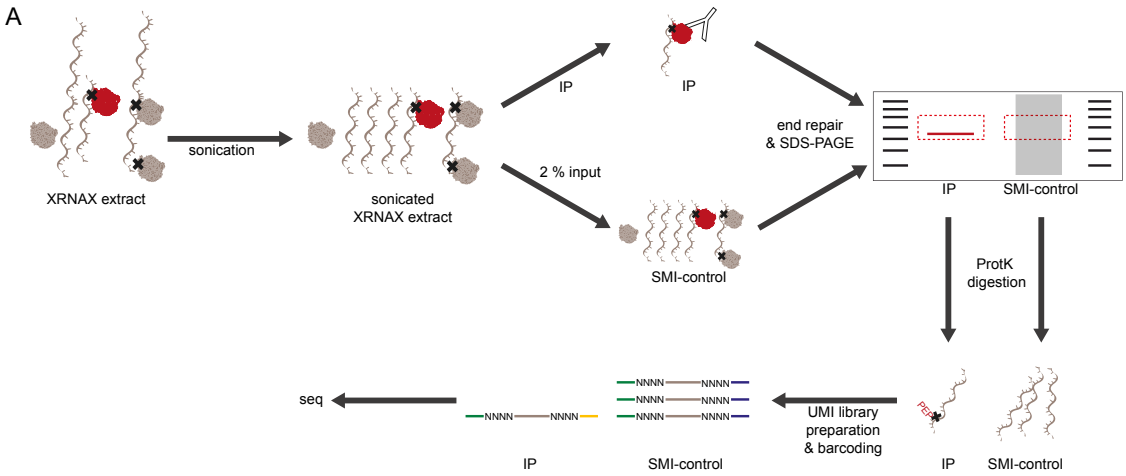


Figure S6. Combination of XRNAX and CLIP-Seq, Related to Figure 6

Experimental scheme for an XRNAX CLIP-Seq experiment. After initial fragmentation of RNA through ultrasonication, IP against a protein of interest co-precipitates crosslinked RNA fragments. Both input and IP are resolved on an SDS-PAGE, blotted, and membrane pieces cut out in a region corresponding to the adequate molecular mass. RNA is released by proteinase K digestion and prepared into a sequencing library using conventional small RNA library preparation with unique molecular identifiers (UMIs). For further details refer to STAR Method section.