Purification of Cross-linked RNA-Protein Complexes by Phenol-Toluol Extraction

Erika C Urdaneta[1], Carlos H Vieira-Vieira[2], Timon Hick[1], Hans-Herrmann Wessels[34], Davide Figini[1], Rebecca Moschall[5], Jan Medenbach[5], Uwe Ohler[34], Sander Granneman[6] Matthias Selbach[2], and Benedikt M Beckmann*[17]

Abstract

Recent methodological advances allowed the identification of an increasing number of RNA-binding proteins (RBPs) and their RNA-binding sites. Most of those methods rely, however, on capturing proteins associated to polyadenylated RNAs which neglects RBPs bound to non-adenylate RNA classes (tRNA, rRNA, pre-mRNA) as well as the vast majority of species that lack poly-A tails in their mRNAs (including all archea and bacteria). To overcome these limitations, we have developed a novel protocol, Phenol Toluol extraction (PTex), that does not rely on a specific RNA sequence or motif for isolation of cross-linked ribonucleoproteins (RNPs), but rather purifies them based entirely on their physicochemical properties. PTex captures RBPs that bind to RNA as short as 30 nt, RNPs directly from animal tissue and can be used to simplify complex workflows such as PAR-CLIP. Finally, we provide a first global RNA-bound proteome of human HEK293 cells and *Salmonella* Typhimurium as a bacterial species.

[1]IRI Life Sciences, Humboldt University, 10115 Berlin, Germany

[2]Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany

[3]Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine, Berlin, Germany

[4]Department of Biology, Humboldt University, 10115 Berlin, Germany

[5]Biochemistry I, University of Regensburg, Universitätsstrasse 31, 93053 Regensburg, Germany

[6]Centre for Systems and Synthetic Biology (SynthSys), University of Edinburgh, Max Born Crescent, EH9 3BF, Edinburgh, UK

[7]Correspondence: benedikt.beckmann@iri-lifesciences.de

Introduction

RNA binding proteins are key factors in the post transcriptional regulation of gene expression. Spurred by recent technological advances such as RNA interactome capture (Castello et al., 2012; Baltz et al., 2012; Beckmann et al., 2015), the number of RBPs has been greatly increased. A powerful tool to study ribonucleoproteins (RNPs) is UV cross-linking: irradiation of cells with short wavelength UV light results in covalent cross-links of proteins in direct contact with the RNA (Fig. 1a) (Favre et al., 1986; Hockensmith et al., 1986; Brimacombe et al., 1988). Exploiting the stability of cross-linked RNPs, new methods have been developed to identify and analyse RNPs: i) RNA interactome capture in which poly-A RNA and its bound proteins are first selected by oligo-dT beads and co-purified proteins subsequently identified by mass spectrometry. This lead to the discovery of hundreds of hitherto unknown RBPs (Beckmann et al., 2016; Hentze et al., 2018). ii) CLIP (cross-linking and immunoprecipitation) and similar methods in which, after UV cross-linking, individual RBPs are immunoprecipitated, and co-precipitated transcripts are identified by RNA-Seq, yielding high resolution data on the RNA binding sites of the RBPs of interest (Ule et al., 2003; Granneman et al., 2009; Hafner et al., 2010; Zhang and Darnell, 2011; Ascano et al., 2012; Huppertz et al., 2014; Van Nostrand et al., 2016).

As RNA interactome capture relies on the purification of cross-linked RNPs based on hybridisation of oligo-dT beads to oligo-A sequences typically found in eukaryotic messenger RNAs, RBPs that exclusively associate with non-adenylate RNA species such as e.g. rRNA, tRNAs, snRNAs, histone mRNAs, or numerous lncRNAs cannot be identified. The same limitations apply to mRNA from bacteria and archaea that lack poly-A tails in general. Recently, RNA interactome using click chemistry (RICK; Bao et al. (2018)) has been introduced in which labeled RNA along with UV-cross-linked interacting proteins was purified in a poly-A-independent fashion. However, the method requires efficient *in vivo* labeling of RNA, limiting its application to suitable (cell culture) systems. Consequentially, no RNA-bound proteomes of prokaryotes have been determined by biochemical means to date.

A commonly used protocol to purify RNA from whole cell lysates is the single step method

(Chomczynski and Sacchi, 1987), also known as AGPC and marketed as e.g. "Trizol". First, chaotropic conditions and ionic detergents are employed to denature cellular components, followed by a biphasic extraction using the organic compound phenol. During this treatment, nucleic acids are specifically enriched in the aqueous phase. Furthermore, the pH during extraction allows to control if DNA and RNA (neutral pH) or only RNA (acidic pH) accumulate in the aqueous phase (Fig. 1b, right panel).

Here, we describe a new method that builds on the single step principle to separate RNA, proteins and cross-linked RNA-protein complexes in biphasic extractions according to their physicochemical differences. We modified the extraction chemistry using a mixture of phenol:toluol which alters enrichment of the biomolecule classes in the extraction and furthermore enabled us to "shift" cross-linked RNPs into the aqueous or interphase, respectively, This allowed us to sequentially deplete the sample of DNA and lipids, as well as non-cross-linked RNA and -proteins, highly enriching for cross-linked RNPs (clRNPs) that then can be directly analysed or further processed in more complex workflows.

Results

PTex enriches for cross-linked RNPs

The poly-A RNA interactome of human HEK293 cells has been mapped in great depth, providing an ideal reference to establish PTex-based (Phenol Toluol extraction) purification of cross-linked RNPs (Baltz et al., 2012). After irradiation with UV light at 254 nm to induce covalent RNA-protein cross-links, cells were subjected to the PTex procedure, a series of three consecutive organic extractions:

- Step 1: Phenol:Toluol (PT; 50:50), pH 7.0, no detergents
- Step 2: Phenol, pH 4.8, (chaotropic) detergents
- Step 3: Phenol, pH 4.8, (chaotropic) detergents

During the first extraction with Phenol/Toluol, RNA, proteins and cross-linked RNPs (clRNPs) are accumulating in the upper aqueous phase while DNA and membranes are predominantly found in the interphase (Fig. 1b left panel). The aqueous phase is subsequently extracted twice under chaotropic and acidic conditions using phenol (Chomczynski and Sacchi, 1987). Now, free RNA accumulates in the upper aqueous phase, free proteins in the lower organic phase and clRNPs migrate to the interphase (Favre et al., 1986) (Fig. 1b right panel). Finally, the complexes in the interphase are precipitated using water and ethanol (Chey et al., 2011) (Fig. 1c). To track the distribution of the diverse cellular molecules from total HEK293 cells during the purification procedure, we probed all phases from the intermediary steps by western blotting against HuR (ELAVL1), a well established 35 kDa RBP (Lebedeva et al., 2011; Mukherjee et al., 2011) (Fig. 1c). UV-cross-linking produces an additional band at high molecular weight which indicates the RNA-cross-linked fraction of HuR (clHuR; Fig. S1). In the PTex fraction (interphase 3), clHuR was highly enriched whereas free HuR is significantly reduced (to less than 1% compared to the input), resulting in an overall enrichment of cross-linked over non-crosslinked protein by at least 100-fold (note that this is a conservative estimate since we expect a fraction of cross-linked HuR to migrate at different molecular mass in the PAGE or not leaving the gel pocket at all, depending on the size of the bound RNA). Furthermore, abundant cellular proteins unrelated to RNA-binding such as beta-actin are not detectable in the PTex fraction. To further demonstrate the efficient removal of non-cross-linked proteins, we spiked into the cell lysates a recombinant RNA-binding protein (the central domain of *Drosophila melanogaster* Sex-lethal, denoted Sxl-RBD4 (Moschall et al., 2018)), after UV-cross-linking. PTex efficiently removes ∼99% of the free spike-in RBP (as determined by densitometry compared to the input; Fig. 1C). Similarly, removal of "free" RNA was demonstrated in an *in vitro* assay in which ^{32}P-5' labeled RNA was subjected to PTex (Fig. 1d). We next tested for depletion of DNA by PCR targeting genomic DNA (exon 5 of the IL3 gene). DNA is removed during the first two PTex steps (Fig. 1d). In sum, PTex highly enriches only cross-linked RNPs while efficiently depleting non-cross-linked proteins or nucleic acids.

To analyse the minimal length of RNA required for PTex-mediated enrichment of RNPs, we again employed the recombinant and highly purified 20 kDa Sxl-RBD4 protein which associates

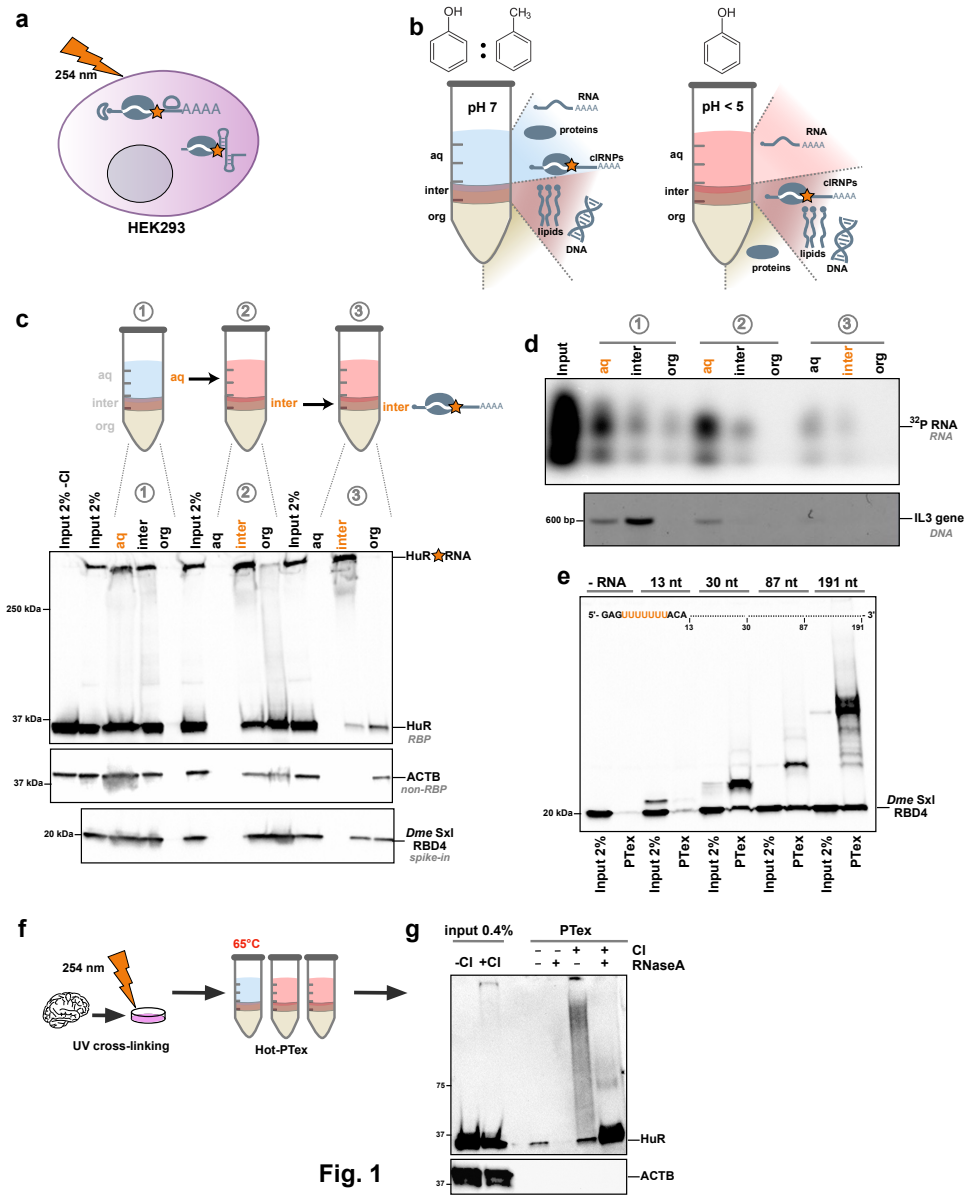

Figure 1: PTex is a fast method to purify cross-linked RNPs a) *In vivo* cross-linking of HEK293 cells using UV light at 254 nm wavelength results in covalent bonds between RNA and proteins in direct contact. Cross-linked RNPs are indicated by an orange star. b) Schematic of the separation principle of biphasic organic extractions used in PTex. Left panel: Phenol-Toluol (50:50) and neutral pH results in an accumulation of proteins and RNA in the upper aqueous phase (aq) while DNA and lipids are retained at the interphase (inter). Right panel: under acidic phenol and chaotropic conditions, non-cross-linked RNA accumulates in the aqueous phase (aq), non-cross-linked proteins in the lower organic phase (org) and cross-linked RNPs (clRNPs) are enriched at the interphase (inter). c) Step-by-step analysis of proteins in 9 intermediary steps of the PTex protocol (3 extractions with 3 phases each). Western blot agains HuR (ELAVL1, 35 kDa) demonstrates that UV-cross-linking-stabilised HuR-RNA complexes (upper edge of the blot; cross-linking is denoted with an orange star) are largely enriched after PTex (step 3 interphase). Abundant cellular non-RNA-binders such as beta-actin (ACTB) are efficiently removed by PTex. A purified fly protein (Sxl RBD4) served as spike-in as 100% non-cross-linked RBP; ∼99% of the free protein is removed by PTex. d) Step-by-step analysis of nucleic acids. 5'- end radioactive-labeled RNA was subjected to PTex *in vitro*. Like proteins, non-cross-linked RNA is efficiently depleted in the PTex fraction (step 3 interphase). Testing for DNA by PCR with specific primers against exon 5 of the interleukin 3 (IL3) gene demonstrates efficient removal of genomic DNA. e) The *Drosophila melanogaster* RBP Sxl RBD4 was bound to RNA of several length carrying the Sxl U7 recognition site and UV-cross-linked *in vitro*. PTex efficiently enriches clRNPs with RNA as short as 30 nt. f,g) Mouse brain tissue was cryo-grinded and UV-irradiated before (HOT)-PTex was performed. Western blot against HuR (ELAVL1) demonstrates recovery of cross-linked HuR from mouse tissue while beta-actin (ACTB) is efficiently depleted. For full blots see Fig. S5.

with Uracil stretches of 7 nucleotides or longer (Singh et al., 1995). We produced by *in vitro* transcription RNAs with lengths varying between 13 and 191 nucleotides, all of which contained the same Sxl-binding motif at their 5'-end. After binding and UV cross-linking *in vitro*, samples were PTex-purified and analysed by western blotting (Fig. 1e). PTex efficiently recoveres cross-linked Sxl-RBD4 complexes with RNA as short as 30 nt. Furthermore, the enrichment of cross-linked protein was also demonstrated when comparing ratios of cross-linked (shifted) and free Sxl-RBD4 in the input versus PTex samples.

Purification of RBPs from animal tissue

To test if PTex can be directly applied to tissues, we UV cross-linked whole mouse brain samples and performed PTex to extract clRNPs directly (Fig. 1f). Since brain tissue is particularly rich in lipids which accumulate in the interphase during step 1 of the PTex protocol, we increased the temperature during extractions to 65°C.The extracted clRNPs were analysed by western blotting. Cross-linked HuR is largely enriched over non cross-linked HuR after PTex (Fig. 1g). Weak detectable bands at 70 or ∼110 kDa, respectively, are likely HuR dimers and trimers as observed before (Kishore et al., 2011). This example demonstrates that PTex is not only suited for cell culture but can also extract RNPs directly from animal tissue samples; an advantage over purification protocols that depend on RNA labeling (such as RICK (Bao et al., 2018)) which are not feasible in a varietey of biological systems.

A simplified CLIP protocol

HuR (ELAVL1) has been shown to interact with mRNA and pre-mRNA in several CLIP studies and has a well-documented binding motif (5'- UUUUUU -3') (Mukherjee et al., 2014). After *in vivo* labeling of cellular RNA using 4-thiouridine (4SU) and UV irradiation at 365 nm, we performed i) classical PAR-CLIP analysis (PAR-CLIP-classic)(Hafner et al., 2010; Maatz et al., 2017) of HuR, ii) a PAR-CLIP variant using on-bead ligation of adapters (PAR-CLIP-on-beads) (Benhalevy et al., 2017), and iii) a version in which we use phenol extraction (pCLIP) for removal of unbound RNA instead of PAGE/membrane excision (Fig. 2a). We found that

pCLIP libraries contained a larger fraction of longer reads than the PAR-CLIP classic/PAR-CLIP-on-beads libraries (Fig. 2b). All three approaches identify the canonical 5'- UUUUUU -3' motif and similar profiles of HuR-bound RNA clusters map to intronic and 3'UTR regions (Mukherjee et al., 2014) in all three variants (Fig. 2c-e). The clusters could also be mapped to the same 3'UTR loci when comparing HuR binding sites in tubulin and splicing factor Srsf6 mRNA. Although we only performed a low-read-coverage experiment as proof-of-principle, our results demonstrate that phenolic extractions of RNPs such as PTex can be integrated into more complex workflows such as (PAR-)CLIP and have the potential to simplify CLIP-type approaches by enriching for clRNPs or remove unbound RNA transcripts.

A global snapshot of human RNA-protein complexes

Despite the recent advances in mapping RBPs in many species, two general issues have not been addressed to date: i) the fact that RNA interactome capture (Castello et al., 2012; Baltz et al., 2012; Beckmann et al., 2015) targets only polyadenylated RNAs suggests that many RBPs that bind non-adenylate RNAs are missed by this experimental approach (Bao et al., 2018); and ii) although UV cross-linking has been widely used to research RNA-protein interactions, no systematic study has been performed to determine the optimal irradiation conditions for efficient cross-linking of individual RNPs.

Using PTex as an unbiased approach, we set out to explore RNA-protein interactions cell-wide in HEK293 cells. To test the effect of different energy UV-irradiation, we employed besides the most commonly used 0.15 J/cm^2 irradiation at 254 nm wavelength also irradiation at two additional energy levels: 0.015 and 1.5 J/cm^2; spanning two order of magnitude of UV irradiation dosage. This setup was then used to comprehensively map RNA-protein interactions beyond the established poly-A RNA-bound proteomes (Baltz et al., 2012; Hentze et al., 2018). We independently irradiated HEK293 cells at all 3 energies and performed PTex purification of cross-linked and non-cross-linked RNPs from whole cells. Transcriptomics by RNA-Seq (Fig. 3a), and proteomics by mass spectrometry (Fig. 4a) and label free quantification (LFQ) were performed using total RNA and protein preparations as input controls.

Fig. 2

Figure 2: pCLIP: a fast PAR-CLIP variant employing phenolic extraction a) Schematic comparison of PAR-CLIP variants. b) Read length distribution of uniquely mapping reads utilized for determine binding sites (cluster) of HuR (ELAVL1). PAR-CLIP samples were processed using PARpipe (see methods). c) Relative proportion of PARalyzer-derived cluster annotation. d) Heatmap of relative positional binding preference for intron-containing mRNA transcripts for each of the six HuR PAR-CLIP samples. Sample-specific binding preferences were averaged across selected transcripts (see methods). The relative spatial proportion of 5'UTR, coding regions and 3'UTR were averaged across all selected transcript isoforms. For TES (regions beyond transcription end site), 5' splice site, and 3' splice site, we chose fixed windows (250 nt for TES and 500 nt for splice sites). For each RBP, meta-coverage was scaled between 5'UTR to TES. The 5' and 3' intronic splice site coverage was scaled separately from other regions but relative to each other. e) Overrepresented motifs in HuR clusters annotating to 3'utr and intronic regions. Only the most enriched 6mer motif is shown. Motifs were determined including RNA-structure using Zagros software (Bahrami-Samani et al., 2015). f) and g) Genome browser shots of TUBB and SRSF6 example genes showing reproducible 3'UTR binding sites. Track y-axes represent uniquely mapping read count.

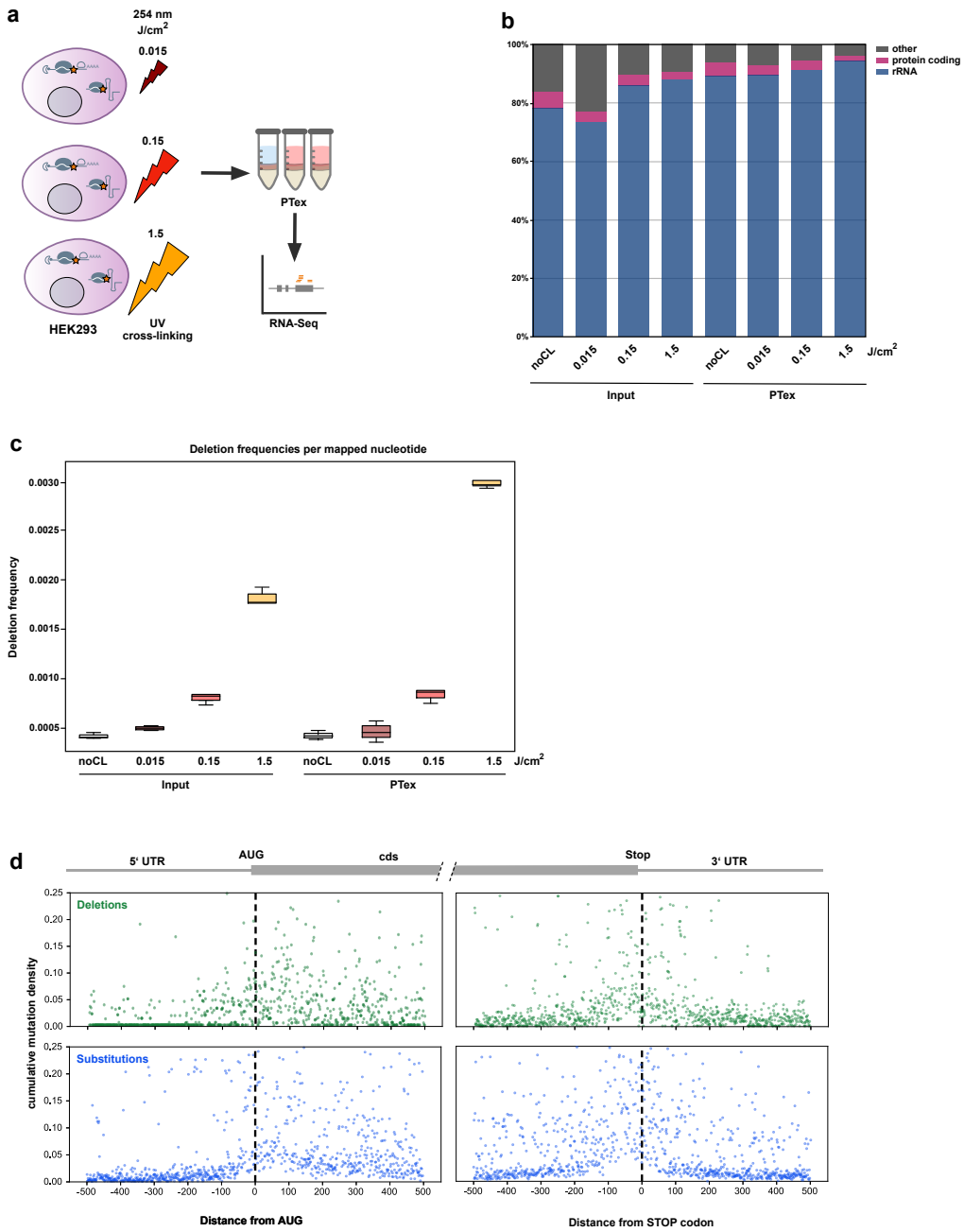

Fig. 3

Figure 3: A global snapshot of RNPs in HEK293 cells (RNA) a) Schematic of the experimental setup: HEK293 cells were UV-cross-linked using 0 (noCL),0.015 (dark red), 0.15 (red) and 1.5 (dark yellow) J/cm^{2}254 nm light in triplicates. Total RNA from input (whole cell lysate) and PTex-purified samples were analysed by RNA-Seq. b) Distribution of RNA classes in input and PTex samples. c) Deletions in RNA from input and PTex samples; frequency of mutations in transcripts correlate with higher UV doses. d) Mutations (deletions = green, substitutions = blue) enriched in UV-irradiated samples were plotted to their position relative to AUG and Stop codon in coding sequences and serve as indicator for protein-binding sites. Note that we cannot delineate which protein bound to which position. Plots for PTex are shown; for input see Fig. S3.

PTex-purified RNPs: RNA We first analysed PTex-purified transcripts. Unlike proteins which can be grouped into RNA-interactors and non-interactors, all cellular RNA can be expected to be associated to proteins (Keene, 2007; Hogan et al., 2008; Gehring et al., 2017). In line with this, we find a similar distribution of RNA classes when comparing RNA from inputs and PTex by RNASeq with the vast majority of transcripts being ribosomal RNA (Fig. 3b). Protein-cross-linked RNA is known to enrich for mutations during reverse transcription in the RNASeq workflow (Fig. 3c,S2) (Hafner et al., 2010; Granneman et al., 2009). Such mutations can then be used as "beacons" to map protein binding sites in transcripts (Hafner et al., 2010; Baltz et al., 2012; Freeberg et al., 2013). We used pyCRAC (Webb et al., 2014) to map deletion and substitution mutations enriched in UV-treated PTex samples to 100 nt windows around the 5' AUG start codon and 3' Stop codon of mRNA reads (Fig. 3d,S3). We find that most mutations (indicating protein binding) are within the first 100 nt after the AUG and the last 100 nt before the Stop codon. This was observed before in a global protein occupancy profiling study (Schueler et al., 2014) and can potentially be attributed to cross-linking of ribosomal proteins or translation initiation/termination factors to mRNA, as ribosomal profiling experiments show increased ribsomal footprint densities at these regions, indicating longer dwell times and a higher potential for cross-linking at these sites (Ingolia et al., 2011).

PTex-purified RNPs: proteins Proteins which were not identified by MS in all 3 replicates after PTex were removed (Fig. 4a). For the remaining proteins, ratios of cross-linked over non-cross-linked (CL/-CL) LFQ intensities (from the PTex experiments) were calculated. P-values from a moderated t-test were then used for multiple testing (Benjamini-Hochberg). Using these stringent tests, we identify 3188 shared among the three conditions; out of these 3037 proteins are significantly enriched in a UV-irradiation-dependent fashion in all samples using a cut-off of FDR 0.01 (Fig. 4b,c). Analysis of general protein features (molecular mass, pI, cellular abundance, hydrophobicity, Fig. 4d,e) demonstrates that the PTex procedure does not enrich for a particular subgroup of cellular proteins based on chemical properties or expression level.
In principle, extended UV exposure should increase the chance for cross-linking events and thus for subsequent protein recovery by PTex (Beckmann et al., 2015). We therefore expected a

gradual increase in enrichment of RNA-interacting proteins after PTex with higher UV dose. However, we only find increased recovery for most proteins when comparing low and medium energies. Surprisingly, irradiation with 1.5 J/cm^2 on the other hand leads to a significant decrease in recovery (Fig. 4f). We observed RNA degradation when analysing total RNA from HEK293 cells irradiated with 1.5 J/cm^2 254 nm UV light (Fig. S4; (Beckmann et al., 2015)). If this degradation was due to damage of nucleic acids induced by high energy UV light or a result of secondary processes during the extended time of treatment is unclear; in any case extensive shortening of RNAs will cause a loss of RNPs purified by PTex.

HEK293 RNA-interacting proteins So far, 700 - 2000 well-established and recently identified eukaryotic RBPs have been described (reviewed in Gerstberger et al. (2014) and Hentze et al. (2018)). To find more than 3000 proteins to be enriched as RNA-associated by PTex is unexpected. Considering that deep proteome studies detect around 10,500 proteins in HEK293 cells (Geiger et al., 2012), we find nearly a third of the expressed cellular proteome to be associated with RNA.

To test sensititvity and specificity of our approach, we first performed global GO enrichment analysis showing that terms from all aspects of RNA biology are the most enriched among PTex-purified proteins (Fig. 5a). At the same time, protein classes with no general role in RNA biology such as transporters and (trans-)membrane proteins were depleted by PTex (Table S2). Known RNA-binding domains (RBDs) such as the RNA recognition motif (RRM) or helicase folds (DEXDc, HELICc) were significantly enriched among PTex-purified proteins (Fig. 5b; Table S2). 82 PTex-enriched proteins contain a WD40 fold; a domain found to directly bind snRNA in Gemin 5 (Jin et al., 2016), taking part in rRNA biogenesis (Erb1 (Wegrecki et al., 2015)) and found in RBPs (Castello et al., 2012). Other enriched domains are: PCI (Proteasome, COP9, Initiation factor 3; domain function unknown) present e.g. in eIF3A, AAA (ATPase, see below) fold and tetratrico peptide repeat region (TPR) as found in the yeast Clf1p splicing factor (Wang et al., 2003), Ski complex (Halbach et al., 2013) and the translation terminator Nro1 (Rispal et al., 2011).

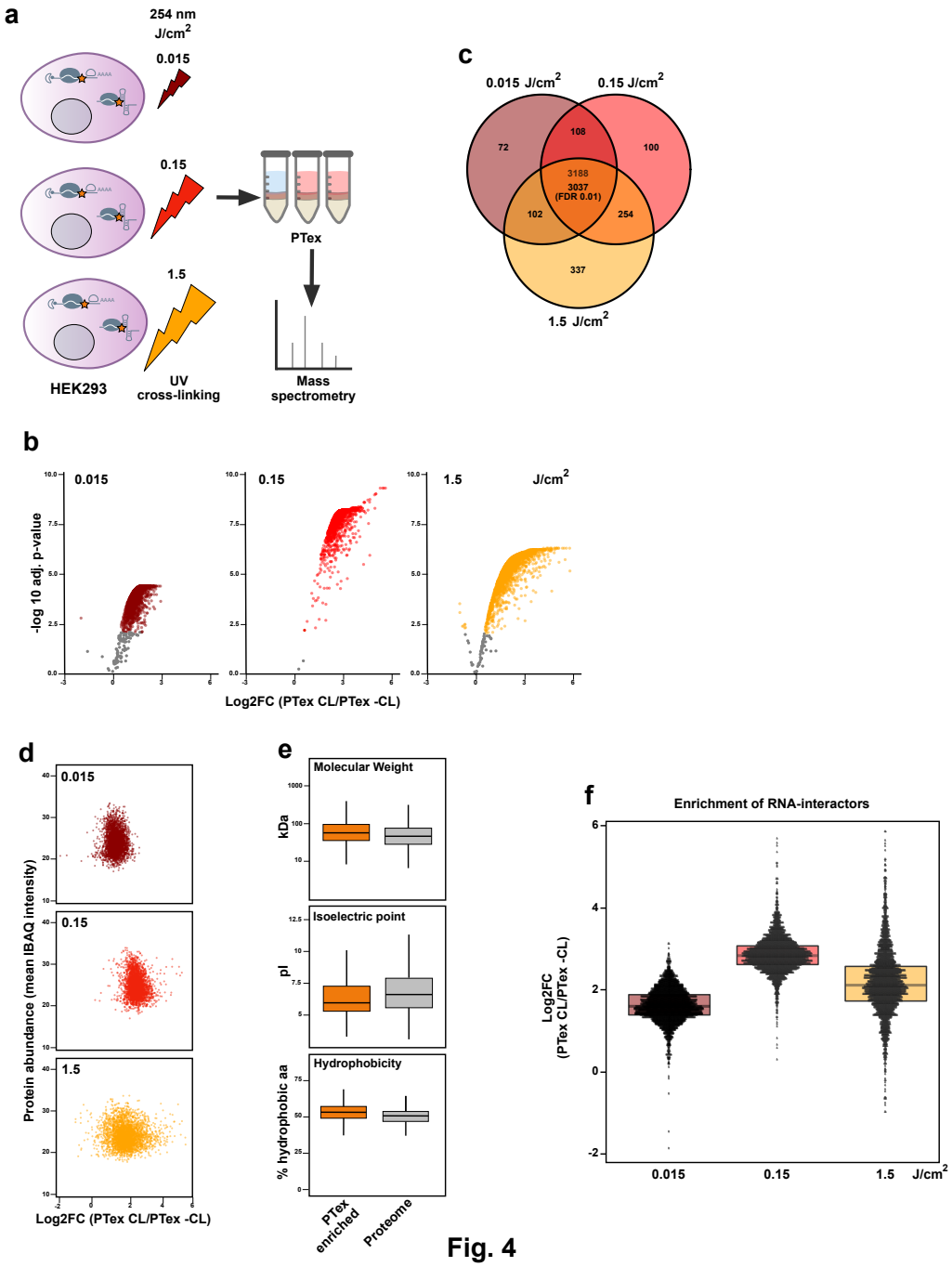

Figure 4: A global snapshot of RNPs in HEK293 cells (proteins) a) Schematic of the experimental setup: HEK293 cells were UV-cross-linked using 0 (noCL),0.015 (dark red), 0.15 (red) and 1.5 (dark yellow) J/cm^{2}254 nm light in triplicates. From each, input (whole cell lysate) and PTex-purified sample were analysed by label-free mass spectrometry. b) Volcano plots of proteins enriched by PTex (FDR 0.01) under the three cross-linking conditions. c) Overlap of PTex-enriched proteins (enriched in all 9 replicates, FDR 0.01) is 3037 (these PTex proteins are from here on colored in orange). d) Protein abundance (IBAQ intensities of input samples) does not correlate with PTex enrichment (log2-fold change of intensities [CL/-CL]) e) PTex does not select for a subset of proteins based on general features such as molecular weight, pI or hydrophobicity. f) Enrichment of proteins by PTex drops at 1.5 J/cm^{2}254 nm light compared to lower doses.

To test that we are enriching for RNA-binders specifically, we calculated the probability of recovering known RBPs interacting with different RNA classes using the hypergeometric test: we enrich for ribosomal proteins (42/47 large subunit and 30/33 small subunit (Ban et al., 2014); p-value: 1.36×10^{-30}), NSUN2 and tRNA synthetases (19/20 cytosolic; p-value: 7.4×10^{-10}) indicating that our approach indeed captured cellular RNPs in a poly-A-independent fashion. We recover 70% of poly-A RNA-binding proteins found in HEK293 cells by Baltz et al. (2012) (Fig. 5c) although different UV irradiation strategies were used (254 vs. 365 nm cross-linking; p-value: 1.26×10^{-139}). Importantly, the largest overlap was with RBPs recently found in HeLa cells using the also unbiased RICK technique (Bao et al., 2018) (92% for high confidence RBPs and 86% of non-poly-A RBPs, respectively). Recent studies show that the boundaries between RNA- and DNA-binding are rather blurry and nuclear DNA-binders were found to interact with RNA (Conrad et al., 2016). We identified proteins involved in replication and response to DNA damage (Table S2) such as DDX54 (Milek et al., 2017) but in general, DNA-binders such as transcription factors were underrepresented, demonstrating that PTex does not select for DNA-specific binding proteins in particular (Fig. 5c). To rule out that previously described RBPs are more efficiently recovered in PTex than the newly identified RNA-associated proteins (which could indicate carry-over of proteins unrelated to RNA interactions), we compared the distribution of the HEK293 mRNA-binding proteins (Baltz et al., 2012) in the PTex enrichment. The established RBPs are similarly enriched along the dynamic range (from no enrichment to log2FC PTex [CL/-CL] of 6; Fig. 5d) of PTex and hence display no difference to the novel RNA-interactors.

The presented results demonstrate that PTex is specific for RNPs. But why weren't the same proteins discovered to associate with RNA before? The majority of recently discovered RBPs are interacting with mRNA (Hentze et al., 2018); a RNA class which is highly heterogeneous in its sequence but represents only ∼5% of the cellular RNA pool. The differences in between interactome capture (poly-A RNPs) and PTex (RNPs in general) is best demonstrated in the case of the eukaryotic RNA exosome (Houseley and Tollervey, 2009; Makino et al., 2013). The core exosome complex consists of ten protein subunits (Exo-10) from which only one protein

(Rrp44) is catalytically acting on RNA as exo- and endoribonuclease (Fig. 5f). The remaining nine proteins (Exo-9) are forming a barrel-like complex in which RNA can be channelled through before it is degraded by Rrp44, but Exo-9 proteins do not degrade or modify the RNA itself (Lorentzen et al., 2008; Makino et al., 2013). Still, all ten subunits are positioned to directly interact with RNA and multiple interactions with the individual subunits have been demonstrated in high resolution structure studies (Wasmuth et al., 2014; Kowalinski et al., 2016; Zinder et al., 2016; Schuller et al., 2018). Hence, all ten subunits are amenable to UV-cross-linking and, as a result, 9 out of the 10 subunits were identified by PTex (Fig. 5g).

Previously it was demonstrated that interactome capture enriches for mRNA-binding proteins with a high isoelectric point (pI; Fig. 5e) (Castello et al., 2012; Beckmann et al., 2015). However, sequence- and oligo-dT-independent approaches such as RICK (Bao et al., 2018) or PTex identify more proteins with a pI <6. Proteins with a low pI are overall negatively charged at cellular pH and thus unlikely to interact with RNA in an unspecific manner due to electrostatic repulsion of likewise negatively charged RNA. Indeed, 7 of the Exo-10 protein subunits have the mRBP-untypical isoelectric point below pH 6 (Fig. 5h). Inside the central exosome channel, RNA of 30-33 nt or 9-10 nt length has been found in *in vitro* and in CRAC analyses (Delan-Forino et al., 2017); the former being of sufficient length for efficient recovery by PTex (Fig. 1e). In sum, PTex enriches for the (near) complete exosome core complex while interactome capture from the same cell line only found a single subunit (Baltz et al., 2012) (Fig. 5g).

But we also enrich for proteins which have no established role in RNA biology such as subunits of the human proteasome (Table S2). However, from the 28 proteins of the 20S core complex, Psma5 and Psma6 were reported to display RNase activity in purified complexes (Petit et al., 1997). Importantly, proteasome-associated RNAs were shown to lack poly-A stretches (Horsch et al., 1989; Brooks, 2010) and ATPases of the AAA family in the 19S proteasome regulatory particle were found to be recruited to RNA polymerase I (rRNA) transcription sites (Fátyol and Grummt, 2008). Hence, none of the proteasome-related RNA activities are approachable via poly-A RNA-mediated purification. The proteasome is a multi-protein complex with structural similarities to the exosome (Makino et al., 2013) and RNAs interacting with Psma5/6 are likely

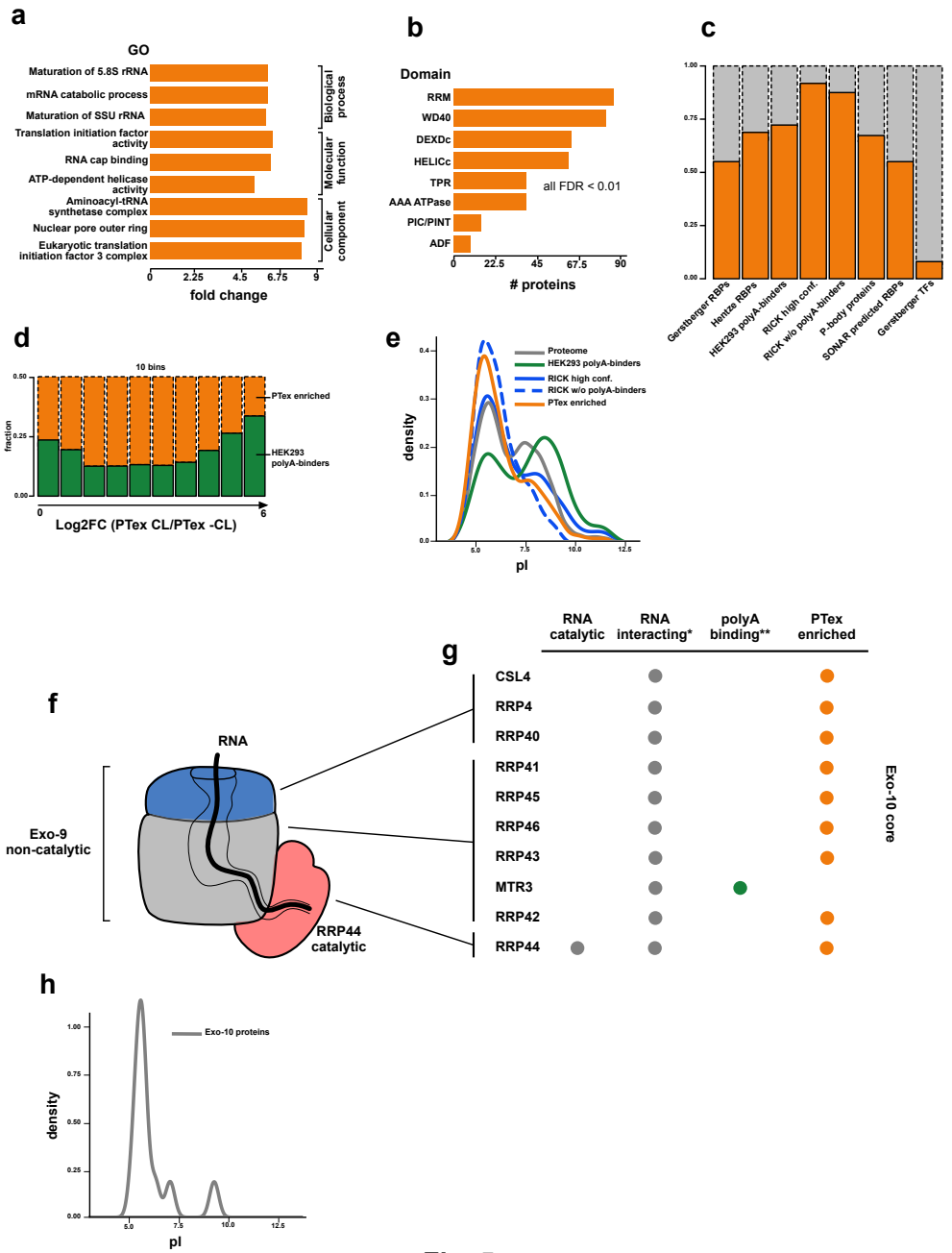

Fig. 5

Figure 5: Features of RNA-interacting proteins found by PTex. a) Top 3 enriched GO terms (CC, MF, BP) and b) enriched protein domains in PTex-purified proteins from HEK293 cells. c) PTex-purified proteins overlap with well-described RBPs but not transcription factors. Recovery of Gersberger RBPs and transcripiton factors (TFs) reviewed in Gerstberger et al. (2014), a recent review on RBPs by Hentze et al. (2018), HEK293 poly-A binders (Baltz et al., 2012), RBPs found by RNA interactome using click chemistry (RICK; (Bao et al., 2018)), P-body components (Hubstenberger et al., 2017) and a recent prediction of candidate RBPs (SONAR, (Brannan et al., 2016)). d) Distribution of previously identified HEK293 mRNA-binding proteins (green; (Baltz et al., 2012)) in PTex; each bin represents 10% of the 3037 PTex proteins from lowest to highest enrichment. e) mRNA-binding proteins display a bimodal pI distribution pattern with peaks at pH 5.5 and 9.5 (Baltz et al., 2012; Castello et al., 2012). RNA-interactors in general peak at pI 5-6 as found by PTex and RICK (Bao et al., 2018). **Proteins of the RNA exosome are prototype PTex proteins.** f) The RNA exosome core consists of ten subunits: nine non-catalytically active proteins (Exo-9) forming a barrel-like structure and an additional RNase (Rrp44; Exo-10); modified from Makino et al. (2013). g) *All exosome subunits are labeled "RNA-binding" (Uniprot.org); **green = identified via poly-A selection in Baltz et al. (2012); orange = enriched in PTex. h) The isoelectric point (pI) from 7 out of 10 exosome subunits is <6 as regularly found in PTex-enriched proteins.

to be UV-cross-linked to other subunits as well.

PTex allows a first snapshot of RNA-associated proteins in bacteria

Prokaryotic mRNAs lack poly-A tails and are thus not approachable by oligo-dT-based methods. We used the pathogen *Salmonella* Typhimurium harbouring a chromosomally FLAG-tagged Hfq protein (Holmqvist et al., 2016) to test PTex in bacteria. Hfq is an abundant RNA-binder facilitating mRNA:ncRNA interactions in Gram-negatives (Holmqvist et al., 2016; Gorski et al., 2017). We used a slightly modified PTex protocol (Hot-PTex) in which RNP extraction from *Salmonella* grown to OD_{600} 3.0 was performed at 65°C thereby supporting cell lysis (Fig. 6a). Hfq is a 17 kDa protein and forms a homo-hexamer in bacterial cells; the complex has been shown to resist normal Laemmli PAGE conditions (Holmqvist et al., 2016) and is also visible in western blots (Fig. 6b). After UV irradiation, PTex purification and RNase treatment, we observe a shifted Hfq monomer band which we attribute to residual cross-linked RNA fragments resulting in a slightly higher molecular mass. The physiologically relevant Hfq hexamer is also strongly enriched compared to non-UV samples, indicating that also the complex is still bound to remaining RNA fragments.

We next used the Hot-PTex fraction of *Salmonella* cells to map RNA-associated proteins by mass spectrometry (Fig. 6c). Comparing recovered protein intensities from UV-irradiated versus control cells (biological duplicates), we find 172 proteins, among them 33 ribosomal proteins, components of the RNA polymerase complex (subunit α, σ factor RpoD, DksA) and 4 out of the 5 established mRNA-binding proteins of *Salmonella* (Hfq, ProQ, CspC/CspE) (Holmqvist et al., 2016; Smirnov et al., 2016; Michaux et al., 2017). 113 of the enriched proteins are so-far unknown to interact with RNA. To validate our findings, we picked YihI; a putative GTPase-activating protein which was speculated to play a role in ribosome biogenesis (Hwang and Inouye, 2010). Using a YihI-FLAG-tagged mutant strain, we performed UV-crosslinking, immunoprecipitation and radioactive labeling of co-purified RNA (known as "PNK assay" as described in (Tawk et al., 2017)), validating that YihI is indeed associated to RNA *in vivo* (Fig. 6d). We furthermore find proteins with known RNA-binding domains (RBDs) such

as the nucleic acid binding OB-fold (present in RpsA, RpsL, RplB, CspC, CspE, Pnp, RNaseE, Ssb, NusA) and domains which were also detected in RBPs when screening eukaryotic cells: the afforementioned AAA ATPase fold in the ATP-dependent protease ATPase subunits HslU, ClpB and ClpX (the latter was also validated by PNK assay; manuscript in preparation), or thioredoxin domains in the Alkyl hydroperoxide reductase c22 protein (ahpC), Thiol:disulfide interchange protein (dsbA) and Bacterioferritin comigratory protein (bcp) (Castello et al., 2012; Beckmann et al., 2015). As in many other species, we find glycolytic enzymes to associate to RNA (Pgk, Pgi) (Beckmann et al., 2015; Matia-González et al., 2015). Using GO terms for functional annotation of the RNA-associated proteins, the most significant terms are "translation" and other terms connected to the ribosome as expected (Fig. 6e). To the best of our knowledge, the periplasmic space is generally considered to be devoid of RNA. We still recover RNA-associated proteins which localise to the outer membrane (Table S3). As in the case of HEK293, we cannot distinguish here between RBPs that actively act on RNA and proteins which are associated to RNA for e.g. structural reasons. However, recent studies in several Gram-negative bacteria demonstrate that secreted outer membrane vesicles (OMVs) contain RNA which indicates that bacterial transcripts must be sorted to the outer membrane via a yet to be determined pathway (reviewed in Tsatsaronis et al. (2018)). It is tempting to speculate if the here identified proteins are involved in such a process or are mere bystander proteins.

Overall, we noticed that enrichments even from known RBPs were lower in *Salmonella* compared to HEK293 cells and we anticipate that additional modifications in UV-crosslinking and/or cell lysis could improve sensitivity when applying (Hot-)PTex in bacteria. To the best of our knowledge, PTex is the first approach that can purify bacterial RNPs in an unbiased fashion without the necessity of immunoprecipitation or introduction of modifications (tag, overexpression, etc.), rendering it a novel tool for cell-wide RBP identification and studying bacterial RNA-protein interactions.

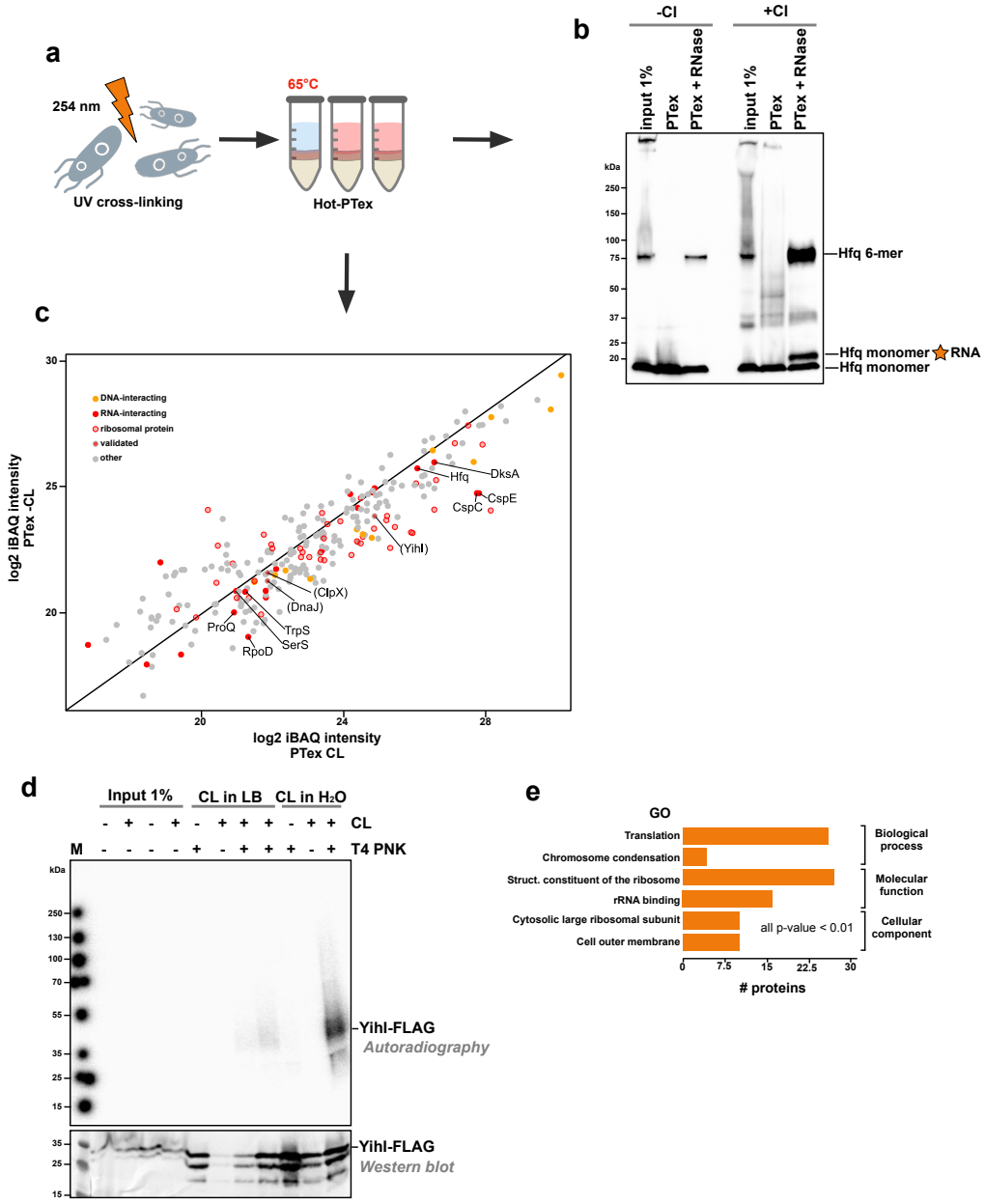

Fig. 6

Figure 6: PTex recovers bacterial RNPs a) *Salmonella* Typhimurium SL1344 Hfq-FLAG was UV-cross-linked and HOT-PTex was performed to purify bacterial RNPs. b) Western blot using an anti-FLAG antibody demonstrates recovery of Hfq monomers linked to RNA. Note that the physiologically active Hfq hexamer partially withstands SDS-PAGE conditions (Holmqvist et al., 2016) and that this complex is also enriched after PTex. c) RNPs in *Salmonella* were purified by PTex globally. 172 Proteins enriched after UV-crosslinking (PTex CL) contain ribosomal proteins (transparent red), known RBPs (red) and DNA-binders (orange). Individual enriched proteins not known to associate with RNA before were used for validation (in parentheses). d) Validation of YihI-FLAG: Western blot of cell lysates (input) and immunoprecipitates +/- UV irradiation (+/- CL). RNA-association is confirmed by radioactive labeling of RNA 5' ends by polynucleotide kinase (PNK) using autoradiography; a signal is exclusively detectable after UV-crosslinking and radiolabeling of precipitated RNA. e) GO terms significantly enriched among the RNA-associated proteins.

Discussion

UV cross-linking of RNPs appears rather inefficient. Even after high UV doses only ∼1-10% of any given RBP can be covalently coupled to ribonucleic acids in human cell culture (Favre et al., 1986; Castello et al., 2012) and yeast (Granneman et al., 2009; Beckmann et al., 2015; Matia-González et al., 2015; Beckmann, 2017). Importantly: this reflects both cross-linking efficiency and fraction of protein bound to RNA; in other words how much of the protein is in a steady state associated with the RNA? What is of experimental interest is therefore only a minor fraction of the RBP (the one cross-linked to RNA), while the vast excess of protein stays in a non-cross-linked state.

With PTex, we are exploiting the physicochemical differences between cross-linked hybrid RNA-protein molecules on the one hand and the non-cross-linked proteins and RNA of a cell on the other for selective purification of complexes only. The method was designed to select against secondary RNA-binders by using denaturing and chaotropic conditions in which RNA-protein interactions which were not covalently cross-linked are not preserved (Chomczynski and Sacchi, 1987), and by selecting proteins which were enriched in a UV-dependent fashion. PTex is a fast and simple modular protocol which can be performed in about 3 hrs. Our approach is independent of the UV wavelength applied for irradiation (254/365 nm) and the type of biomaterial used (human/bacterial cell culture, animal tissue), and does not rely on presence of a particular RNA sequence such as poly-A tails.

Our work provides a first systematic analysis of the effects of different UV irradiation dosage on RNA-protein cross-linking (Table S1). We hope that this resource will aid researches to establish suitable conditions for cross-linking of individual RNPs. The decrease in recovery of proteins after using 1.5 J/cm^2 254 nm light (Fig. 4f) demonstrates that extensive irradiation/cross-linking can have adverse effects on protein recovery. Next to the observed RNA degradation, cross-linked peptides released by trypic digestion are notoriously difficult to identify in MS experiments and hence increasing the amount of protein cross-linking might negatively impact protein identification (Kramer et al., 2014; Castello et al., 2016; He et al., 2016; Mullari et al., 2017). This is not the case for PTex-purified and input control transcripts in which deletion mutations

accumulate at high UV settings which can then serve as marker for protein interaction sites (Fig. 3c,d) (Granneman et al., 2009). Our results demonstrate that when investigating RNPs on a global scale, the crosslinking strategy should be adapted to the biological question: not all proteins are interacting with RNA and increasing the UV dose can be disadvantageous for RBP recovery since severely degraded RNA will cause less-efficient purification by PTex and cross-links will impair identification by mass spectrometry. Conversely, allmost all RNA can be expected to be bound by a set of proteins under physiological conditions (Hogan et al., 2008; Gehring et al., 2017) which explains why we also observe an increase in mutations in input RNA (Fig. 3c,S2,S3). In contrast to protein recovery, (partial) *in vivo* RNA degradation will not impair recovery of cross-linked transcripts since RNase treatment/RNA fragmentation is part of CLIP and RNASeq workflows already and cross-linked RNA will not be lost during cDNA preparation or sequencing (Ule et al., 2003; Granneman et al., 2009; Hafner et al., 2010; Zhang and Darnell, 2011; Ascano et al., 2012; Huppertz et al., 2014; Van Nostrand et al., 2016).

Our findings indicate that up to a third of a cell's proteins can associate with RNA *in vivo* which raises the question of the underlying biological function of these interactions. In this respect, it is intriguing to see that the Exo-9 proteins are interacting with and can be cross-linked to RNA although none of these subunits diplay RNase activity themselves. By increasing the detection efficiency for UV-cross-linked complexes e.g. by recovery of proteins interacting with RNA as short as 30 nt, we now have to separate classical RBP functionalities such as RNA degradation, transport or modification from RNA-interactors which are in physical contact with RNA due to structural organisation as in the case for ribosomal or exosome proteins. With PTex, we have developed a tool for fast recovery of RNPs in general which will allow us interrogate the functionality of individual proteins in RNA biology. So far, eukaryotic proteomes have been extensively scrutinised for RNA-binding proteins in the recent years. The two other kingdoms of life - archea and prokaryotes - could not be investigated for technical reasons. We here provide a first RNA-bound proteome from *Salmonella* Typhimurium, demonstrating that PTex will now allow to expand global RNP analysis to species in all three branches of the tree of life.

Methods

Human cell celture and in vivo cross-linking Human embryonic kidney cells (HEK293) were grown on 148 cm^2 dishes using DMEM high glucose (Dulbecco's Modified Eagle, glucose 4.5g/L, Gibco, 41966-029) supplemented with 10% bovine serum (Gibco, 10270-106), penicillin/streptomycin (100 U/mL 0,1 mg/mL; Gibco, 15140-122) at 37°C with 5% CO2. After reaching 80% confluence, cells in monolayer were washed once with cold phosphate buffer saline (DPBS; Gibco, 10010-015) and placed on ice. Then, DPBS was removed completely and cells were irradiated with 0.015 - 1.5 J/cm2 UV light (λ= 254 nm) in a CL-1000 ultraviolet crosslinker device (Ultra-Violet Products Ltd), collected in 15 mL tubes, pelleted by centrifugation (1000 x g, 3 min, 4 °C), aliquoted in 2 mL tubes and stored at -20/-80°C (+CL). Non-irradiated cells were used as non-cross-link control (-CL). Additionally, the potential UV damage on the RNA after exposure to the different radiation energies was assessed: RNA isolated from HEK293 cells before and after exposure UV light by phenol extraction (Chomczynski and Sacchi, 1987) were analysed with the DNA 1000 Chip Kit in a Bioanalyzer 2100 (Agilent, 5067-1504).

Bacterial cell culture and in vivo cross-linking *Salmonella* Typhimurium Hfq::x3FLAG (Holmqvist et al., 2016) was grown on LB medium to stationary phase (OD_{600} = 3). Aliquots of 20 ml were pelleted (4,000 x g, 15 min, 4°C), and resuspended with 2 ml of water. Cells were cross-linked, on ice, with 5 J/cm^2 UV light (λ= 254 nm) in a CL-1000 ultraviolet crosslinker device (Ultra-Violet Products Ltd), snap-frozen and stored at -80°C. Bacterial suspensions equivalent to 2.5 ml of initial culture were used as input in HOT-PTex (see below) for Western Blot in figure 6b. *Salmonella* enterica subsp. enterica Serovar Typhimurium strain SL1344 used for the mapping of RNA-protein interactions was grown in LB medium to OD_{600} 2.0. Half of the cultures were cross-linked in a Vari-X-linker (UVO3, www.vari-x-link.com), using UV light (λ= 254 nm) lamps for 90 seconds. Fractions of 10 ml from each, cross-linked and non-cross-linked cultures, were harvested by filtration as described in van Nues et al. (2017).

Construction of bacterial strains Yihi::x3FLAG::KmR awas constructed following the procedure based on the Lambda Red system developed by Uzzau et al. (2001). The system is based on two plasmids: pKD46, a temperature-sensitive plasmid that carries gamma, beta and exo genes (the bacteriophage λ red genes) under the control of an Arabinose-inducible promoter, and pSUB11, carrying the x3FLAG::KmR cassette. The cassette in pSUB11 was PCR-amplified with primers (5'-GAA GCA GGA AGA TAT GAT GCG CCT GCT AAG AGG CGG CAA CGA CTA CAA AGA CCA TGA CG-3' and 5'-GGG TTA TAA GCA GGA CGG GCA AGC CCA CGG TGT AAA CCC GCA TAT GAA TAT CCT CCT TAG-3'), the 5' ends of which were designed to target the 3'end of the gene of interest, digested with DpnI at 37°C for one hour and, upon purification, used for subsequent electroporation. *Salmonella* Typhimurium SL1344 harbouring plasmid pKD46 was grown in LB containing Ampicillin (100 µg/ml) and L-Arabinose (100 mM) at 30°C to an OD_{600} of 0.8. Cells were incubated on ice for 15 min, and centrifuged for 30 min at 3220 x g at 4°C and resuspended in ice-cold water. The wash was repeated three times. On the final wash, cells were resuspended in 300 µl water and electroporated with 200 ng of PCR product. Cells were recovered for one hour in LB at 37°C on a tabletop thermomixer at 600 rpm, plated on LB agar with Kanamycin (50 µg/ml) overnight. The following day, 10 colonies per strain were picked, resuspended in PBS and streaked on plates containing Ampicillin or Kanamycin and incubated at 40°C. Colonies that showed resistance to Kanamycin but not to Ampicillin were selected for further analysis, and the correct expression of the epitope tag was verified by western blot.

PTex (phenol-toluol extraction) HEK293 suspensions in 600 µl of DPBS (5-8 x 10^6 cells, +/-CL) were mixed with 200 µl of each: neutral phenol (Roti-Phenol, Roth 0038.3), toluol (Th.Geyer, 752.1000) and 1,3-bromochloropropane (BCP) (Merck, 8.01627.0250) for 1 min (21°C, 2.000 r.p.m, Eppendorf ThermoMixer) and centrifuged 20.000 x g 3 min, 4°C. The upper aqueous phase (Aq1) was carefully removed and transferred to a new 2 ml tube containing 300 µl of solution D (5.85 M guanidine isothiocyanate (Roth, 0017.3); 31.1 mM sodium citrate (Roth, 3580.3); 25.6 mM N-lauryosyl-sarcosine (PanReac AppliChem, A7402.0100); 1% 2-mercaptoethanol (Sigma)). Then, 600 µl neutral phenol and 200 µl BCP were added, mixed

and centrifuged as before. After phase separation, the upper 3/4 of Aq2 and the lower 3/4 of Org2 were removed. The resulting interphase (Int2) was kept and then mixed with 400 µl water, 200 µl ethanol p.a., 400 µl neutral phenol and 200 µl BCP (1 min, 21°C, 2.000 r.p.m, Eppendorf ThermoMixer) and centrifuged as previously. Aq3 and Org3 were carefully removed, while Int3 was precipitated with 9 volumes of ethanol (-20°C, 30 min to overnight). Samples were centrifuged during 30 min at 20.000 x g, pellets dried under the hood for max. 10 min and solubilised with 30 µL Laemmli buffer at 95°C for 5 min; or the indicated buffer/temperature according to the downstream application. For global mapping of RNA-protein interactions in HEK293 cells, 3 replicates from each UV irradiation energy were used.

HOT-PTex from mouse tissue 30 mg of mouse-brain tissue was resuspended in 600 µL of DPBS, mixed during 5 min at 65°C in the presence of 0.5 g of 100 µ low-binding zirconium beads (OPS Diagnostics, LLC) and the phenol-toluol-BCP mix described above. After centrifuging (20.000 x g 3 min, 4°C.) Aq1 was carefully removed. Consecutive extractions were done as described before with the solely difference that mixing steps were performed at 65°C (2.000 r.p.m, Eppendorf ThermoMixer).

HOT-PTex from Salmonella For global mapping of RNA-protein interactions in *Salmonella* cells, 2 biological replicates were used. Bacteria attached to filters (one per replicate, per treatment) were collected with 12 ml of DPBS, aliquots of 4 ml were pelleted at 20.000 x g, 2 min, 4°C. A modification of the step 1 was introduced in HOT-PTex in order to improve removal of free-proteins, as follows: bacterial pellets (+Cl/-CL) resuspended in 400 µl of DPBS were mixed with phenol-toluol-BCP (200 µl each), and 0.5g of zirconium beads during 5 min at 65 °C (2.000 r.p.m, Eppendorf ThermoMixer). After centrifugation at 20.000 x g, 3 min, 4°C, the upper aqueous phase (Aq1) was mixed again with the same volumes of phenol-toluol-BCP (without beads) during 1 min at 65°C. Then the aqueous phase was carefully transfered to a third tube where steps 2 and 3 were performed as described in the PTex protocol but at 65°C.

Precipitation of individual PTex steps Extractions were carried as described before using sets of three tubes containing synthetic 30-50 nt RNAs 5´-labelled ^{32}P-ATP or 2-3x10^6 HEK293 cells (+/- CL). Spike-in controls were: 0.25 µg of Sxl-RBD4 or 10 µg pUC19-sxl-target per tube. PCRs were performed using primers designed to amplify an endogenous DNA fragment from the il-3 gene (576bp, forward primer: 5´-GAT CGG ATC CTA ATA CGA CTC ACT ATA GGC GAC ATC CAA TCC ATA TCA AGG A-3´. Reverse primer: 5´-GAT CAA GCT TGT TCA GAG TCT AGT TTA TTC TCA CAC-3´). RNA samples were electrophoresed in TBE-Urea PAGE 12% and radioactivity detected by phosphoimaging. Sxl-RBD4, endogenous HuR or ACTB in HEK293/Sxl-RBD4 spiked-in samples were detected by Western blot with specific antibodies.

RNaseA Digestion and Electrophoretic Mobility Assay Crosslinked and non-crosslinked HEK293 cell suspensions were subjected to PTex as described before. Resulting pellets were solubilised in 150 µl of buffer TED (20 mM Tris, 1 mM EDTA, 0.03% DDM [n-Dodecyl -D-maltoside]) at 56°C during 20 min, followed by incubation at 75°C during 20 min. Samples were mixed with RNaseA (2 ng) and incubated at 37°C; aliquots of 20 L were taken at different time points: 0, 1, 5, 10, 30 and 60 min, immediately mixed with 5 µL of 6x Laemmli buffer and heated at 95°C for 5 min before SDS-PAGE.

In-vitro transcription of Sxl-RBD4 target RNAs The T7 promoter and a sxl-target DNA sequence 5´-GAT CCG GTC ATA GGT GTA AAA AAA GTC TCC ATT CCT ATA GTG AGT CGT ATT AA-3´ was cloned into pUC19 using the restriction enzymes BamHI and HindIII and the resulting plasmid was named pUC19-sxl-target. The 30 nt RNA was synthesised as described before (Beckmann et al., 2010) by hybridising two complementary sequences containing the T7 RNApolymerase promoter 5´-TAA TAC GAC TCA CTA TAG-3´ and the template sequence 5´-GGT CAT AGG TGT AAA AAA ACT CTC CAT TCC TAT AGT GAG TCG TAT TAA-3´, followed by T7 run-off transcription. Templates for RNAs of 87 and 191 nt length were generated using DNA restriction fragments from the pUC19-sxl-target plasmid (HindIII+EcoRI - 87 bp; HindIII+PvuI - 191 bp). T7 RNA polymerase and restriction enzymes were purchased from

New England Biolabs. Plasmids were purified using the NucleoBond Xtra Midi kit (Macherey-Nagel, 740410.100), DNA fragments by NucleoSpin Gel and PCR clean-up (Macherey-Nagel, 740609.50) and RNA by acidic phenol extraction (Chomczynski and Sacchi, 1987). RNA 5'-GAG UUU UUU UAC A-3' (13 nt) was synthesised by Biomers (Ulm, Germany).

In vitro cross-linking assays 48 µg of Sxl-RBD4 in 100 µl crosslinking buffer (CLB: 10 mM Tris pH 7.4, 50 mM KCl, 1 mM EDTA, 1 mM DDT) were mixed with *in vitro*-transcribed RNA (13, 30, 87 and 191 nt, final concentration 500 nM) harbouring one copy of the target motif 5'-GAG UUU UUU UAC A-3', incubated at 4°C for 30 min and cross-linked with 0.25 J/cm2 of UV-254 nm, on ice. Afterwards, 80% of each sample was used for PTex extraction, while 20% of the sample were kept as input control and used for SDS-PAGE and western blotting to detect Sxl-RBD4.

Western blotting Western blotting was performed using standard techniques. Samples were electrophoreses on SDS-PAGE gradient gels 4-20% (TGX stain free, BioRad) and proteins transferred onto a nitrocellulose membranes 0.2µ (BioRad). Membranes were blocked during 30 min with PBST-M (10mM phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, pH 7.4 0.1% tween 20 (Sigma), 5% milk) and incubated with dilutions 1:1000 of anti-HuR (proteintech, 11910-1-AP 0.42 µg/µL) ACTB (proteintech, 66009-1-Ig 0.13 µg/µL), anti-FLAG (Sigma, F1804 1µg/µL) or Sxl-RBD4 (DHSB, antiSxl hybridome culture supernatant M114 1:20) overnight at 4° C (or 2h room temperature). Monoclonal mouse anti-Sxl antibodies (M18 and M114) were developed by P. Schedl and obtained from the Developmental Studies Hybridoma Bank (DHSB, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA52242). Antibody binding was detected using the respective anti-mouseHRP or anti-rabbitHRP secondary antibodies (Proteintech) and Clarity ECL Western Blotting Substrate for chemiluminescence in a ChemiDocMP imaging system (BioRad).

Immunoprecipitation and PNK assay Immunoprecipitation of bacterial FLAG-tagged proteins and radioactive labeling of RNA by PNK was performed as described by Holmqvist et al. (2016).

Protein purification Recombinant Sxl-RBD4 protein (Sxl amino acids 122-301) was purified essentially as described before (Moschall et al., 2018). In brief, after IPTG induction for 4h at 23°C in *E. coli* (BL21Star [Invitrogen] transformed with the Rosetta 2 plasmid [Merck]), cells were lysed in a buffer containing 20 mM Tris/Cl pH 7.5, 1M NaCl, 0.2 mM EDTA, 1 mM DTT, cOmplete Protease Inhibitor Cocktail [Roche] followed by centrifugation for 20min at 12,000 x g. The cleared lysate was then subjected to GSH-affinity chromatography using an ÄKTA FPLC system. Bound protein was eluted in a buffer containing 100 mM HEPES/KOH pH 8.0, 50 mM glutathione, 50 mM KCl, and 1 mM DTT. Fractions containing the protein were supplemented with 3C protease and dialyzed overnight against IEX buffer (20 mM HEPES/KOH pH 8.0, 50 mM KCl, 10% Glycerol, 0.2 mM EDTA, 0.01% NP-40) followed by ion exchange chromatography using a MonoS column. Fractions containing the pure protein were pooled, dialyzed against storage buffer (20 mM HEPES/KOH pH 8.0, 20% Glycerol, 0.2 mM EDTA, 0.01% NP-40, 1 mM DTT) and stored at 80°C.

MS sample preparation Human HEK293 and *Salmonella* cells were cultivated as described above and pellets used in PTex protocol. A minor fraction of initial input material was lysed and proteins denatured in 1% SDS and 0.1 M DTT Phosphate Buffer Solution (PBS) by boiling for 10 min at 95°C. After cooling, Benzonase was added for 30 min at 37°C before cell lysates were spun down and supernatants transferred to fresh tubes. Remaining input material was used for RBPs enrichment with PTex as described above. After PTex, RBPs were precipitated in 90% ethanol solution at -20°C and 30 minutes centrifugation at 20,000 x g at 4°C. Protein pellets were ressuspended in 2 M urea in 50mM ammonium bicarbonate (ABC) buffer and Benzonase was added for 30 min at 37°C to remove RNA. For silver staining, protein samples were directly loaded and separated on a pre-casted SDS-PAGE gel. Gels were fixed in 30% ethanol, 15% acetic acid solution in MiliQ water and incubated for 1 hour in 500 mM sodium acetate, 12 mM

sodium thiosulfate, 0.125% glutaraldehyde, 25% ethanol solution. Gels were washed 3 times with MiliQ water for 10 minutes and stained with 0.1% Silver nitrate, 0.011% formaldehyde in MiliQ water solution for 30 minutes. Finally, gels were briefly rinsed with MiliQ water and reaction developed in 240 mM sodium carbonate, 0.01% Formaldehyde in MiliQ water solution. Reaction was stopped by addition of 50 mM EDTA.

For mass spectrometry analysis, proteins were precipitated with methanol-chloroform extraction (Wessel and Flugge 1984) and ressuspended in 8 M urea and 0.1 M Tris pH 8 solution. Proteins were reduced with 10 mM DTT at room temperature for 30 min and alkylated with 55 mM iodoacetamide at room temperature for 30 min in the dark. Proteins were first digested by lysyl endopeptidase (LysC) (Wako) at a LysC-to-protein ratio of 50:1 (w/w) at room temperature for 3 h. Afterwards, samples were diluted to 2 M final concentration of urea with 50 mM ammonium bicarbonate. Trypsin (Promega) digestion was performed at a trypsin-to-protein ratio of 50:1 (w/w) under constant agitation at room temperature for 16 h. Peptides were desalted with C18 Stage Tips (Rappsilber et al., 2003) prior to LC-MS/MS analysis. Peptides were separated on a monolithic column (100 µm ID x 2,000 mm, MonoCap C18 High Resolution 2000 [GL Sciences] kindly provided by Dr. Yasushi Ishihama [Kyoto University]) using 6 hour gradient of increasing acetonitrile concentration at a flow rate of 300 nl/min, or on an in-house made C18 15cm microcolumns (75 µm ID packed with ReproSil-Pur C18-AQ 3-µm resin, Dr. Maisch GmbH) using 2 or 4 hours gradient of 5 to 50% increasing acetonitrile concentration at a flow rate of 200 nl/min. The Q Exactive instrument (Thermo Fisher Scientific) was operated in the data dependent mode with a full scan in the Orbitrap followed by top 10 MS/MS scans using higher-energy collision dissociation (HCD). All raw files were analyzed with MaxQuant software (v1.5.1.2) (Cox and Mann, 2008) using the label free quantification (LFQ) algorithm (Cox et al., 2014) with default parameters and match between runs option on. Database search was performed against the human reference proteome (UNIPROT version 2014-10, downloaded in October 2014) or the Salmonella Typhimurium reference proteome (UNIPROT version 2017, downloaded in August 2017) with common contaminants. False discovery rate (FDR) was set to 1% at peptide and protein levels. The mass spectrometry proteomics data (HEK293) have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org)

via the PRIDE partner repository with the dataset identifier PXD009571.

Bioinformatic analysis of PTex-purified proteins (HEK293) The complete analysis is available as R notebook (Suppl_MS-HEK293_analysis.pdf). In short, we used LFQ MS intensities normalised to trypsin (which is constant in all samples). Potential contaminants, reverse and peptides only identified by modification were excluded from analysis. Fold changes were calculated by subtraction of the log2 values of LFQ intensity for proteins from UV cross-linked samples and non-cross-linked samples. Only proteins which were found in all replicates were processed further. Enrichment (CL/-CL) was calculated as described before (Beckmann, 2017): P-values were calculated from an Ebayes moderated t-test using the limma package (Ritchie et al., 2015) followed by Benjamini-Hochberg False Discovery Rate (FDR) correction. Only proteins with an adjusted p-value of 0.01 or smaller in all 3 cross-linking intensities were considered being enriched. GO analysis was performed using PANTHER V.11 (Mi et al., 2017). Since GO terms can be very broad terms (e.g. "localization" (GO:0051179)) or just represent small specialised groups (e.g. "small protein activating enzyme binding" (GO:0044388)) with 10 or less protein members, we ignored GO terms with total protein members in the lower and upper decile. Domain enrichment was done using DAVID (Huang et al., 2009) searching the SMART (Letunic and Bork, 2018) database.

Bioinformatic analysis of PTex-purified proteins (Salmonella) The complete analysis is available as R notebook (Suppl_MS-Salmonella_analysis.pdf). We used iBAQ-normalised values for the *Salmonella* analysis. Potential contaminants, reverse and peptides only identified by modification were excluded from analysis. Fold changes were calculated by subtraction of the log2 values of iBAQ intensity for proteins from UV cross-linked samples and non-cross-linked samples. Only proteins which were found in both replicates were taken into account (258 proteins; 172 with a log2 fold-change >0). Domain and GO terms were analysed using DAVID (Huang et al., 2009).

RNASeq sample preparation Cross-linked PTex samples and non-cross-linked control (CL-, 0.015, 0.15, 1.5 J/cm^2) were proteinase K digested (1 h, 56°C) and the RNA recovered by acidic phenol extraction (Chomczynski and Sacchi, 1987) using phase lock gel tubes (5Prime, 2302830). Libraries were created according to the "TruSeq Stranded Total RNA LT" protocol (Ilumina, 15032612) with the modification that we skipped the rRNA depletion step. We used adapters AR002,4,5-7,12,13-16,18,19. DNA concentration was determined by Qubit 3.0 Fluorometer (Life Technologies) and the quality of libraries assessed by a Bioanalyzer 2100, DNA 1000 Chip Kit (Agilent, 5067-1504). Sequencing was performed on a Illumina HiSeq4000.

Bioinformatic analysis of PTex-purified RNA RNASeq data were quality controled using fastqc (v0.11.2). We then mapped obtained reads against a single copies of human rRNA and tRNA sequences using bowtie2 (v2.2.6):

```
bowtie2 --no-unal --un $LIBRARY_rRNA_not_aligned_reads.fastq --al
    $LIBRARY_rRNA_aligned_reads.fastq -x rRNA_db -U $LIBRARY.fastq >
    $LIBRARY_rRNA.sam 2>> $LIBRARY_rRNA_alignment_stats.txt
```

We then used the remaining reads and mapped reads to the human genome (GRCh38.p12) and the corresponding comprehensive gene annotation file (gencode.v28.chr_patch_hapl_scaff.annotation.gtf) from GENCODE using the STAR aligner (v020201):

```
STAR --genomeDir PATH/TO/index --readFilesIn
    $LIBRARY_rRNA_not_aligned_reads.fastq --quantMode GeneCounts
```

After aligning the remaining reads with STAR, samtools calmd was used to annotate each read in the bam files with mutation information. Using the SAM.py parser from the pyCRAC suite Webb et al. (2014), chromosomal locations of substitutions and deletions were extracted and counted. Only mutations that were unique to the UV-treated samples were considered. To normalize the data for sequencing depth, for each dataset the counts for substitutions and deletions were divided by the total number of mapped nucleotides, which provided an indication of mutation frequencies. To map the distribution of deletions and substitutions around AUG and Stop codons, CDS coordinates from the gencode.v28.chr_patch_hapl_scaff.annotation.gtf

annotation files were extracted. Tables containing counts and chromosomal positions for each substitution and deletion were converted into gene transfer format (GTF) files using the GTF2 and NGSFormatWriters classes from the pyCRAC package. Subsequently, pyBinCollector from the pyCRAC package was used to map the distribution of substitutions and deletions around start and stop codons of protein-coding genes:

```
pyBinCollector.py -f mutations.gtf --gtf annotationfile_CDS_coordinates.gtf
   -s 5end -a protein_coding --normalize -v -o dist_around_AUG.txt
pyBinCollector.py -f mutations.gtf --gtf annotationfile_CDS_coordinates.gtf
   -s 3end -a protein_coding --normalize -v -o dist_around_STOP.txt
```

For the feature counts, count tables generated by STAR were used in conjunction with the gencode.v28.chr_patch_hapl_scaff.annotation.gtf annotation file. PTex RNA-seq data have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE113655.

PAR-CLIP and pCLIP We performed the PAR-CLIP protocol as described by Hafner et al. (2010); Maatz et al. (2017): HEK293 cells stably expressing FLAG-tagged HuR (ELAVL1), were grown until 90% confluence. The last 16 h of incubation, 200 mM 4SU was added. Living cells were irradiated with 0.15 J/cm^2 365 nm UV light, snap-frozen on dry ice and stored at -80°C until use. Cells ($\sim$ 1.2 x 10^8 cells/replicate) were lysed on ice for 10 min with 3 ml lysis buffer (50 mM Tris-Cl pH 7.5 (Life Tech., 15567027), 100 mM NaCl (Life Tech. AM9760G), 1% (v/v) Nonidet P40 substitute (Sigma 74385), 0.5% (v/v) Sodium deoxycholate (AppliChem A1531) containing 0.04 U/ml RNasin (Promega, N2515) and 2x Complete Protease Inhibitor (Roche, 11697498001) and centrifuged 20,000 x g, 10 min, 4°C. Cleared lysates (1.5 mL/replicate) were digested with 8 U/mL TURBO DNase (ThermoFisher, AM2238) and 2 U/µL RNase I (ThermoFisher, AM2294) at 37°C for 4 min (replicate 1) or 3 min 15 sec (replicates 2 and 3). FLAG-tagged HuR was immunoprecipitated with 10 µg of anti-FLAG monoclonal antibody (Sigma, F1804) bound to 40 µl of Protein G Dynabeads (Life Tech, 10004D). After extensive washes with high-salt buffer (50mM Tris-HCL pH 7.5, 1mM EDTA, 1% Nonidet P40, 0.1% (v/v) SDS, 1 M NaCl) beads were incubated with 1 U/µl of T4 polynucleotide kinase (T4-PNK,

NEB) and 0.5 µCi/µL of ^{32}P-ATP. After radiolabelling, samples were splitted into 4 tubes and underwent three versions of CLIP:

PAR-CLIP classic Briefly, cIHuR-RNA complexes were resolved by Nu-PAGE MOPS (invitrogen) transferred to nitrocellulose and excised at a defined size-range (50 to 60 kDa). Proteins were digested from the membrane with proteinase K and the RNA recovered by acidic phenol/chloroform extraction and ethanol precipitation. RNA was ligated with 3' adapter (5'App-NNN NTG GAA TTC TCG GGT GCC AAG G-3'InvdT) gel-slice isolated, ligated with 5´adapter (5'-GUU CAG AGU UCU ACA GUC CGA CGA UCN NNN-3') and purified again from PAGE-Urea gels by elution and ethanol precipitation.

PAR-CLIP on-beads As described in (Benhalevy et al., 2017) on-beads 3´ adapter ligation was done by incubating the 3´adapter with Rnl2(1–249)K227Q ligase, in presence of PEG-8000 overnight at 4°C. After washes, the 5´adapter was ligated using the Rnl1 enzyme, 2 h at 37°C.

pCLIP Adapter ligation was performed as described above (Benhalevy et al., 2017). Immediately after the 5´- adapter ligation, beads were resuspended in 600 µl of solution D, incubated at 95°C during 10 min and beads were separated with a magnet. cIHuR-RNA complexes in the supernatant were purified applying the last two steps of the PTex protocol. Ethanol-precipitated pellets were digested with proteinase K and RNA isolated by phenol/chloroform extraction.
All the ligated RNAs were retro-transcribed into cDNA with the reverse transcription primer 5'-GCC TTG GCA CCC GAG AAT TCC A-3'. After the PCR cycles were determined for each case, cDNA libraries were created by PCR using the forward primer 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA-3' and the Ilumina adapter index RPI 1-6, 8,10-11. Bands obtained at 150 bp were excised from 2 % agarose gels and purified using the Zymoclean Gel Recovery Kit (Zymo, D4002). DNA concentration and library quality was determined by Qubit Fluorometer dsDNA HS assay (Life Tech, Q32854) and BioAnalizer DNA HS Kit (Agilent 2100 Bioanalyzer; Agilent, 5067-4626). Libraries were sequenced in a NextSeq 500. PARCLIP RNA-seq data have been submitted to the NCBI Gene Expression

Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE113628.

CLIP data processing The PAR-CLIP data was processed and annotated using the PARpipe pipeline (https://github.com/ohlerlab/PARpipe) around the PAR-CLIP data tailored peak caller PARalyzer (Corcoran et al., 2011) as described previously (Mukherjee et al., 2014), with one modification. In brief, adapter sequences were trimmed retaining the four randomized adapter nucleotides on both read ends included during adapter ligation to serve as unique molecular identifiers during PCR-duplicate removal (read collapsing). Differences in the numbers of uniquely aligning reads were balanced by random subsampling prior to PARalyzer cluster identification. Annotation of identified binding sites (cluster) was simplified by grouping closely related sub-annotation categories (Table S3).

De novo motif discovery For de novo motif finding we used Zagros (Bahrami-Samani et al., 2015) with default settings including RNA secondary structure information. As the majority of HuR cluster resided in intronic and 3'utr regions (70-80%, Table S4) according to its reported functions, we use intronic and 3'utr cluster sequences as input.

Transcriptomic metacoverage For depicting spatial preferences for mRNA binding, we selected genes previously used for RNA classification based on processing and turnover dynamics (Mukherjee et al. 2016) (n=15120) present in GENCODE v19. To select transcripts, we ran RSEM (Li B. et al., Bioinformatics 2011) and retained transcripts with TPM >3. For each gene, we selected the transcript isoform with the highest isoform percentage or chose one randomly in case of ties (n=8298). The list of selected transcript isoforms was used to calculate the median 5'UTR, CDS and 3'UTR length proportions (5'UTR=0.06, CDS=0.53, 3'UTR=0.41) using R Bioconductor packages GenomicFeatures and GenomicRanges (Lawrence et al., 2013). For regions post annotated transcription ends and splice sites we chose windows of fixed sizes (TES 500 nt, 5' and 3' splice sites 250 nt each). We generated coverage tracks from the PARalyzer output alignment files and intersected those with the filtered transcripts. Each annotation category was binned according to its relative length and the coverage averaged within each bin.

For intronic coverage, we averaged across all introns per gene, given a minimal intron length of 500 nt. All bins were stitched to one continuous track per transcript (n=6632 intron containing transcripts). Each library bam file was filtered to retain only PARalyzer cluster overlapping alignments. We required transcripts to have a minimal coverage maximum of >2. For each transcript we scaled the binned coverage dividing by its maximal coverage (min-to-1 scaling) to emphasize on spatial patterns independent from transcript expression levels. Next, we split transcript coverage in two parts, separating 5'UTR to TES regions and intronic regions. To generate the scaled meta coverage across all targeted transcripts per RBP, we used the heatMeta function from the Genomation package (Akalin et al., 2015). For the 5'UTR to TES part, we scaled each RBP meta-coverage track independent of other libraries. For intronic sequences, we scaled each sample relative to all other sample. Finally, we clustered the meta-coverage tracks using ward.D clustering with euclidean distance.

Genome browser vizualizations PAR-CLIP alignments were visualized using Gviz (Hahne and Ivanek, 2016).

Acknowledgements

The authors wish to thank Florian Heyd and Marco Preussner for providing mouse brain samples, and Jörg Vogel for providing Hfq-FLAG-tagged *Salmonella*. We are grateful to Maja Köhn for discussions and to Markus Landthaler for critically reading the manuscript. Work in the Beckmann lab is supprted by the German Research Foundation (DFG; IRTG 2290 and ZUK 75/1 Project 0190-854599). Work in the Medenbach lab is supported by the DFG (SFB 960/2, B11), the Bavarian State Ministry for Education, Science and the Arts (Bavarian Research Network for Molecular Biosystems, BioSysNet), and the German Federal Ministry of Education and Research (BMBF, 01ZX1401D). The Granneman lab is funded by a Medical Research Council non-clinical Senior Research fellowship (MR/R008205/1). The Ohler lab is funded by NIH TRA R01 GM104962.

References

Akalin, A., Franke, V., Vlahoviček, K., Mason, C. E. and Schübeler, D. (2015), Genomation: a toolkit to summarize, annotate and visualize genomic intervals., *Bioinformatics (Oxford, England)* **31**(7), 1127–1129.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25417204*

Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. and Tuschl, T. (2012), Identification of rna-protein interaction networks using par-clip., *Wiley interdisciplinary reviews. RNA* **3**(2), 159–177.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22213601*

Bahrami-Samani, E., Penalva, L. O. F., Smith, A. D. and Uren, P. J. (2015), Leveraging cross-link modification events in clip-seq for motif discovery., *Nucleic acids research* **43**(1), 95–103.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25505146*

Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C. and Landthaler, M. (2012), The mrna-bound proteome and its global occupancy profile on protein-coding transcripts., *Molecular cell* **46**(5), 674–690.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22681889*

Ban, N., Beckmann, R., Cate, J. H. D., Dinman, J. D., Dragon, F., Ellis, S. R., Lafontaine, D. L. J., Lindahl, L., Liljas, A., Lipton, J. M., McAlear, M. A., Moore, P. B., Noller, H. F., Ortega, J., Panse, V. G., Ramakrishnan, V., Spahn, C. M. T., Steitz, T. A., Tchorzewski, M., Tollervey, D., Warren, A. J., Williamson, J. R., Wilson, D., Yonath, A. and Yusupov, M. (2014), A new system for naming ribosomal proteins., *Current opinion in structural biology* **24**, 165–169.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24524803*

Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., Wang, X., Ji, L., Khan, M. J., Zhu, X., Luo, Z., Shao, C., Lim, D.-H., Liu,

X., Li, N., Wang, W., He, M., Liu, Y.-L., Ward, C., Wang, T., Zhang, G., Wang, D., Yang, J., Chen, Y., Zhang, C., Jauch, R., Yang, Y.-G., Wang, Y., Qin, B., Anko, M.-L., Hutchins, A. P., Sun, H., Wang, H., Fu, X.-D., Zhang, B. and Esteban, M. A. (2018), Capturing the interactome of newly transcribed rna., *Nature methods* **15**(3), 213–220.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/29431736*

Beckmann, B. M. (2017), Rna interactome capture in yeast., *Methods (San Diego, Calif.)* **118-119**, 82–92.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27993706*

Beckmann, B. M., Castello, A. and Medenbach, J. (2016), The expanding universe of ribonucleoproteins: of novel rna-binding proteins and unconventional interactions., *Pflugers Archiv : European journal of physiology* **468**(6), 1029–1040.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27165283*

Beckmann, B. M., Grünweller, A., Weber, M. H. W. and Hartmann, R. K. (2010), Northern blot detection of endogenous small rnas (approximately14 nt) in bacterial total rna extracts., *Nucleic acids research* **38**(14), e147.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/20504856*

Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.-M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., Krijgsveld, J. and Hentze, M. W. (2015), The rna-binding proteomes from yeast to man harbour conserved enigmrbps., *Nature communications* **6**, 10127.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/26632259*

Benhalevy, D., McFarland, H. L., Sarshad, A. A. and Hafner, M. (2017), Par-clip and streamlined small rna cdna library preparation protocol for the identification of rna binding protein target sites., *Methods (San Diego, Calif.)* **118-119**, 41–49.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27871973*

Brannan, K. W., Jin, W., Huelga, S. C., Banks, C. A. S., Gilmore, J. M., Florens, L., Washburn,

M. P., Van Nostrand, E. L., Pratt, G. A., Schwinn, M. K., Daniels, D. L. and Yeo, G. W. (2016), Sonar discovers rna-binding proteins from analysis of large-scale protein-protein interactomes., *Molecular cell* **64**(2), 282–293.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27720645*

Brimacombe, R., Stiege, W., Kyriatsoulis, A. and Maly, P. (1988), Intra-rna and rna-protein cross-linking techniques in escherichia coli ribosomes., *Methods in enzymology* **164**, 287–309.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/3071669*

Brooks, S. A. (2010), Functional interactions between mrna turnover and surveillance and the ubiquitin proteasome system., *Wiley interdisciplinary reviews. RNA* **1**(2), 240–252.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21935888*

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J. and Hentze, M. W. (2012), Insights into rna biology from an atlas of mammalian mrna-binding proteins., *Cell* **149**(6), 1393–1406.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22658674*

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., *Molecular cell* **63**(4), 696–710.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27453046*

Chey, S., Claus, C. and Liebert, U. G. (2011), Improved method for simultaneous isolation of proteins and nucleic acids., *Analytical biochemistry* **411**(1), 164–166.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21094121*

Chomczynski, P. and Sacchi, N. (1987), Single-step method of rna isolation by acid guanidinium thiocyanate-phenol-chloroform extraction., *Analytical biochemistry* **162**(1), 156–159.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2440339*

Conrad, T., Albrecht, A.-S., de Melo Costa, V. R., Sauer, S., Meierhofer, D. and Ørom, U. A. (2016), Serial interactome capture of the human cell nucleus., *Nature communications* **7**, 11212.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27040163*

Corcoran, D. L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R. L., Keene, J. D. and Ohler, U. (2011), Paralyzer: definition of rna binding sites from par-clip short-read sequence data., *Genome biology* **12**(8), R79.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21851591*

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N. and Mann, M. (2014), Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq., *Molecular & cellular proteomics : MCP* **13**(9), 2513–2526.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24942700*

Cox, J. and Mann, M. (2008), Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification., *Nature biotechnology* **26**(12), 1367–1372.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19029910*

Delan-Forino, C., Schneider, C. and Tollervey, D. (2017), Rna substrate length as an indicator of exosome interactions., *Wellcome open research* **2**, 34.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28748221*

Fátyol, K. and Grummt, I. (2008), Proteasomal atpases are associated with rdna: the ubiquitin proteasome system plays a direct role in rna polymerase i transcription., *Biochimica et biophysica acta* **1779**(12), 850–859.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/18804559*

Favre, A., Moreno, G., Blondel, M. O., Kliber, J., Vinzens, F. and Salet, C. (1986), 4-thiouridine photosensitized rna-protein crosslinking in mammalian cells., *Biochemical and biophysical*

research communications **141**(2), 847–854.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2432896*

Freeberg, M. A., Han, T., Moresco, J. J., Kong, A., Yang, Y.-C., Lu, Z. J., Yates, J. R. and Kim, J. K. (2013), Pervasive and dynamic protein binding sites of the mrna transcriptome in saccharomyces cerevisiae., *Genome biology* **14**(2), R13.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23409723*

Gehring, N. H., Wahle, E. and Fischer, U. (2017), Deciphering the mrnp code: Rna-bound determinants of post-transcriptional gene regulation., *Trends in biochemical sciences* **42**(5), 369–382.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28268044*

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., *Molecular & cellular proteomics : MCP* **11**(3), M111.014050.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22278370*

Gerstberger, S., Hafner, M. and Tuschl, T. (2014), A census of human rna-binding proteins., *Nature reviews. Genetics* **15**(12), 829–845.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25365966*

Gorski, S. A., Vogel, J. and Doudna, J. A. (2017), Rna-based recognition and targeting: sowing the seeds of specificity., *Nature reviews. Molecular cell biology* **18**(4), 215–228.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28196981*

Granneman, S., Kudla, G., Petfalski, E. and Tollervey, D. (2009), Identification of protein binding sites on u3 snorna and pre-rrna by uv cross-linking and high-throughput analysis of cdnas., *Proceedings of the National Academy of Sciences of the United States of America* **106**(24), 9613–9618.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19482942*

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M. and Tuschl, T. (2010), Transcriptome-wide identification of rna-binding protein and microrna target sites by par-clip., *Cell* **141**(1), 129–141.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/20371350*

Hahne, F. and Ivanek, R. (2016), Visualizing genomic data using gviz and bioconductor., *Methods in molecular biology (Clifton, N.J.)* **1418**, 335–351.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27008022*

Halbach, F., Reichelt, P., Rode, M. and Conti, E. (2013), The yeast ski complex: crystal structure and rna channeling to the exosome complex., *Cell* **154**(4), 814–826.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23953113*

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., *Molecular cell* **64**(2), 416–430.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27768875*

Hentze, M. W., Castello, A., Schwarzl, T. and Preiss, T. (2018), A brave new world of rna-binding proteins., *Nature reviews. Molecular cell biology* .
URL: *http://www.ncbi.nlm.nih.gov/pubmed/29339797*

Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R. and von Hippel, P. H. (1986), Laser cross-linking of nucleic acids to proteins. methodology and first applications to the phage t4 dna replication system., *The Journal of biological chemistry* **261**(8), 3512–3518.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/3949776*

Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. and Brown, P. O. (2008), Diverse rna-binding proteins interact with functionally related sets of rnas, suggesting an extensive regulatory system., *PLoS biology* **6**(10), e255.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/18959479*

Holmqvist, E., Wright, P. R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., Backofen, R. and Vogel, J. (2016), Global rna recognition patterns of post-transcriptional regulators hfq and csra revealed by uv crosslinking in?vivo., *The EMBO journal* **35**(9), 991–1011.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27044921*

Horsch, A., Martins de Sa, C., Dineva, B., Spindler, E. and Schmid, H. P. (1989), Prosomes discriminate between mrna of adenovirus-infected and uninfected hela cells., *FEBS letters* **246**(1-2), 131–136.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/2565252*

Houseley, J. and Tollervey, D. (2009), The many pathways of rna degradation., *Cell* **136**(4), 763–776.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19239894*

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009), Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists., *Nucleic acids research* **37**(1), 1–13.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19033363*

Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.-B., Munier, A., Fradet, M., Daunesse, M., Bertrand, E., Pierron, G., Mozziconacci, J., Kress, M. and Weil, D. (2017), P-body purification reveals the condensation of repressed mrna regulons., *Molecular cell* **68**(1), 144–157.e5.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28965817*

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M., König, J. and Ule, J. (2014), iclip: protein-rna interactions at nucleotide resolution., *Methods (San Diego, Calif.)* **65**(3), 274–287.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24184352*

Hwang, J. and Inouye, M. (2010), A bacterial gap-like protein, yihi, regulating the gtpase of der, an essential gtp-binding protein in escherichia coli., *Journal of molecular biology* **399**(5), 759–

772.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/20434458*

Ingolia, N. T., Lareau, L. F. and Weissman, J. S. (2011), Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes., *Cell* **147**(4), 789–802.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/22056041*

Jin, W., Wang, Y., Liu, C.-P., Yang, N., Jin, M., Cong, Y., Wang, M. and Xu, R.-M. (2016), Structural basis for snrna recognition by the double-wd40 repeat domain of gemin5., *Genes & development* **30**(21), 2391–2403.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27881601*

Keene, J. D. (2007), Rna regulons: coordination of post-transcriptional events., *Nature reviews. Genetics* **8**(7), 533–543.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/17572691*

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M. and Zavolan, M. (2011), A quantitative analysis of clip methods for identifying binding sites of rna-binding proteins., *Nature methods* **8**(7), 559–564.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21572407*

Kowalinski, E., Kögel, A., Ebert, J., Reichelt, P., Stegmann, E., Habermann, B. and Conti, E. (2016), Structure of a cytoplasmic 11-subunit rna exosome complex., *Molecular cell* **63**(1), 125–134.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27345150*

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., *Nature methods* **11**(10), 1064–1070.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25173706*

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T. and Carey, V. J. (2013), Software for computing and annotating genomic ranges., *PLoS computational biology* **9**(8), e1003118.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23950696*

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M. and Rajewsky, N. (2011), Transcriptome-wide analysis of regulatory interactions of the rna-binding protein hur., *Molecular cell* **43**(3), 340–352.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21723171*

Letunic, I. and Bork, P. (2018), 20 years of the smart protein domain annotation resource., *Nucleic acids research* **46**(D1), D493–D496.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/29040681*

Lorentzen, E., Basquin, J. and Conti, E. (2008), Structural organization of the rna-degrading exosome., *Current opinion in structural biology* **18**(6), 709–713.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/18955140*

Maatz, H., Kolinski, M., Hubner, N. and Landthaler, M. (2017), Transcriptome-wide identification of rna-binding protein binding sites using photoactivatable-ribonucleoside-enhanced crosslinking immunoprecipitation (par-clip)., *Current protocols in molecular biology* **118**, 27.6.1–27.6.19.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28369676*

Makino, D. L., Halbach, F. and Conti, E. (2013), The rna exosome and proteasome: common principles of degradation control., *Nature reviews. Molecular cell biology* **14**(10), 654–660.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23989960*

Matia-González, A. M., Laing, E. E. and Gerber, A. P. (2015), Conserved mrna-binding proteomes in eukaryotic organisms., *Nature structural & molecular biology* **22**(12), 1027–1033.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/26595419*

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. and Thomas, P. D. (2017), Panther version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements., *Nucleic acids research* **45**(D1), D183–D189.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27899595*

Michaux, C., Holmqvist, E., Vasicek, E., Sharan, M., Barquist, L., Westermann, A. J., Gunn, J. S. and Vogel, J. (2017), Rna target profiles direct the discovery of virulence functions for the cold-shock proteins cspc and cspe., *Proceedings of the National Academy of Sciences of the United States of America* **114**(26), 6824–6829.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28611217*

Milek, M., Imami, K., Mukherjee, N., Bortoli, F. D., Zinnall, U., Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., Selbach, M. and Landthaler, M. (2017), Ddx54 regulates transcriptome dynamics during dna damage response., *Genome research* **27**(8), 1344–1359.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28596291*

Moschall, R., Strauss, D., García-Beyaert, M., Gebauer, F. and Medenbach, J. (2018), Sister-of-sex-lethal is a repressor of translation., *RNA (New York, N.Y.)* **24**(2), 149–158.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/29089381*

Mukherjee, N., Corcoran, D. L., Nusbaum, J. D., Reid, D. W., Georgiev, S., Hafner, M., Ascano, M., Tuschl, T., Ohler, U. and Keene, J. D. (2011), Integrative regulatory mapping indicates that the rna-binding protein hur couples pre-mrna processing and mrna stability., *Molecular cell* **43**(3), 327–339.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21723170*

Mukherjee, N., Jacobs, N. C., Hafner, M., Kennington, E. A., Nusbaum, J. D., Tuschl, T., Blackshear, P. J. and Ohler, U. (2014), Global target mrna specification and regulation by the rna-binding protein zfp36., *Genome biology* **15**(1), R12.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24401661*

Mullari, M., Lyon, D., Jensen, L. J. and Nielsen, M. L. (2017), Specifying rna-binding regions

in proteins by peptide cross-linking and affinity purification., *Journal of proteome research* **16**(8), 2762–2772.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28648085*

Petit, F., Jarrousse, A. S., Dahlmann, B., Sobek, A., Hendil, K. B., Buri, J., Briand, Y. and Schmid, H. P. (1997), Involvement of proteasomal subunits zeta and iota in rna degradation., *The Biochemical journal* **326 (Pt 1)**, 93–98.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/9337855*

Rappsilber, J., Ishihama, Y. and Mann, M. (2003), Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and lc/ms sample pretreatment in proteomics., *Analytical chemistry* **75**(3), 663–670.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/12585499*

Rispal, D., Henri, J., van Tilbeurgh, H., Graille, M. and Séraphin, B. (2011), Structural and functional analysis of nro1/ett1: a protein involved in translation termination in s. cerevisiae and in o2-mediated gene control in s. pombe., *RNA (New York, N.Y.)* **17**(7), 1213–1224.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21610214*

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015), limma powers differential expression analyses for rna-sequencing and microarray studies., *Nucleic acids research* **43**(7), e47.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25605792*

Schueler, M., Munschauer, M., Gregersen, L. H., Finzel, A., Loewer, A., Chen, W., Landthaler, M. and Dieterich, C. (2014), Differential protein occupancy profiling of the mrna transcriptome., *Genome biology* **15**(1), R15.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24417896*

Schuller, J. M., Falk, S., Fromm, L., Hurt, E. and Conti, E. (2018), Structure of the nuclear exosome captured on a maturing preribosome., *Science (New York, N.Y.)* .
URL: *http://www.ncbi.nlm.nih.gov/pubmed/29519915*

Singh, R., Valcárcel, J. and Green, M. R. (1995), Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins., *Science (New York, N.Y.)* **268**(5214), 1173–1176.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/7761834*

Smirnov, A., Förstner, K. U., Holmqvist, E., Otto, A., Günster, R., Becher, D., Reinhardt, R. and Vogel, J. (2016), Grad-seq guides the discovery of proq as a major small rna-binding protein., *Proceedings of the National Academy of Sciences of the United States of America* **113**(41), 11591–11596.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27671629*

Tawk, C., Sharan, M., Eulalio, A. and Vogel, J. (2017), A systematic analysis of the rna-targeting potential of secreted bacterial effector proteins., *Scientific reports* **7**(1), 9328.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28839189*

Tsatsaronis, J. A., Franch-Arroyo, S., Resch, U. and Charpentier, E. (2018), Extracellular vesicle rna: A universal mediator of microbial communication?, *Trends in microbiology* **26**(5), 401–410.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/29548832*

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R. B. (2003), Clip identifies nova-regulated rna networks in the brain., *Science (New York, N.Y.)* **302**(5648), 1212–1215.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/14615540*

Uzzau, S., Figueroa-Bossi, N., Rubino, S. and Bossi, L. (2001), Epitope tagging of chromosomal genes in salmonella., *Proceedings of the National Academy of Sciences of the United States of America* **98**(26), 15264–15269.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/11742086*

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., Stanton, R., Rigo, F., Guttman, M. and Yeo, G. W. (2016), Robust transcriptome-wide discovery of rna-binding

protein binding sites with enhanced clip (eclip)., *Nature methods* **13**(6), 508–514.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/27018577*

van Nues, R., Schweikert, G., de Leau, E., Selega, A., Langford, A., Franklin, R., Iosub, I., Wadsworth, P., Sanguinetti, G. and Granneman, S. (2017), Kinetic crac uncovers a role for nab3 in determining gene expression profiles during stress., *Nature communications* **8**(1), 12.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/28400552*

Wang, Q., Hobbs, K., Lynn, B. and Rymond, B. C. (2003), The clf1p splicing factor promotes spliceosome assembly through n-terminal tetratricopeptide repeat contacts., *The Journal of biological chemistry* **278**(10), 7875–7883.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/12509417*

Wasmuth, E. V., Januszyk, K. and Lima, C. D. (2014), Structure of an rrp6-rna exosome complex bound to poly(a) rna., *Nature* **511**(7510), 435–439.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25043052*

Webb, S., Hector, R. D., Kudla, G. and Granneman, S. (2014), Par-clip data indicate that nrd1-nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast., *Genome biology* **15**(1), R8.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/24393166*

Wegrecki, M., Marcin, W., Neira, J. L. and Bravo, J. (2015), The carboxy-terminal domain of erb1 is a seven-bladed ß-propeller that binds rna., *PloS one* **10**(4), e0123463.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/25880847*

Zhang, C. and Darnell, R. B. (2011), Mapping in vivo protein-rna interactions at single-nucleotide resolution from hits-clip data., *Nature biotechnology* **29**(7), 607–614.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21633356*

Zinder, J. C., Wasmuth, E. V. and Lima, C. D. (2016), Nuclear rna exosome at 3.1?å reveals substrate specificities, rna paths, and allosteric inhibition of rrp44/dis3., *Molecular cell*

64(4), 734–745.

URL: *http://www.ncbi.nlm.nih.gov/pubmed/27818140*