

ARTICLE

<https://doi.org/10.1038/s41467-019-08942-3>

OPEN

# Purification of cross-linked RNA-protein complexes by phenol-toluol extraction

Erika C. Urdaneta<sup>1</sup>, Carlos H. Vieira-Vieira<sup>2</sup>, Timon Hick<sup>1</sup>, Hans-Herrmann Wessels<sup>3,4</sup>, Davide Figini<sup>1</sup>, Rebecca Moschall<sup>5</sup>, Jan Medenbach<sup>5</sup>, Uwe Ohler <sup>3,4</sup>, Sander Granneman<sup>6</sup>, Matthias Selbach <sup>2</sup> & Benedikt M. Beckmann <sup>1</sup>

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 archaea and bacteria). We have developed the Phenol Toluol extraction (PTex) protocol 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 global RNA-bound proteome of human HEK293 cells and the bacterium *Salmonella Typhimurium*.

<sup>1</sup>IRI Life Sciences, Humboldt University, Philippstr. 13, 10115 Berlin, Germany. <sup>2</sup>Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany. <sup>3</sup>Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany. <sup>4</sup>Department of Biology, Humboldt University, Philippstr. 13, 10115 Berlin, Germany. <sup>5</sup>Biochemistry I, University of Regensburg, Universitätsstrasse 31, 93053 Regensburg, Germany. <sup>6</sup>Centre for Systems and Synthetic Biology (SynthSys), University of Edinburgh, Max Born Crescent, Edinburgh EH9 3BF, UK. Correspondence and requests for materials should be addressed to B.M.B. (email: [benedikt.beckmann@iri-lifesciences.de](mailto:benedikt.beckmann@iri-lifesciences.de))

**R**NA-binding proteins are key factors in the post transcriptional regulation of gene expression. Spurred by recent technological advances such as RNA interactome capture (RIC)<sup>1–3</sup>, the number of RBPs has 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)<sup>4–6</sup>. 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 (MS). This led to the discovery of hundreds of hitherto unknown RBPs<sup>7,8</sup>. (ii) Cross-linking and immunoprecipitation (CLIP) 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<sup>9–11</sup>.

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<sup>12</sup> and CARIC<sup>13</sup>) has been introduced in which labelled RNA along with UV cross-linked interacting proteins was purified in a poly-A-independent fashion. However, the method requires efficient *in vivo* labelling 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<sup>14</sup>, also marketed as “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 (acidic conditions shown in Fig. 1b, right panel).

Here, we describe a 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. Following the rationale that phenol and toluol (toluene) share a similar chemical structure but toluol is less water soluble due to the lack of the OH group (Fig. 1b), we modified the extraction chemistry using a mixture of phenol:toluol. This alters enrichment of the biomolecule classes in the extraction and furthermore enabled us to shift cross-linked RNPs into the aqueous or interphase, respectively (Fig. 1b, left panel). Combining both separation strategies 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 (Phenol Toluol extraction)-based purification of cross-linked RNPs<sup>2</sup>. 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
- Step 2: Phenol, pH 4.8, (chaotropic) detergents
- Step 3: Phenol, EtOH, water, pH 4.8

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<sup>14</sup>. Now, free RNA accumulates in the upper aqueous phase, free proteins in the lower organic phase and cLRNPs migrate to the interphase<sup>4</sup> (Fig. 1b right panel). Finally, the complexes in the interphase are precipitated using ethanol<sup>15</sup>. 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<sup>16</sup> (Fig. 1c). UV-cross-linking produces an additional band at high molecular weight which indicates the RNA-cross-linked fraction of HuR (cHuR; Supplementary Fig. 1). In the PTex fraction (interphase 3), cHuR was highly enriched whereas free HuR was strongly reduced. 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-RBD<sup>4</sup><sup>17</sup>), 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 <sup>32</sup>P-5' labelled 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) or plasmid DNA (pUC19). DNA is removed during the first two PTex steps (Fig. 1e). We then tested for additional well-established RBPs (Fig. 1f), namely polypyrimidine tract binding protein 1 (PTB1), fused in sarcoma (FUS), and the more recently identified RNA-binding enzymes glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and enolase (Eno1);<sup>3</sup> all of which are enriched by PTex in a UV-irradiation-dependent manner whereas the highly abundant DNA-binding histone H3 is depleted. We additionally used PTB1 and FUS to demonstrate that RNase digestion prior to PTex abolishes efficient enrichment of RBPs (Fig. 1g, Supplementary Fig. 9), consistent with selectivity for RBPs complexed with RNA. In sum, PTex highly enriches for cross-linked RNPs while efficiently depleting non-cross-linked proteins or nucleic acids.

**PTex performance.** We then set out to critically assess the performance of PTex to purify RBPs. However, while cross-linked HuR generates a signal at the height of the gel pocket (compare Fig. 1c and Supplementary Fig. 1), we have no a priori knowledge about the amount of protein which is efficiently cross-linked by UV light to RNA. Thus, we performed RNA interactome capture<sup>1</sup> on HEK293 cells to obtain proteins which are 100% cross-linked to poly(A) RNA. We used this sample as input for PTex (Fig. 2a) and quantified purification of HuR by densitometry (Fig. 2b), resulting in a ~30% recovery of the cross-linked protein input. 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 with Uracil stretches of 7 nucleotides or longer<sup>18</sup>. We produced *in vitro* transcribed RNAs with lengths varying between 13 and



**Fig. 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 (cRNPs) 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 against HuR (ELAVL1, 35 kDa) demonstrates that UV-cross-linking-stabilised HuR-RNA complexes (upper edge/gel pocket of the blot) are largely enriched after PTex (step 3 interphase). A purified fly protein (Sxl RBD4) served as spike-in as 100% non-cross-linked RBP. **d** 5'-end radioactive-labelled RNA was subjected to PTex in vitro. **e** PCR with specific primers against exon 5 of the interleukin 3 (IL3) gene demonstrates efficient removal of genomic DNA after either full HEK293 cells (upper panel) or pre-purified genomic DNA (middle panel) were subjected to PTex. A PCR product derived from linear pUC19 DNA (lower panel) is also removed. **f** Enrichment of known RBPs by PTex tested by western-blot against PTBP1, FUS, or against non-classical RNA-binding enzymes Eno1 and GAPDH. Note that RNaseA treatment was performed after PTex as it removes partially shifted bands (smear) for some RBPs. **g** PTex enriches for cross-linked RBPs. RNase treatment before PTex strongly reduces recovery of known RNA-binders (PTBP1, FUS). Non-RBP controls Histone H3 and actin (ACTB) are efficiently depleted by PTex (**c**, **f**, **g**). For full gels/blots see Supplementary Figures 1–8

Finally, we estimated the relative enrichment of free vs. cross-linked protein in input and PTex (Fig. 2f, g). For HuR (in vivo), PTex largely enriches for RNA-bound HuR with a background of 10–15% non-UV-cross-linked HuR. For Sxl, PTex-mediated enrichment is even larger if compared to the input; however, free Sxl is more prominently detected in the artificial in vitro conditions only but not when spiked-in to cell lysates (Fig. 1c) where the free protein is removed almost completely. Interestingly, we only recover non-cross-linked proteins in the PTex fraction when interrogating RNA-binding proteins but not in the case of other abundant cellular proteins such as histone H3 or beta-actin (see Fig. 1c–g). We attribute this to either stable RNA-protein complexes which resist complete denaturation and separation during the PTex procedure or an artefact of the gel system as RNA strand breaks after PTex purification could result in a western blot signal akin to the free/unbound RBP. In both cases, only bona-fide RBPs will be enriched.

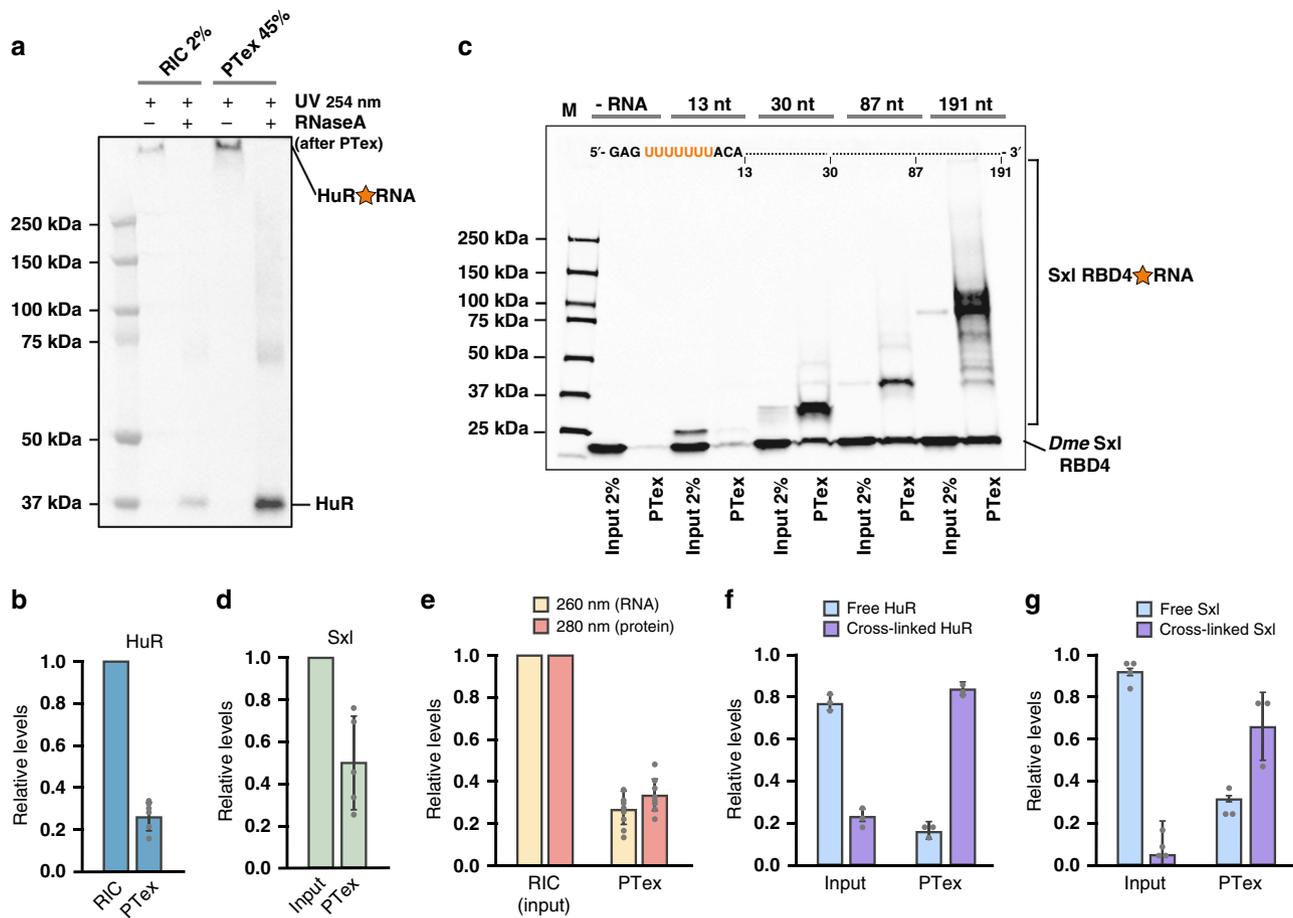
**Purification of RBPs from animal tissue.** To test if PTex can be directly applied to tissues, we UV cross-linked whole mouse brain samples and performed PTex to extract cRNPs directly (Fig. 3). Since brain tissue is particularly rich in lipids which accumulate in the interphase during step 1 of the PTex protocol, we increased the temperature during extractions to 65 °C (Hot-PTex). The extracted cRNPs were analysed by western blotting. Cross-linked HuR is largely enriched over non cross-linked HuR after PTex (Fig. 3 lower panel). Weak detectable bands at 70 or ~110 kDa, respectively, are likely HuR dimers and trimers as observed before<sup>19</sup>. 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 labelling (such as PAR-CLIP<sup>11</sup>, RICK<sup>12</sup>, or CARIC<sup>13</sup>) for which efficient uptake and incorporation of nucleotide analogues can be challenging. Brain tissue can be efficiently UV cross-linked, however, as demonstrated by HITS-CLIP experiments<sup>20</sup>.

**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')<sup>16</sup>. After in vivo labelling of cellular RNA using 4-thiouridine (4SU) and UV irradiation at 365 nm, we performed (i) classical PAR-CLIP analysis (PAR-CLIP-classic)<sup>11</sup> of HuR, (ii) a PAR-CLIP variant using on-bead ligation of adaptors (PAR-CLIP-on-beads)<sup>21,22</sup>, and (iii) a version in which we use phenol extraction (termed pCLIP) for removal of unbound RNA instead of PAGE/membrane excision (Fig. 4a). We found that pCLIP libraries contained a larger fraction of longer reads than the PAR-CLIP classic/PAR-CLIP-

on-beads libraries (Fig. 4b). 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 in all three variants (Fig. 4c–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 (Fig. 4f, g). 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 cRNPs 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<sup>1–3</sup> targets only polyadenylated RNAs suggests that many RBPs that bind non-adenylated RNAs are missed by this experimental approach<sup>12,13</sup> 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<sup>2</sup> irradiation at 254 nm wavelength also irradiation at two additional energy levels: 0.015 and 1.5 J/cm<sup>2</sup>; spanning two orders 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<sup>2,8</sup>. We independently irradiated HEK293 cells at all 3 energies and performed PTex purification of cross-linked and non-cross-linked RNPs from whole cells (Fig. 5a). Transcriptomics by RNA-Seq (Fig. 5b, c), and proteomics by MS and label-free quantification (LFQ) (Fig. 5d–g) were performed using total RNA and protein preparations as input controls.

**RNA in PTex-purified RNPs.** 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<sup>23,24</sup>. In line with this, we find a similar distribution of RNA classes when comparing RNA from inputs and PTex by RNA-Seq with the vast majority of transcripts being ribosomal RNA (Supplementary Fig. 21). Protein-cross-linked RNA is known to enrich for mutations during reverse transcription in the RNA-Seq workflow (Fig. 5b, Supplementary Fig. 17)<sup>10,11</sup>. Such mutations can then be used as beacons to map



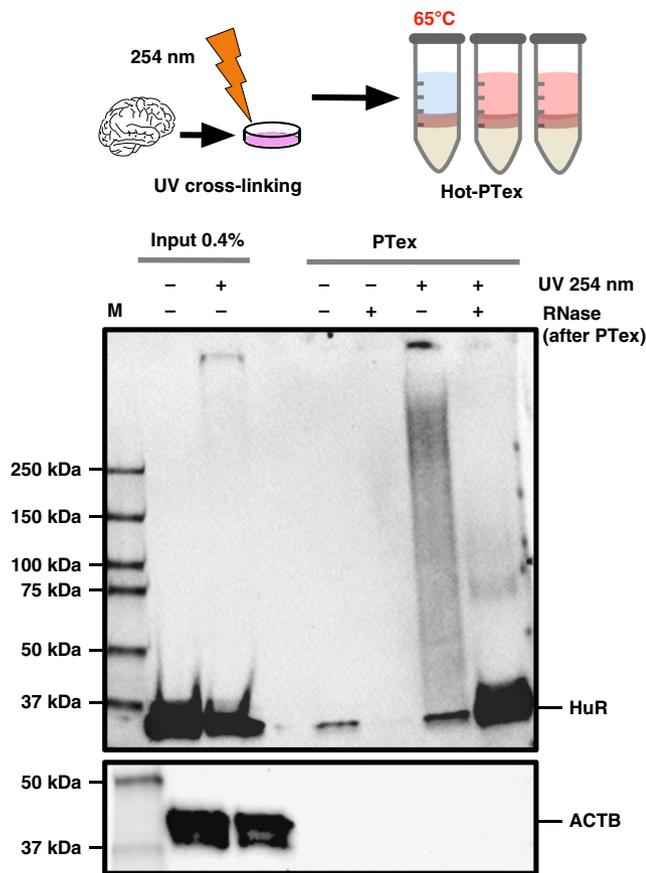
**Fig. 2** Performance of PText. **a** RNA interactome capture of HEK293 was used as 100% cross-linked input material and HuR recovery by PText was assessed. **b** Quantification of **a** (HuR  $n = 6$  biologically independent experiments). **c** *Drosophila melanogaster* RBP Sxl RBD4 was bound to RNA of several length carrying the Sxl U7 recognition site and UV cross-linked in vitro; minimal RNA length for efficient recovery by PText is 30 nt. **d** Quantification of **c** ( $n = 5$  biologically independent experiments). **e** PText recovery of RNA (260 nm) and proteins (280 nm) determined by UV spectroscopy ( $n = 9$ ; 5 biologically independent experiments; 2 of those with 3 technical repeats each). **f, g** Relative enrichment of cross-linked HuR (source Fig. 1c) and Sxl (source Fig. 2c,  $n = 3$  biologically independent experiments) by PText. Error bars represent SD. All data are in Supplementary Data 4. For full blots see Supplementary Figures 10–14

protein binding sites in transcripts<sup>2,11</sup>. We used pyCRAC<sup>25</sup> to map deletion and substitution mutations enriched in UV-treated PText samples to 100 nt windows around the 5' AUG start codon and 3' stop codon of mRNA reads (Fig. 5c, Supplementary Fig. 18). 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<sup>26</sup> and can potentially be attributed to cross-linking of ribosomal proteins or translation initiation/termination factors to mRNA, as ribosomal profiling experiments show increased ribosomal footprint densities at these regions, indicating longer dwell times and a higher potential for cross-linking at these sites<sup>27</sup>.

**Proteins in PText-purified RNPs.** Proteins which were not identified by MS in all 3 replicates after PText were removed. For the remaining proteins, ratios of cross-linked over non cross-linked (CL/-CL) LFQ intensities (from the PText experiments) were calculated (Fig. 5d). *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. 5e, Supplementary Data 1). Analysis of general protein features (molecular mass, pI, cellular abundance, hydrophobicity, Fig. 5f, g) demonstrates that the PText procedure does not enrich for a particular subgroup of cellular proteins based on chemical properties or expression level. We picked two of the 3037 proteins which have not been reported to bind RNA: ATP-binding cassette sub-family F member 2 (ABCF2), a member of the AAA + ATPase family (see below) and T-complex protein 1 subunit eta (CCCT7) as part of the chaperonin CCT/TRiC which is involved in telomere maintenance<sup>28</sup>. Both are enriched after UV-irradiation in vivo and PText purification (Fig. 5h, i), indicating RNA-association.

In principle, extended UV exposure should increase the chance for cross-linking events and thus for subsequent protein recovery by PText<sup>3</sup>. We therefore expected a gradual increase in enrichment of RNA-interacting proteins after PText 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<sup>2</sup> on the other hand leads to a significant decrease in recovery (Supplementary Fig. 23). We observed RNA degradation when analysing total RNA from HEK293 cells irradiated with 1.5 J/cm<sup>2</sup> 254 nm UV light (Supplementary Figs. 20, 21;<sup>3</sup>). If this degradation was due to damage of nucleic acids induced by high energy UV light or a result of secondary



**Fig. 3** RNP purification from animal tissue. 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 Supplementary Figure 15

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 refs. <sup>29</sup> and <sup>8</sup>) and PTex-purified protein patterns differ from the whole proteome in silver stains (Supplementary Fig. 22). However, 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<sup>30</sup>, we find nearly a third of the expressed cellular proteome to be associated with RNA.

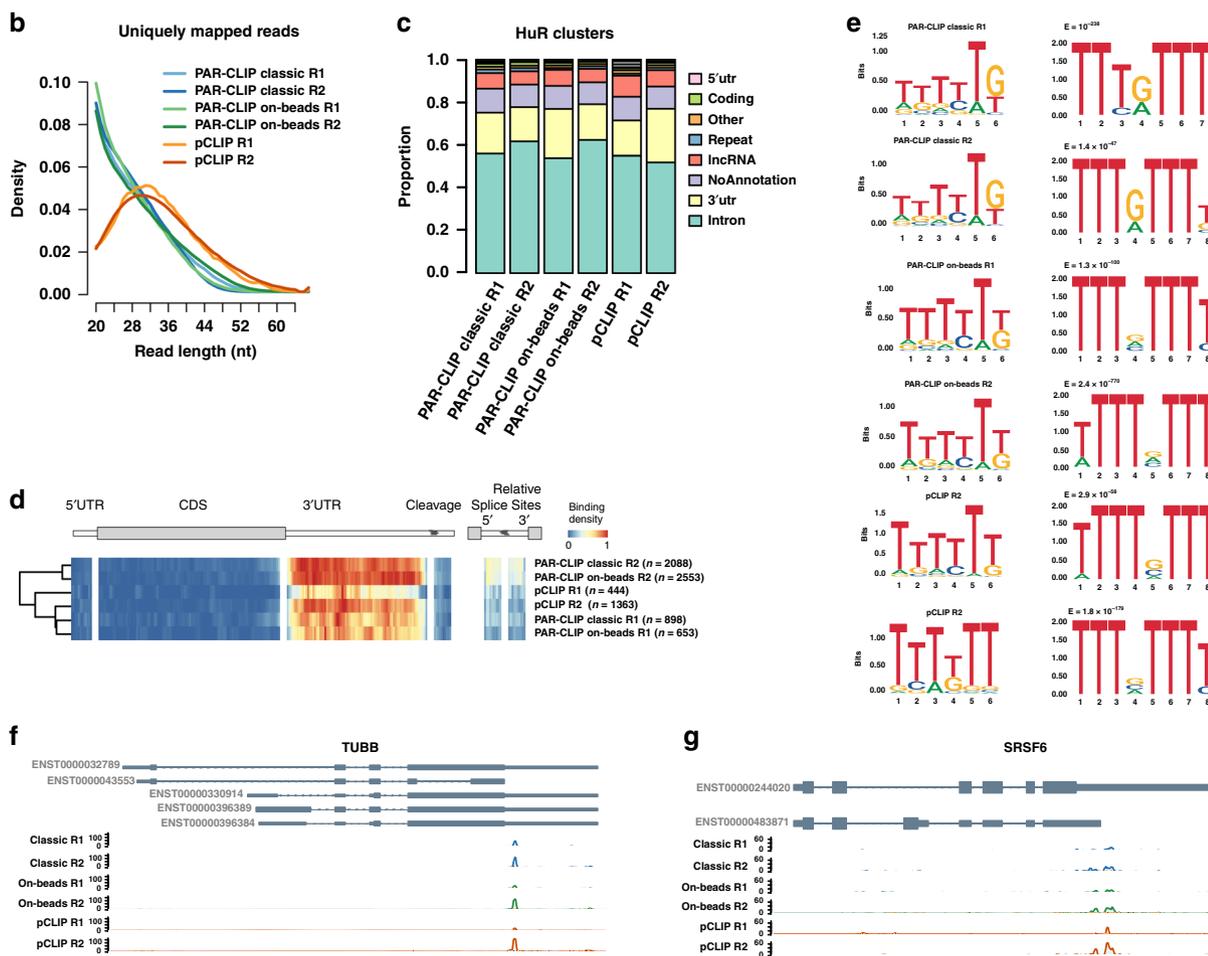
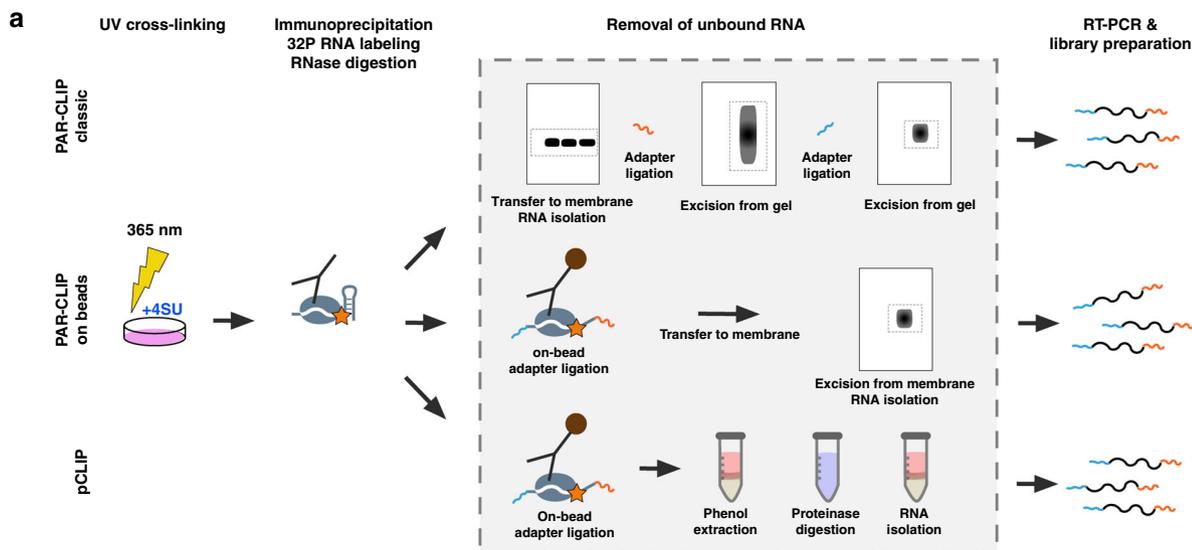
To test sensitivity and specificity of our approach, we first performed global GO enrichment analysis showing that terms from all aspects of RNA biology are the most enriched among PTex-purified proteins (Fig. 6a). At the same time, protein classes with no general role in RNA biology such as transporters and (trans-)membrane proteins were depleted by PTex (Supplementary Data 2). Known RNA-binding domains (RBDs) such as the RNA recognition motif (RRM), helicase folds (DEXDC, HELICc) or K homology (KH) domain were significantly enriched among PTex-purified proteins (Fig. 6b; Supplementary Data 2). 89 PTex-enriched proteins contain a WD40 fold; a domain found to directly bind snRNA in Gemin 5<sup>31</sup>, taking part in rRNA biogenesis (Erb1<sup>32</sup>) and found in RBPs<sup>1</sup>. Other enriched domains are: AAA (ATPase, see below) fold, tetratricopeptide repeat region (TPR) as found in the yeast Clf1p splicing factor<sup>33</sup>, Ski

complex<sup>34</sup> and the translation terminator Nro1<sup>35</sup>, and the CH domain which is found among actin-binding proteins<sup>3</sup>.

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;<sup>36</sup>  $p$ -value:  $1.36 \times 10^{-30}$ ), NSUN2 and tRNA synthetases (19/20 cytosolic;  $p$ -value:  $7.4 \times 10^{-10}$ ) 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 ref. <sup>2</sup> (Fig. 6c) although different UV irradiation strategies were used (254 vs. 365 nm cross-linking;  $p$ -value:  $1.26 \times 10^{-139}$ ). Importantly, the largest overlap was with RBPs recently found in HeLa cells using the also unbiased RICK technique<sup>12</sup> (94% for high confidence RBPs and 86% of non-poly-A RBPs, respectively) and the CARIC approach (84% overlap)<sup>13</sup> (Fig. 6c). Recent studies show that the boundaries between RNA- and DNA-binding are rather blurry and nuclear DNA-binders were found to interact with RNA<sup>37–40</sup>. We identified proteins involved in replication and response to DNA damage (Supplementary Data 2) such as DDX54<sup>38</sup> but in general, DNA-binders such as transcription factors were under-represented, demonstrating that PTex does not select for DNA-specific binding proteins in particular (Fig. 6c). 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<sup>2</sup> in the PTex enrichment. The established RBPs are similarly enriched along the dynamic range (from no enrichment to  $\log_2\text{FC PTex [CL/-CL]}$  of 6; Fig. 6d) of PTex and hence display no difference to the novel RNA-interactors.

The presented results demonstrate that PTex is specific for RNPs. But why were not the same proteins discovered to associate with RNA before? The majority of recently discovered RBPs are interacting with mRNA;<sup>8</sup> 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<sup>41</sup>. The core exosome complex consists of 10 protein subunits (Exo-10) from which only one protein (Rrp44) is catalytically acting on RNA as exo- and endoribonuclease (Fig. 6f). 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<sup>41,42</sup>. Still, all 10 subunits are positioned to directly interact with RNA and multiple interactions with the individual subunits have been demonstrated in high-resolution structure studies<sup>43,44</sup>. Hence, all 10 subunits are amenable to UV cross-linking and, as a result, 9 out of the 10 subunits were identified by PTex (Fig. 6g).

Previously it was demonstrated that interactome capture enriches for mRNA-binding proteins with a high isoelectric point (pI; Fig. 6e)<sup>1,3</sup>. However, sequence- and oligo-dT-independent approaches such as RICK<sup>12</sup> 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 (Supplementary Fig. 24). 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;<sup>45</sup> the former being of sufficient length for efficient recovery by PTex (Fig. 2c). In sum, PTex enriches for the (near) complete exosome core complex while interactome capture from the same cell line only found a single subunit<sup>2</sup> (Fig. 6g).



Yet we also enrich for proteins which have no established role in RNA biology such as subunits of the human proteasome (Supplementary Data 2). However, from the 28 proteins of the 20S core complex, Psma5 and Psma6 were reported to display RNase activity in purified complexes<sup>46</sup>. Importantly, proteasome-associated RNAs were shown to lack poly-A stretches<sup>47</sup> and ATPases of the AAA family in the 19S proteasome regulatory particle were found to be recruited to RNA polymerase I (rRNA) transcription sites<sup>48</sup>. 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<sup>41</sup> and RNAs interacting with Psma5/6 are likely to be UV-cross-linked to other subunits as well.

**A 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

**Fig. 4** A fast PAR-CLIP variant employing phenolic extraction (pCLIP). **a** Schematic comparison of PAR-CLIP variants. **b** Read length distribution of uniquely mapping reads utilised 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** We applied de novo motif discovery for PARalyzer derived clusters using ZAGROS (left) and DREME (right). For Zagros, we found a T-rich motif scoring the highest in all cases. As ZAGROS does not return E-values we analysed the cluster sequences using DREME. For all but classic PAR-CLIP R2 we found a T-rich motif scoring the highest. For classic PAR-CLIP R2 however, the T-rich motif scored second with a similar E-value to a less frequent primary motif (Supplementary Fig. 16). **f, g** Genome browser shots of TUBB and SRSF6 example genes showing reproducible 3'UTR binding sites. Track y-axes represent uniquely mapping read count

protein<sup>49</sup> to test PTex in bacteria. Hfq is an abundant RNA-binder facilitating mRNA:ncRNA interactions in Gram-negatives<sup>49</sup>. As for animal tissue culture (see above), we used the slightly modified PTex protocol (Hot-PTex) in which RNP extraction from *Salmonella* grown to OD<sub>600</sub> 3.0 was performed at 65 °C thereby supporting cell lysis (Fig. 6a). Hfq is a 17-kDa protein and forms a homo-hexamers in bacterial cells; the complex has been shown to resist normal Laemmli PAGE conditions<sup>49</sup> 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 MS (Fig. 6c). Comparing recovered protein intensities from UV-irradiated versus control cells (biological duplicates), we find 172 proteins (Supplementary Data 3), among them 33 ribosomal proteins, components of the RNA polymerase complex (subunit  $\alpha$ ,  $\sigma$  factor RpoD, DksA) and 4 out of the 5 established mRNA-binding proteins of *Salmonella* (Hfq, ProQ and CspC/CspE)<sup>49–51</sup>. One hundred and thirteen 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<sup>52</sup>, the alkyl hydroperoxide reductase c22 protein (AhpC)<sup>53</sup> and the cell invasion protein SipA<sup>54</sup>. Using mutant strains carrying FLAG-tags fused to the C-terminus of PTex candidate RNA-interactors, we performed UV cross-linking, immunoprecipitation and radioactive labelling of co-purified RNA (known as PNK assay as described in ref. <sup>55</sup>), validating that YihI, SipA and AhpC are indeed associated to RNA in vivo (Fig. 6d, Supplementary Figs. 25, 26). In addition, we recently validated PTex-enriched ClpX and DnaJ as well<sup>56</sup>. 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 and NusA) and domains which were also detected in RBPs when screening eukaryotic cells: the aforementioned AAA ATPase fold in the ATP-dependent protease ATPase subunits HslU, ClpB, and ClpX, or thioredoxin domains as in AhpC, Thiol:disulphide interchange protein (DsbA) and Bacterioferritin comigratory protein (Bcp)<sup>1,3</sup>. As in many other species, we find glycolytic enzymes to associate to RNA (Pgi, Pgi)<sup>3,57</sup>. 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 (Supplementary Data 3). 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 ref. <sup>58</sup>). It is tempting to speculate if the here identified proteins are involved in such a process or are mere bystander proteins.

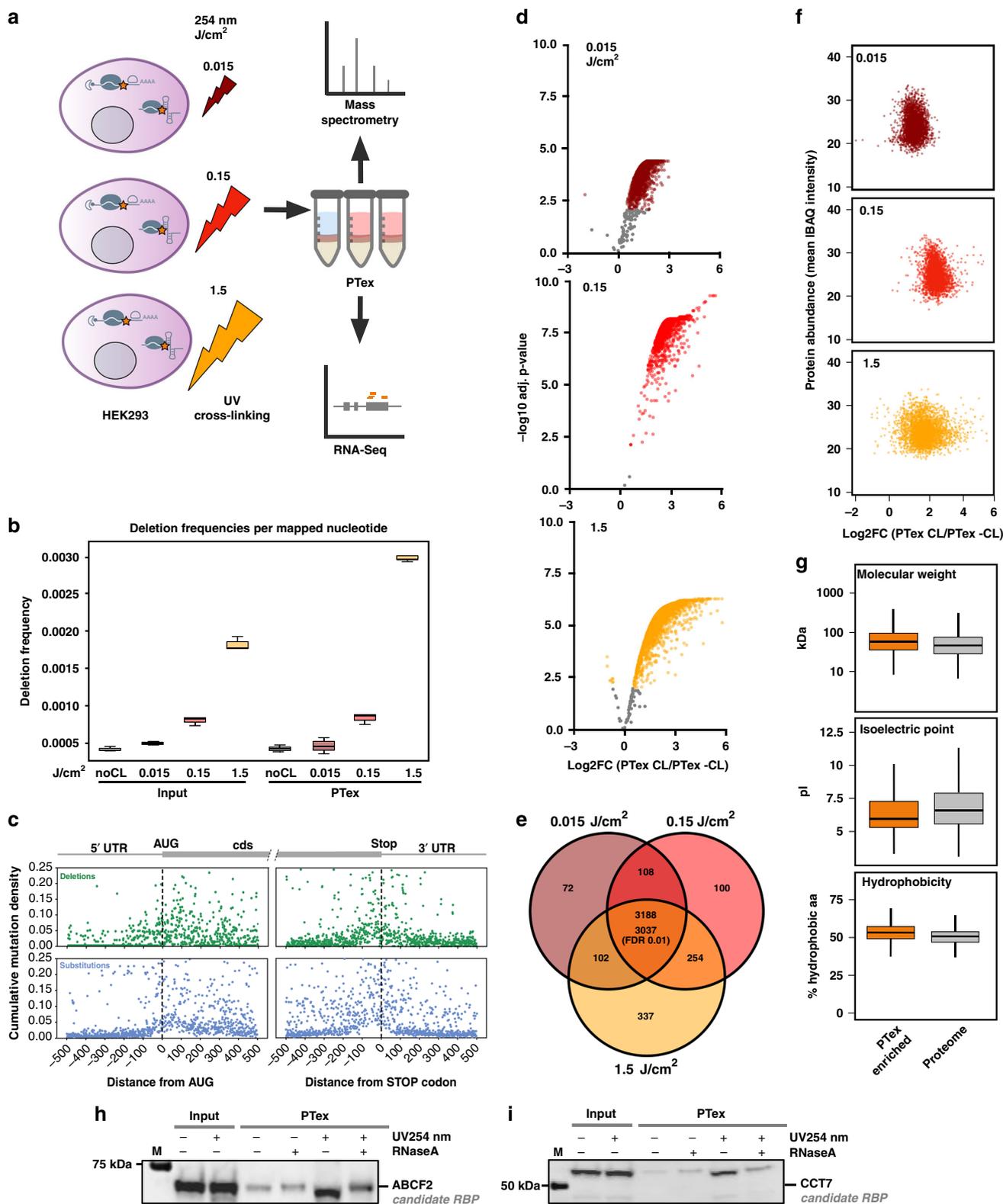
Overall, we noticed that enrichments even from known RBPs were lower in *Salmonella* compared to HEK293 cells and we anticipate that additional modifications in UV cross-linking and/or cell lysis could improve sensitivity when applying (Hot-)PTex in bacteria. PTex is the first approach that can purify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

## 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<sup>1,4</sup> and yeast<sup>3,57,59</sup>. 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<sup>14</sup>, 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 h. 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. Cross-linked proteins of 30–50% could be recovered by PTex compared to the starting material (Fig. 2b–e). At the same time, PTex drastically improves the relative enrichment of cross-linked over free protein (Fig. 2f, g).

Next to PTex, RBR-ID<sup>60</sup>, RICK<sup>12</sup> and CARIC<sup>13</sup> are methods which allow for unbiased purification of RNPs. However, there are also differences: All, RBR-ID, RICK and CARIC rely on RNA labelling using either 4SU or 5-ethynyl uridine (EU) and subsequent UV irradiation. Using pulse/chase experiments, these methods thus allow to determine newly transcribed RNA species. For PTex, UV irradiation at 254 nm wavelength is sufficient (no labelling required) and we anticipate that this is an advantage



when analysing biological material in which uptake of nucleotide analogues is either insufficient or too cost-intensive. It did not escape our attention however that PText requires fewer input material compared to the other methods (Table 1). Since all of these methods rely on UV irradiation, however, biases introduced by different cross-linking efficiencies for individual proteins remain a general issue. During the revision of this manuscript, two related protocols were published termed OOPS<sup>61</sup> and

XRNAX<sup>62</sup>. Both also exploit organic biphasic extraction as principle to purify RNPs.

Our work provides a cell-wide analysis of the effects of different UV irradiation dosage on RNA-protein cross-linking (Supplementary Data 1, Supplementary Fig. 23). 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<sup>2</sup> 254 nm light (Supplementary

**Fig. 5** A global snapshot of RNPs in HEK293 cells. **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<sup>2</sup> 254 nm light in triplicates. Total RNA from input (whole-cell lysate) and PTex-purified samples were analysed by RNA-Seq. **b** Deletions in RNA from input and PTex samples; frequency of mutations in transcripts correlate with higher UV doses. Boxplot centre line represents median, bounds are first and third quantile, and whiskers extend to 1.5 times the inter-quantile range ( $n = 3$  biologically independent experiments except for input 0.015 J/cm<sup>2</sup> ( $n = 2$ )). **c** 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 Supplementary Fig. 17, 18. **d–g** Input (whole-cell lysate) and PTex-purified sample were analysed by label-free mass spectrometry. **d** Volcano plots of proteins enriched by PTex (FDR 0.01) under the three cross-linking conditions. **e** Overlap of PTex-enriched proteins (enriched in all 9 replicates, FDR 0.01) is 3037 (these PTex proteins are from here on coloured in orange). **f** Protein abundance (IBAQ intensities of input samples) does not correlate with PTex enrichment (log<sub>2</sub>-fold change of intensities [CL/-CL]). **g** PTex does not select for a subset of proteins based on general features such as molecular weight, pI or hydrophobicity. Boxplot centre line represents median, bounds are first and third quantile, and whiskers extend to 1.5 times the inter-quantile range. **h, i** PTex of individual predicted RNA-associated proteins. ATP-binding cassette sub-family F member 2 (ABC2) and T-complex protein 1 subunit eta (CCT7) have not been reported to bind RNA. Both are enriched after PTex in a UV-irradiation-dependent fashion, indicating that they indeed associate with RNA in vivo. For full blots see Supplementary Figure 19

Fig. 23) demonstrates that extensive irradiation/cross-linking can have adverse effects on protein recovery. Next to the observed RNA degradation, cross-linked peptides released by tryptic digestion are notoriously difficult to identify in MS experiments and hence increasing the amount of protein cross-linking might negatively impact protein identification<sup>63</sup>. 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. 5b, c)<sup>9–11</sup>. Our results demonstrate that when investigating RNPs on a global scale, the cross-linking 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 MS. Conversely, almost all RNA can be expected to be bound by a set of proteins under physiological conditions<sup>23,24</sup> which explains why we also observe an increase in mutations in input RNA (Fig. 4b, Supplementary Figs. 17, 18). 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 RNA-Seq workflows already and cross-linked RNA will not be lost during cDNA preparation or sequencing<sup>9–11</sup>.

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 display 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 - archaea and prokaryotes - could not be investigated for technical reasons. We here provide a 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

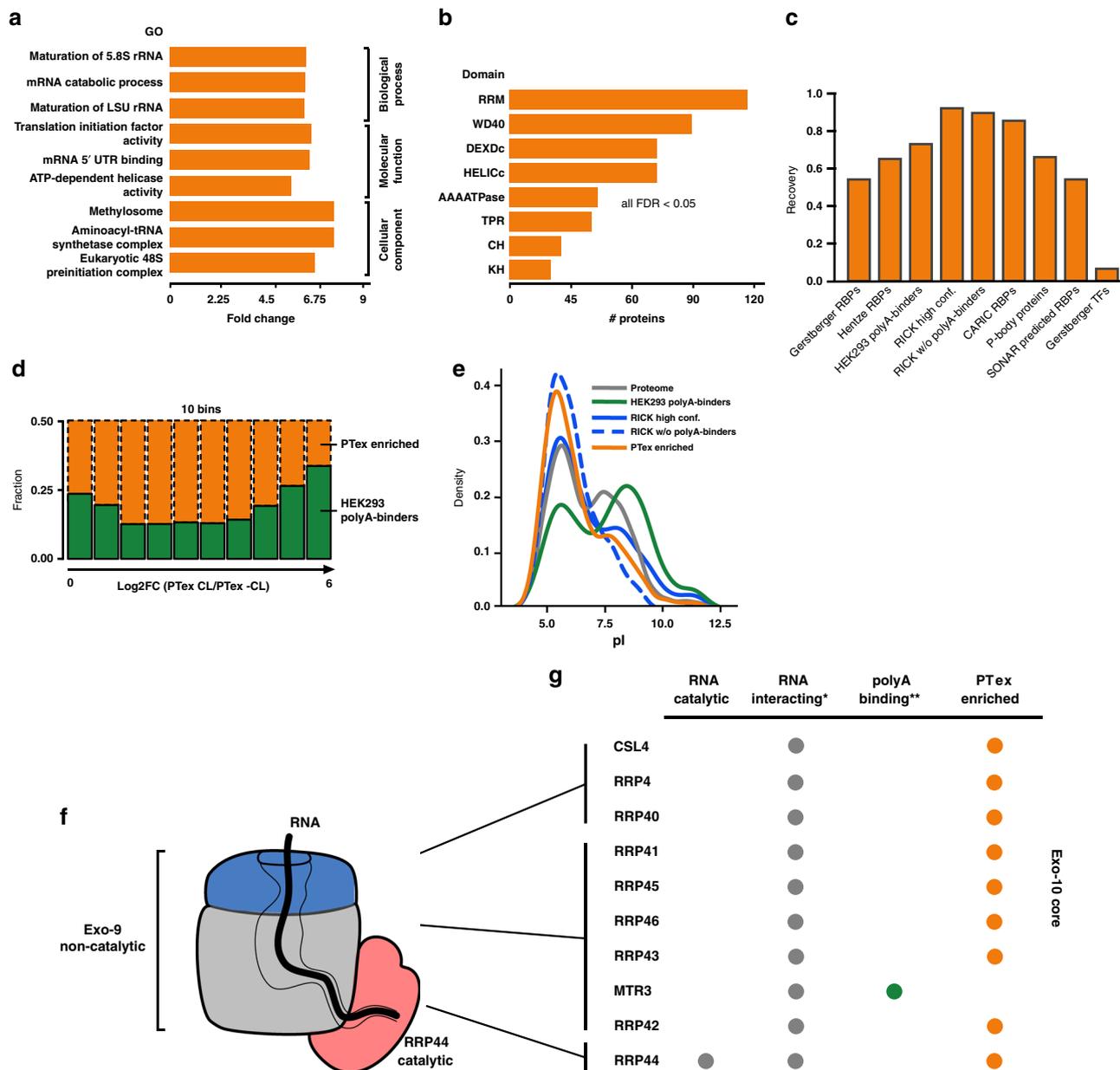
**Human cell culture and in vivo cross-linking.** Human embryonic kidney cells (HEK293, a gift from Markus Landthaler, Max Delbrück Center for Molecular

Medicine, Berlin, Germany) were grown on 148 cm<sup>2</sup> dishes using DMEM high glucose (Dulbecco's Modified Eagle, glucose 4.5 g/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% CO<sub>2</sub>. 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/cm<sup>2</sup> UV light ( $\lambda = 254$  nm) in a CL-1000 ultraviolet cross-linker device (Ultra-Violet Products Ltd), collected in 15 mL tubes, pelleted by centrifugation (1000×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). In addition, 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<sup>14</sup> were analysed with the RNA 6000 Chip Kit in a Bioanalyzer 2100 (Agilent, 5067–1513).

**Bacterial cell culture and in vivo cross-linking.** *Salmonella* Typhimurium SL 1344 Hfq::x3FLAG (a gift from Jörg Vogel,<sup>49</sup>) was grown on LB medium to stationary phase (OD<sub>600</sub> = 3). Aliquots of 20 ml were pelleted (20,000×g, 8 min, 37 °C) and resuspended in 1/10 of water for UV irradiation. Cells were cross-linked on ice with 5 J/cm<sup>2</sup> UV light ( $\lambda = 254$  nm) in a CL-1000 ultraviolet cross-linker 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 Fig. 7b. *Salmonella enterica* subsp. *enterica* Serovar Typhimurium strain SL1344 used for the mapping of RNA-protein interactions was grown in LB medium to OD<sub>600</sub> 2.0. Half of the cultures were cross-linked in a Vari-X-linker (UVO3, [www.vari-x-link.com](http://www.vari-x-link.com)), using UV light ( $\lambda = 254$  nm) lamps for 90 s. Fractions of 10 ml from each, cross-linked and non-cross-linked cultures, were harvested by filtration<sup>64</sup>.

**Construction of bacterial strains.** Yihii::x3FLAG::KmR was constructed following the procedure based on the Lambda Red system developed by ref. <sup>65</sup>. The system is based on two plasmids: pKD46, a temperature-sensitive plasmid that carries gamma, beta and exo genes (the bacteriophage  $\lambda$  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 (forward: 5'-GAA GCA GGA AGA TAT GAT GCG CCT GCT AAG AGG CGG CAA CGA CTA CAA AGA CCA TGA CG-3' and reverse: 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 1 h and, upon purification, used for subsequent electroporation. Similarly, AhpC::6xHis-TEV-3xFLAG::TetR and SipA::6xHis-TEV-3xFLAG::TetR constructs were produced by amplification of the plasmid pJet1.2-Hfq-HTF-TetR<sup>66</sup> with the primers AhpC-forward (5'-AAA GAA GGC GAA GCG ACT CTG GCT CCA TCC TTA GAC CTG GTC GGT AAA ATC CGC TCT GC TGG ATC CAT GGA G-3') and AhpC-reverse (5'-GTG AGC AGG CGA CGC CAA CGC AGC TAT GGC GTG AAA GAC GAC GGA AAT TTA CGC GTG AGG GGA TCT TGA AG-3') or SipA-forward (5'-CCT GGC GTG GAT CGG GTT ATT ACT ACC GTT GAT GGC TTG CAC ATG CAG CGT CGC TCT GCT GGA TCC ATG GAG-3') and SipA-reverse (5'-TTT GAC TCT TGC TTC AAT ATC CAT ATT CAT CGC ATC TTT CCC GGT TAA TTA CGC TGT AGG GGA TCT TGA AG-3'). Prior to electroporation, PCR products were digested with ExoI and DpnI during 1 h at 37 °C followed by ethanol precipitation and verification on Agarose gels.

*Salmonella* Typhimurium SL1344 harbouring the plasmid pKD46 was grown in LB containing Ampicillin (100 µg/ml) and L-Arabinose (100 mM) at 28–30 °C to an OD<sub>600</sub> of 0.8. Cells were incubated on ice for 15 min, centrifuged at 3220×g for 5 min 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 1 h in LB at 37 °C on a tabletop thermomixer at 600 rpm, plated on LB agar with Kanamycin (50 µg/ml) or



**Fig. 6** 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; *p*-value derived from a one-tail Fisher Exact test and FDR determined by the Benjamini-Hochberg method<sup>75</sup>. **c** PTex-purified proteins overlap with well-described RBPs but not transcription factors. Recovery of Gerstberger RBPs and transcription factors (TFs) reviewed in ref. <sup>29</sup>, a recent review on RBPs by ref. <sup>8</sup>, HEK293 poly-A binders<sup>2</sup>, RBPs found by RNA interactome using click chemistry (RICK;<sup>12</sup> CARIC<sup>13</sup>), P-body components<sup>39</sup> and a recent prediction of candidate RBPs (SONAR<sup>40</sup>). **d** Distribution of previously identified HEK293 mRNA-binding proteins (green;<sup>2</sup>) 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<sup>1, 2</sup>. RNA-interactors in general peak at pI 5-6 as found by PTex and RICK<sup>12</sup>. Proteins of the RNA exosome are prototype PTex proteins. **f** The RNA exosome core consists of 10 subunits: nine non-catalytically active proteins (Exo-9) forming a barrel-like structure and an additional RNase (Rrp44; Exo-10); modified from ref. <sup>41</sup>. **g** \*All exosome subunits are labelled "RNA-binding" (Uniprot.org); \*\*green = identified via poly-A selection in ref. <sup>2</sup>; orange = enriched in PTex

Tetracycline (100 µg/ml) overnight. The following day, 10 colonies per strain were picked, resuspended in PBS and streaked on plates containing Ampicillin or Kanamycin (or Tetracycline) and incubated at 37–40 °C. Colonies that showed resistance to Kanamycin (or Tetracycline) but not to Ampicillin were selected for further analysis, and the correct expression of the epitope tag was verified by western blot.

**PTex.** HEK293 suspensions in 600 µl of DPBS (5–8 × 10<sup>6</sup> 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 rpm, Eppendorf ThermoMixer) and centrifuged 20,000×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*-laurylosyl-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 in the same tube and mixed with 400 µl water, 200 µl ethanol p.a., 400 µl neutral phenol and 200 µl BCP (1 min, 21 °C, 2000 rpm, 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×g, pellets dried under the hood for max. 10 min and solubilised with 30 µL

**Table 1 Starting material of RNP purification methods**

Method	Starting material (cells)	Study
PAR-CLIP	2–9×10 <sup>8</sup>	(11)
RIC	2.85×10 <sup>8</sup>	(1)
RBR-ID	N/A	(76)
RICK	2×10 <sup>7</sup>	(12)
CARIC	3.6×10 <sup>8</sup>	(13)
PText	5–8×10 <sup>6</sup>	This study

RNP ribonucleoproteins, N/A not applicable

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. A detailed step-by-step protocol for PText is available in the Supplementary Methods.

**Hot-PText from mouse tissue.** Mice brain samples were a kind gift from M. Preußner (Heyd lab, FU Berlin). All animal experiments with C57BL/6 mice were performed in accordance with institutional and governmental recommendations and laws (Approval no. T0311/13). Mouse-brain tissue (260 mg) was disrupted by cryogenic grinding. Immediately after, 130 mg were irradiated with 0.75 J/cm<sup>2</sup> UV light ( $\lambda = 254$  nm) in a CL-1000 ultraviolet cross-linker device (Ultra-Violet Products Ltd). Samples ( $\pm$ CL) were resuspended in DPBS on ice. 600  $\mu$ l aliquots (32.5 mg) were mixed during 5 min at 65 °C in the presence of 0.5 g of low-binding zirconium beads (100  $\mu$ , OPS Diagnostics, LLC) and the phenol-toluol-BCP mix described above. After centrifuging (20,000 $\times$ g, 3 min, 4 °C) Aq1 was carefully removed. Consecutive extractions were done as described above with the sole difference that mixing steps were performed at 65 °C (2000 r.p.m, Eppendorf ThermoMixer).

**Hot-PText from *Salmonella*.** For global mapping of RNA-protein interactions in *Salmonella* cells from two biological replicates were used. Bacteria attached to filters obtained as described in the section Bacterial cell culture and in vivo cross-linking were collected with 12 ml of DPBS and aliquots of 4 ml were pelleted at 20,000 $\times$ g, 2 min, 4 °C. A modification of the step 1 was introduced in HOT-PText in order to improve removal of free proteins, as follows: bacterial pellets ( $\pm$ CL) resuspended in 400  $\mu$ l of DPBS supplemented with EDTA (5 mM) were mixed with phenol-toluol-BCP (200  $\mu$ l each), and 0.5 g of zirconium beads during 5 min at 65 °C (2000 rpm, Eppendorf ThermoMixer). After centrifugation at 20,000 $\times$ 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 transferred to a third tube where steps 2 and 3 of the PText protocol were performed at 65 °C.

**Analysis of individual PText steps.** PText extractions were carried as described before, using sets of three tubes containing synthetic 30–50 nt RNAs 5'-labelled <sup>32</sup>P-ATP, 200 ng pUC19 LacZ-containing fragment (817 bp, generated by DrdI, NEB) or 2–3  $\times$  10<sup>6</sup> HEK293 cells spiked-in with 0.25  $\mu$ g of Sxl-RBD4 per tube. PCRs were performed using primers designed to amplify endogenous chromosomal DNA, 11-3 gene (574 bp, forward primer: 5'-GAT CCG ATC CTA ATA CGA CTC ACT ATA GGC GAC ATC CAA TCC ATA TCA AGG A-3' and reverse primer: 5'-GAT CAA GCT TGT TCA GAG TCT AGT TTA TTC TCA CAC-3'), or the LacZ gene present in the pUC19 linear fragment (324 nt, forward 5'- AGA GCA GAT TGT ACT GAG-3' and M13-reverse 5'-CAG GAA ACA GCT ATG ACC-3'). DNA and RNA samples were electrophoresed in Agarose 1.0–1.5% or TBE-Urea PAGE 12%, respectively. Radioactivity was detected by phosphorimaging, while Sxl-RBD4, endogenous HuR or ACTB in HEK293/Sxl-RBD4 spiked-in proteins were analysed by western blot with specific antibodies as described in the western Blotting section.

**RNaseA digestion and electrophoretic mobility assay.** Cross-linked and non-cross-linked HEK293 cell suspensions were subjected to PText, the resulting pellets were solubilised in 150  $\mu$ l of buffer TED (20 mM Tris, 1 mM EDTA, 0.03% DDM [n-Dodecyl  $\beta$ -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  $\mu$ l were taken at different time points: 0, 1, 5, 10, 30 and 60 min, immediately mixed with 5  $\mu$ l of 6x Laemmli buffer, heated at 95 °C for 5 min and used for SDS-PAGE.

**RNase treatment prior PText.** For PText extractions shown in Fig. 1g, suspensions of 2–3  $\times$  10<sup>6</sup> HEK293 cells/ml ( $\pm$ CL) were treated with 2000 U/ml benzonase (Merck, 70664) in the recommended buffer (50 mM Tris, 1 mM MgCl<sub>2</sub>, pH 8.0) during 1 h at 37 °C and 1000 rpm (ThermoMixer, Eppendorf). Untreated cells

( $\pm$ CL) were used as controls. PText extraction was then performed as described above. After ethanol precipitation, pellets were directly solubilised in 40  $\mu$ l Laemmli buffer (2 $\times$ ). SDS-PAGE and western blots were performed as indicated below.

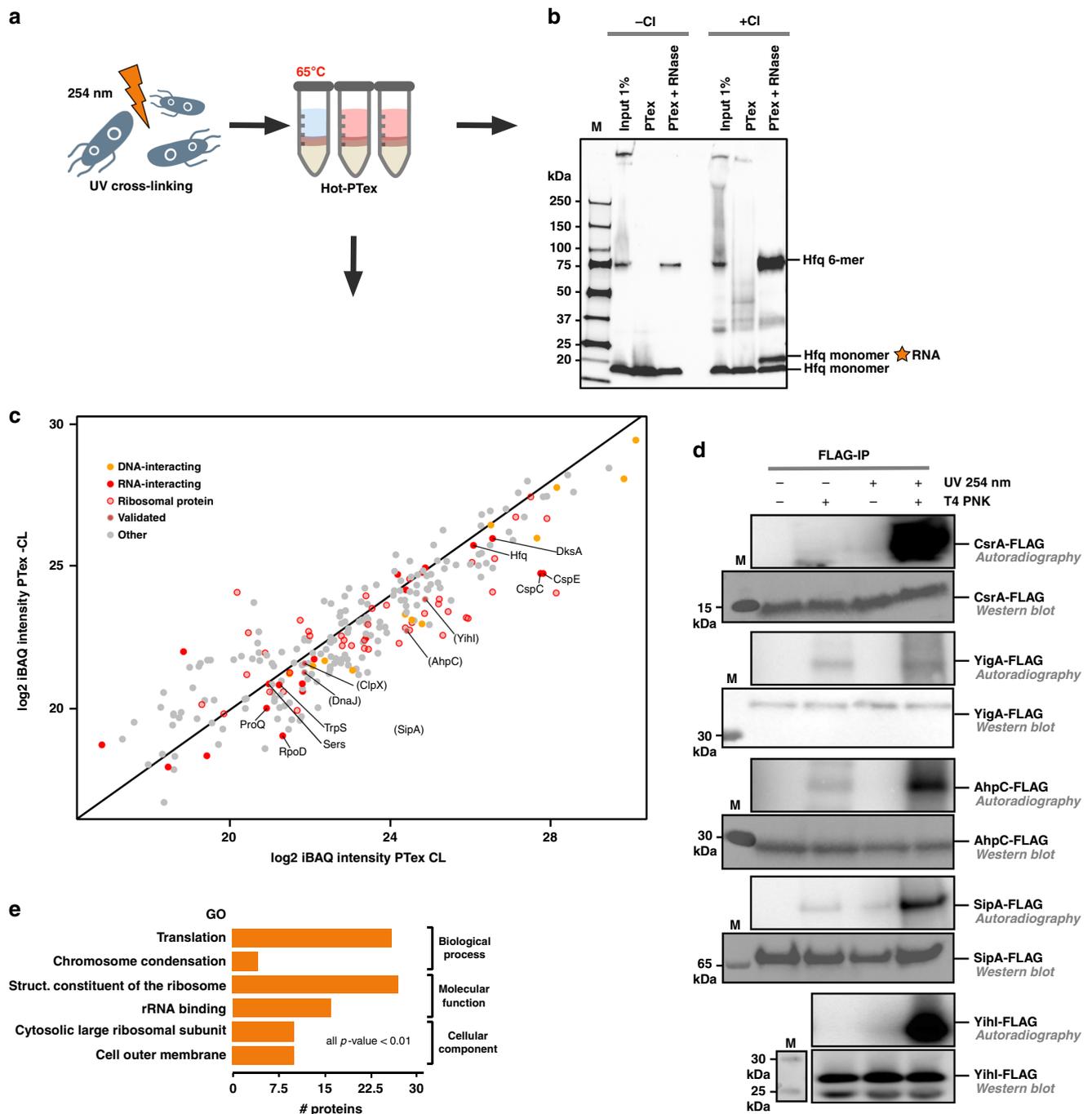
**In vitro transcription of Sxl-RBD4 target RNAs.** The T7 promoter and a sxl-target DNA sequence 5'-GAT CCG GTC ATA GGT GTA AAA AAA GTC TCC ATT CCT ATA GTG AGT CGT ATT AA-3' was cloned into pUC19 using the restriction enzymes BamHI and HindIII. The resulting plasmid was named pUC19-sxl-target. Templates 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). The 30 nt RNA was synthesised as described before<sup>67</sup> by hybridising two complementary sequences containing the T7 RNA polymerase 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. 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<sup>14</sup>. RNA 5'-GAG UUU UUU UAC A-3' (13 nt) was synthesised by Biomers (Ulm, Germany).

**In vitro cross-linking assays.** Sxl-RBD4 (40  $\mu$ g) in 100  $\mu$ l cross-linking 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, 1.7–10  $\mu$ M) 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/cm<sup>2</sup> of UV-254 nm, on ice. Afterwards, 98% of each sample was used for PText extraction, while 2% of the sample were kept as input control for SDS-PAGE and western blotting to detect Sxl-RBD4.

**Quantification of PText.** In order to unbiasedly determine the overall yield of PText, we prepared mRNA interactome capture (RIC) samples from +CL HEK293<sup>1</sup> to serve as a only-clRNPs starting material. RIC samples from five biological replicates were used for PText extraction. After ethanol precipitation, PText samples were washed once with 5 ml of cold ethanol to remove traces of phenol which could interfere with the quantification, and resuspended in 20–50  $\mu$ l of water. The absorbance of RIC and PText samples were measured at  $\lambda$ 280 nm and  $\lambda$ 260/280 nm in a Nanodrop 2000. In addition, 2 and 45% of RIC and PText samples, respectively, were digested with RNaseA (0.1  $\mu$ g/ $\mu$ l) at 37 °C during 40 min. Intact and digested samples were loaded in Bis-Tris-MOPS gels 4–12% (NuPage, Invitrogen), transferred to nitrocellulose membranes and blotted to detect HuR (see below). A selection of western blot images generated during this study (Supplementary Figs 13,14) were used to calculate the performance of PText in terms of yield and specific clRNP enrichment by densitometry analysis using ImageJ<sup>68</sup>.

**Western blotting.** Western blotting was performed following standard techniques. Samples were electrophoresed on SDS-PAGE gradient gels 4–20% (TGX stain free, BioRad) or Bis-Tris 4–12% (NuPage, Invitrogen) and proteins transferred onto nitrocellulose membranes 0.2  $\mu$ m (BioRad). Membranes were blocked during 30 min with PBST-M (10 mM phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, pH 7.4 0.1% tween 20 (Sigma), 5% milk) and incubated with 0.1–1.0  $\mu$ g/ml of the respective antibody overnight at 4 °C (or 2 h room temperature). Primary antibodies targeted the proteins HuR (1:1000, Proteintech, 11910-1-AP), ABCF2 (1:1000, Proteintech, 10226-1-AP), CCT7 (1:1000, Proteintech, 15994-1-AP), FUS (1:1000, abcam, ab124923), GAPDH (1:1000, Proteintech, 10494-1-AP), alpha-enolase (ENO1, 1:1000, Proteintech, 11204-1-AP), PTBP1 (1:1000, abcam, ab133734), PABPC1 (1:1000, Proteintech, 10970-1-AP), ACTB (1:1000, Proteintech, 66009-1-Ig), Histone H3 (1:1000, abcam, ab21054), FLAG-tag (1:5000, Sigma, A8592), or Sxl-RBD4 (DHSB, anti-Sxl hybridome culture supernatant M114, 1:20). 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 anti-mouseHRP (1:5000, Proteintech, SA00001-1), anti-mouse-AlexaFluor680 (1:2000, Invitrogen, A32729) anti-rabbitHRP (1:2500, Proteintech, SA00001-2), or anti-rabbit-AlexaFluor488 (1:2000, Invitrogen, A32731) 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 labelling of RNA by PNK was performed as described<sup>49</sup>. Cell pellets were resuspended in 800  $\mu$ l NP-T buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 0.05% Tween, pH 8.0) together with 1 ml glass beads (0.1 mm). Cells were lysed by shaking at 30 Hz for 15 min at 4 °C and centrifuged for 15 min at 16,000 $\times$ g and 4 °C. Cell lysates were transferred to new tubes and centrifuged for 15 min at 16,000 $\times$ g and 4 °C. The cleared lysates were mixed with one volume of NP-T buffer with 8 M urea, incubated for 5 min at 65 °C while at 900 rpm and diluted 1:10 in ice-cold NP-T buffer. Anti-FLAG magnetic beads (Sigma) were washed three times in NP-T buffer (30  $\mu$ l 50% bead suspension was used for a lysate from 100 ml bacterial culture), added to the lysate, and the mixture



**Fig. 7** 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<sup>49</sup> and that this complex is also enriched after PTex. **c** RNPs in *Salmonella* were purified by PTex globally. 172 Proteins enriched after UV-cross-linking (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 PTex-enriched RNA-interactors: *Salmonella* strains expressing FLAG-tagged proteins were immunoprecipitated ±UV irradiation. RNA-association is confirmed by radioactive labelling of RNA 5' ends by polynucleotide kinase (T4 PNK) using autoradiography; a signal is exclusively detectable after UV-cross-linking and radiolabelling of precipitated RNA. CsrA-FLAG (pos. ctr.), YigA-FLAG (neg. ctr.), AhpC-FLAG, SipA-FLAG and YihI-FLAG are bound to RNA in vivo. **e** GO terms significantly enriched among the RNA-associated proteins; *p*-value derived from a one-tail Fisher Exact test<sup>75</sup>. For full gels/blots see Supplementary Figures 25, 26

was rotated for 1 h at 4 °C. Beads were collected by centrifugation at 800×g, resuspended in 1 ml NP-T buffer, transferred to new tubes, and washed twice with high-salt buffer (50 mM NaH<sub>2</sub> PO<sub>4</sub>, 1 M NaCl, 0.05% Tween, pH 8.0) and twice with NP-T buffer. Beads were resuspended in 100-µl NP-T buffer containing 1 mM MgCl<sub>2</sub> and 2.5 U benzonase nuclease (Sigma) and incubated for 10 min at 37 °C while shaking at 800 rpm, followed by 2 min incubation on ice.

After one wash with high-salt buffer and two washes with CIP buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>), the beads were resuspended in 100 µl CIP buffer with 10 units of calf intestinal alkaline phosphatase (NEB) and incubated for 30 min at 37 °C while shaking at 800 rpm. This was followed by one wash with high-salt buffer and two washes with PNK buffer (50 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>, 0.1 mM spermidine). Next, beads were resuspended in 100 µl

PNK buffer and 1 U T4 PNK (0.1 U/μl, NEB) and 5.5 μCi γ-<sup>32</sup>P-ATP were added and incubated at 37 °C for 30 min. The beads were then washed twice in PNK buffer and resuspended in 50 μl of a 2× denaturing gel loading buffer (Invitrogen). 30 microliter samples were then analysed on a NuPAGE Bis-Tris gel (Invitrogen) and radioactive signal was detected in a Life Science FLA-5100 imaging system (Fujifilm).

**Protein purification.** Recombinant Sxl-RBD4 protein (Sxl amino acids 122–301) was purified essentially as described before<sup>17</sup>. In brief, after IPTG induction for 4 h 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-HCl pH 7.5, 1 M NaCl, 0.2 mM EDTA, 1 mM DTT, cOmplete Protease Inhibitor Cocktail [Roche] followed by centrifugation for 20 min at 12,000×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 3 C protease and dialysed 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, dialysed 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 and bacterial 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, samples were treated with Benzamide for 30 min at 37 °C, spun down to remove cell debris and supernatant containing cellular proteins 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 min centrifugation at 20,000×g at 4 °C. Protein pellets were resuspended in 2 M urea in 50 mM ammonium bicarbonate (ABC) buffer and Benzamide was added for 30 min at 37 °C to digest RNA. Proteins were precipitated with methanol-chloroform extraction<sup>69</sup> and resuspended 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 1:50 (w/w) at room temperature for 3 h, diluted to 2 M final concentration of urea with 50 mM ammonium bicarbonate, and further digested with trypsin (Promega) at a trypsin-to-protein ratio of 1:50 (w/w) under constant agitation at room temperature for 16 h. Peptides were desalted with C18 Stage Tips<sup>70</sup> prior to LC-MS/MS analysis. Peptides were separated on a 2 m long monolithic column (100 μm ID, MonoCap C18 High Resolution 2000 [GL Sciences] kindly provided by Dr. Yasushi Ishihama [Kyoto University]), or on an in-house packed C18 15 cm microcolumns (75 μm ID, packed with ReproSil-Pur C18-AQ 3-μm resin, Dr. Maisch GmbH). Peptides were eluted from columns with an increasing acetonitrile concentration gradient of 5–50% at a flow rate of 200 nL/min for 6 h (long column) or 300 nL/min for 2 h (short column). Q Exactive Plus 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). Full scans were performed with a resolution of 70,000, a target value of 3 × 10<sup>6</sup> ions and a maximum injection time of 20 ms. The MS/MS scans were performed with a 17,500 resolution, a 1 × 10<sup>6</sup> target value and a 60 ms maximum injection time. Isolation window was set to 2 and normalised collision energy was 26. Ions with an unassigned charge state and singly charged ions were rejected.

All raw files were analysed with MaxQuant software (v1.5.1.2)<sup>71</sup> using the label-free quantification (LFQ) algorithm<sup>72</sup> with default parameters and match between runs option on. Search parameters included minimum peptide length of 7, two missed cleavage sites, protease specificity: trypsin (allowing also for cuts after proline), cysteine carbamidomethyl fixed modification and variable modifications including methionine oxidation and protein N-terminal acetylation. The mass tolerance for precursor ions was 6 ppm and the mass tolerance for fragment ions was 20 ppm. Database search was performed against the human reference proteome (UNIPROT, downloaded in October 2014) or the *Salmonella* Typhimurium reference proteome (UNIPROT, downloaded in August 2017) with common serum and enzyme contaminants sequences included. A minimum peptide count required for protein quantification was set to two. False discovery rate (FDR) was set to 1% at peptide spectrum match (PSM) and protein levels.

**Bioinformatic analysis of PTex-purified proteins (HEK293).** The complete analysis is available as a supplementary R notebook (Supplementary Data 8). 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 log<sub>2</sub> 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:<sup>3,59</sup> P-values were calculated from an EBayes moderated *t*-test using the limma package<sup>73</sup>

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<sup>74</sup>. Domain enrichment was done using DAVID<sup>75</sup> searching the SMART<sup>76</sup> database.

**Bioinformatic analysis of PTex-purified proteins (Salmonella).** The complete analysis is available as supplementary R notebook (Supplementary Data 9). We used iBAQ-normalised values for the *Salmonella* analysis. Potential contaminants, reverse peptides and peptides only identified by modification were excluded from analysis. Fold changes were calculated by subtraction of the log<sub>2</sub> 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 log<sub>2</sub> fold-change >0). Domain and GO terms were analysed using DAVID<sup>75</sup>.

**RNA-Seq sample preparation.** Cross-linked PTex samples and non-cross-linked control (–CL, 0.015, 0.15, 1.5 J/cm<sup>2</sup>) were proteinase K digested (1 h, 56 °C) and the RNA recovered by acidic phenol extraction<sup>14</sup> using phase lock gel tubes (5Prime, 2302830). Libraries were created according to the TruSeq Stranded Total RNA LT protocol (Illumina, 15032612) with the modification that we skipped the rRNA depletion step. We used adaptors 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. Note that we sequenced each condition in triplicates with the exception of input 0.015 J/cm<sup>2</sup> for which we have duplicates.

**Bioinformatic analysis of PTex-purified RNA.** RNA-Seq data were quality controlled 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) with the following command line:

```
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) with the following command line:

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

After aligning the remaining reads with STAR, samtools was used to annotate each read in the bam files with mutation information. Using the SAM.py parser from the pyCRAC suite<sup>25</sup>, chromosomal locations of substitutions and deletions were extracted and counted. Only mutations that were unique to the UV-treated samples were considered. To normalise 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. The following command lines were used:

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

**PAR-CLIP and pCLIP.** We performed the PAR-CLIP protocol as described:<sup>11</sup> 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<sup>2</sup> 365 nm UV light, snap-frozen on dry ice and stored at –80 °C until use. Cells were collected on different days, representing biological replicates. Cells (~1.2 × 10<sup>8</sup> 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 No. A1531) containing 0.04 U/ml RNasin (Promega, N2515) and 2x Complete Protease Inhibitor (Roche, 11697498001) and centrifuged 20,000×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 (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 1% Nonidet P40, 0.1% (v/v) SDS, 1 M NaCl) beads were incubated with 1 U/ $\mu$ l of T4 polynucleotide kinase (T4-PNK, NEB) and 0.5  $\mu$ Ci/ $\mu$ l of  $^{32}$ P-ATP. After radiolabelling, samples were split into 4 tubes and underwent three versions of CLIP (classic PAR-CLIP, PAR-CLIP on-beads and pCLIP).

In classic PAR-CLIP, cHuR-RNA complexes were resolved by 4–12% Nu-PAGE MOPS (Invitrogen) transferred to nitrocellulose and excised at a defined size-range (50–60 kDa). Proteins were digested from the membrane with proteinase K and the RNA recovered by acidic phenol/chloroform extraction and ethanol precipitation. Recovered RNA was ligated with 3' adaptor (5'-App-NNN NTG GAA TTC TCG GGT GCC AAG G-3'InvdT) gel-slice isolated, ligated with the 5' adaptor (5'-GUU CAG AGU UCU ACA GUC CGA CGA UCN NNN-3') and purified again from PAGE-Urea gels by elution and ethanol precipitation<sup>11</sup>.

As an alternative, we performed PAR-CLIP on-beads. Here, the ligation of the 3'/5' adaptors can be achieved directly on the beads used in the affinity capture of the selected cRNP<sup>21,22</sup>. On-beads adaptor ligation was done by incubating the FLAG-cHuR-RNA beads with the 3' adaptor in the presence of Rnl2(1–249) K227Q ligase and PEG-8000 overnight at 4 °C. After washes, the 5' adaptor was ligated using the Rnl1 enzyme, 2 h at 37 °C as described in ref. <sup>22</sup>. Followed by protein electrophoresis (4–12% Nu-PAGE MOPS, Invitrogen), blotting to nitrocellulose membranes and band excision at the defined size-range (50–60 kDa). Proteins were digested from the membrane with proteinase K. RNA was then recovered by acidic phenol/chloroform extraction and ethanol precipitation.

pCLIP beads capturing cHuR-RNA complexes were subjected to 3'/5' adaptors ligation as above. Immediately after the 5' adaptor ligation beads were magnetically captured and resuspended in 600  $\mu$ l solution D. After a short denaturation (95 °C during 10 min), beads were separated with a magnet and the cHuR-RNA complexes recovered in the supernatant were subjected to the last two steps of the PTex protocol (20 min). PTex recovered interphases were precipitated with 9 volumes of ethanol on dry ice during 30 min, followed by 30 min centrifuging at 20,000 $\times$ g, 4 °C. Ethanol-precipitated pellets were digested with proteinase K and RNA isolated by phenol/chloroform extraction.

**Library preparation and RNA-Seq.** RNAs obtained from the three PAR-CLIP procedures (classic, on-beads and pCLIP) were reverse-transcribed into cDNA with the reverse transcription primer 5'-GCC TTG GCA CCC GAG AAT TCC A-3' and the minimal 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 Illumina adaptor index RPI 1–6, 8,10–11. Bands obtained at 150 bp were excised from 2% agarose gels and purified using the Zymo-clean Gel Recovery Kit (Zymo, D4002). DNA concentration and library quality was determined by Qubit Fluorometer ds-DNA HS assay (Life Tech, Q32854) and BioAnalyzer DNA HS Kit (Agilent 2100 Bioanalyzer; Agilent, 5067–4626). Libraries were sequenced in a NextSeq 500.

**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<sup>60</sup>, with one modification. In brief, adaptor sequences were trimmed retaining the four randomised adaptor nucleotides on both read ends included during adaptor 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 (Supplementary Data 6).

**De novo motif discovery.** For de novo motif finding we used Zagros with default settings including RNA secondary structure information. As the majority of HuR cluster resided in intronic and 3'UTR regions (70–80%, Supplementary Data 7) according to its reported functions, we use intronic and 3'utr cluster sequences as input. For DREME motif analysis, we used all PARalyzer-derived cluster sequences and ran DREME with default settings against shuffled background sequences allowing only sense strand motif search.

**Transcriptomic metacoverage.** For depicting spatial preferences for mRNA binding, we selected genes previously used for RNA classification based on processing and turnover dynamics<sup>77</sup> ( $n = 15,120$ ) present in GENCODE v19. To select transcripts, we ran RSEM<sup>78</sup> 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<sup>79</sup>. 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 emphasise 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<sup>80</sup>. 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 visualisations.** PAR-CLIP alignments were visualised using Gviz<sup>81</sup>.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

RNA-seq data have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE113628 for PAR-CLIP RNA-seq and GSE113655 for PTex RNA-seq. Read count data for RNA classes (Supplementary Fig. 20 are available in Supplementary Data 5). The proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD009571. MaxQuant analysed-proteomics data (protein groups for HEK293 and *Salmonella*) are available in Supplementary Data 8 and 9, respectively. A reporting summary for this Article is available as a Supplementary Information file. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Received: 6 December 2018 Accepted: 11 February 2019

Published online: 01 March 2019

## References

- Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 393–1406 (2012).
- Baltz, G. A. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
- Beckmann, B. M. et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **6**, 10127 (2015).
- Favre, A. et al. 4-thiouridine photosensitized RNA-protein crosslinking in mammalian cells. *Biochem. Biophys. Res. Commun.* **141**, 847–854 (1986).
- Hockensmith, J. W. et al. Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. *J. Biol. Chem.* **261**, 3512–3518 (1986).
- Brimacombe, R. et al. Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol.* **164**, 287–309 (1988).
- Beckmann, B. M., Castello, A. & Medenbach, J. The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *PLoS Arch.* **468**, 1029–1040 (2016).
- Hentze, M. W., Castello, A., Schwarzl, T., & Preiss T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018).
- Ule J., Hwang H. W. & Darnell R. B. The future of cross-linking and immunoprecipitation (CLIP). *Cold Spring Harb. Perspect. Biol.* **10**, a032243 (2018)
- Lee, F. C. Y. & Ule, J. Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* **69**, 354–369 (2018).
- Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
- Bao, X. et al. Capturing the interactome of newly transcribed RNA. *Nat. Methods* **15**, 213–220 (2018).
- Huang R., Han M., Meng L. & Chen X. Capture and identification of RNA-binding proteins by using click chemistry-assisted RNA-interactome capture (CARIC) strategy. *J. Vis. Exp.* 10.3791/58580 (2018).
- Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159 (1987).
- Chey, S., Claus, C. & Liebert, U. G. Improved method for simultaneous isolation of proteins and nucleic acids. *Anal. Biochem.* **411**, 164–166 (2011).
- Lebedeva, S. et al. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein Hur. *Mol. Cell* **43**, 340–352 (2011).

17. Moschall, R. et al. Sister-of-sex-lethal is a repressor of translation. *RNA* **24**, 149–158 (2018).
18. Singh, R., Valcárcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173–1176 (1995).
19. Kishore, S. et al. A quantitative analysis of clip methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* **8**, 559–564 (2011).
20. Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
21. Porter, D. F. et al. Target selection by natural and redesigned Puf proteins. *Proc. Natl Acad. Sci. USA* **112**, 15868–15873 (2015).
22. Benhalevy, D., McFarland, H. L., Sarshad, A. A. & Hafner, M. PAR-CLIP and streamlined small RNA cDNA library preparation protocol for the identification of RNA binding protein target sites. *Methods* **118–119**, 41–49 (2017).
23. Hogan, D. J. et al. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* **6**, e255 (2008).
24. Gehring, N. H., Wahle, E. & Fischer, U. Deciphering the mRNP code: RNA-bound determinants of post-transcriptional gene regulation. *Trends Biochem. Sci.* **42**, 369–382 (2017).
25. Webb, S., Hector, R. D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* **15**, R8 (2014).
26. Schueler, M. et al. Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol.* **15**, R15 (2014).
27. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
28. Freund, A. Proteostatic control of telomerase function through tric-mediated folding of Tcab1. *Cell* **159**, 1389–1403 (2014).
29. Gerstberger, S., Hafner, M. & Tuschl, M. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
30. Geiger, T. et al. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteom.* **11**, M111.014050 (2012).
31. Jin, W. et al. Structural basis for snRNA recognition by the double-WD40 repeat domain of Gemin5. *Genes Dev.* **30**, 2391–2403 (2016).
32. Wegrecki, M., Neira, J. L. & Bravo, J. The carboxy-terminal domain of Erb1 is a seven-bladed  $\beta$ -propeller that binds RNA. *PLoS One* **10**, e0123463 (2015).
33. Wang, Q., Hobbs, K., Lynn, B. & Rymond, B. C. The Clfp splicing factor promotes spliceosome assembly through n-terminal tetratricopeptide repeat contacts. *J. Biol. Chem.* **278**, 7875–7883 (2003).
34. Halbach, F., Reichelt, P., Rode, M. & Conti, E. The yeast Ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* **154**, 814–826 (2013).
35. Rispal, D. et al. Structural and functional analysis of Nro1/Ett1: a protein involved in translation termination in *S. cerevisiae* and in O<sub>2</sub>-mediated gene control in *S. pombe*. *RNA* **17**, 1213–1224 (2011).
36. Ban, N. et al. A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* **24**, 165–169 (2014).
37. Conrad, A. et al. Serial interactome capture of the human cell nucleus. *Nat. Commun.* **7**, 11212 (2016).
38. Milek, M. et al. Ddx54 regulates transcriptome dynamics during DNA damage response. *Genome Res.* **27**, 1344–1359 (2017).
39. Hubstenberger, A. et al. P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157 (2017).
40. Brannan, K. W. et al. Sonar discovers RNA-binding proteins from analysis of large-scale protein-protein interactomes. *Mol. Cell* **64**, 282–293 (2016).
41. Makino, D. L., Halbach, F. & Conti, E. The RNA exosome and proteasome: common principles of degradation control. *Nat. Rev. Mol. Cell Biol.* **14**, 654–660 (2013).
42. Lorentzen, E., Basquin, J. & Conti, E. Structural organization of the RNA-degrading exosome. *Curr. Opin. Struct. Biol.* **18**, 709–713 (2008).
43. Wasmuth, E. V., Januszkyk, K. & Lima, C. D. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* **511**, 435–439 (2014).
44. Zinder, J. C., Wasmuth, E. V. & Lima, C. D. Nuclear RNA exosome at 3.1 Å reveals substrate specificities, RNA paths, and allosteric inhibition of Rrp44/Dis3. *Mol. Cell* **64**, 734–745 (2016).
45. Delan-Forino, C., Schneider, C. & Tollervy, D. RNA substrate length as an indicator of exosome interactions in vivo. *Wellcome Open Res.* **2**, 34 (2017).
46. Petit, F. Involvement of proteasomal subunits zeta and iota in RNA degradation. *Biochem. J.* **326**, 93–98 (1997).
47. Horsch, A. et al. Proteasomes discriminate between mRNA of adenovirus-infected and uninfected HeLa cells. *FEBS Lett.* **246**, 131–136 (1989).
48. Fátol, K. & Grummt, I. Proteasomal ATPases are associated with rDNA: the ubiquitin proteasome system plays a direct role in RNA polymerase I transcription. *Biochim. Biophys. Acta* **1779**, 850–859 (2008).
49. Holmqvist, E. et al. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* **35**, 991–1011 (2016).
50. Smirnov, A. Grad-seq guides the discovery of Proq as a major small RNA-binding protein. *Proc. Natl Acad. Sci. USA* **113**, 11591–11596 (2016).
51. Michaux, C. et al. RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. *Proc. Natl Acad. Sci. USA* **114**, 6824–6829 (2017).
52. Hwang, J. & Inouye, M. A bacterial gap-like protein, YihI, regulating the GTPase of Der, an essential GTP-binding protein in *Escherichia coli*. *J. Mol. Biol.* **399**, 759–772 (2010).
53. Jacobson, F. S., Morgan, R. W., Christman, M. F. & Ames, B. N. An alkyl hydroperoxide reductase from *Salmonella Typhimurium* involved in the defense of DNA against oxidative damage. Purification and properties. *J. Biol. Chem.* **264**, 1488–1496 (1988).
54. Zhou, D., Mooseker, M. S. & Galán, J. E. Role of the *S. Typhimurium* actin-binding protein SipA in bacterial internalization. *Science* **283**, 2092–2095 (1999).
55. Tawk, C., Sharan, M., Eulalio, A. & Vogel, J. A systematic analysis of the RNA-targeting potential of secreted bacterial effector proteins. *Sci. Rep.* **7**, 9328 (2017).
56. Bressin A. et al. Tripepsvm - de novo prediction of RNA-binding proteins based on short amino acid motifs. *bioRxiv* <https://doi.org/10.1101/466151> (2018)
57. Matia-González, A. M., Laing, E. E. & Gerber, A. P. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat. Struct. Mol. Biol.* **22**, 1027–1033 (2015).
58. Tsatsaronis, J. A., Franch-Arroyo, S., Resch, U. & Charpentier, E. Extracellular vesicle RNA: a universal mediator of microbial communication? *Trends Microbiol.* **26**, 401–410 (2018).
59. Beckmann, B. M. RNA interactome capture in yeast. *Methods* **118–119**, 82–92 (2017).
60. Corcoran, D. L. et al. Paralyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12**, R79 (2011).
61. Queiroz R. M. L. et al. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* **37**, 169–178 (2019).
62. Trendel, J. et al. The human RNA-binding proteome and its dynamics during translational arrest. *Cell* **176**, 391–403 (2019).
63. Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070 (2014).
64. van Nues, R. et al. Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. *Nat. Commun.* **8**, 12 (2017).
65. Uzzau, S., Figueroa-Bossi, N., Rubino, S. & Bossi, L. Epitope tagging of chromosomal genes in Salmonella. *Proc. Natl Acad. Sci. USA* **98**, 15264–15269 (2001).
66. Tree, J. J. et al. Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *J. Mol. Cell* **55**, 199–213 (2014).
67. Beckmann, B. M., Grünweller, A., Weber, M. H. W. & Hartmann, R. K. Northern blot detection of endogenous small RNAs (approximately 14 nt) in bacterial total RNA extracts. *Nucleic Acids Res.* **38**, e147 (2010).
68. Rueden, C. T. et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinform.* **18**, 529 (2017).
69. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
70. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and lc/ms sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
71. Cox, J. & Mann, M. Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 367–372 (2008).
72. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlq. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
73. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
74. Mi, H. et al. Panther version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
75. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
76. Letunic, I. & Bork, P. 20 years of the smart protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).

77. Mukherjee, N. et al. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**, 86–96 (2017).
78. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
79. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comp. Biol.* **9**, e1003118 (2013).
80. Akalin, A. et al. Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127–1129 (2015).
81. Hahne, F. & Ivanek, R. Visualizing genomic data using gviz and bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).

## Acknowledgements

We wish to thank Florian Heyd and Marco Preussner for providing mouse brain samples, Jörg Vogel for providing Hfq-FLAG-tagged *Salmonella*, Stuart McKellar for help with cross-linking, and Ira Iosub and Julie Bohl for assistance. We are grateful to Maja Köhn and Matthias Hentze for discussions and to Markus Landthaler for critically reading the manuscript. Work in the Beckmann lab is supported 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. E.C.U. is supported by the Joachim Herz Foundation and an EMBO Short-Term Fellowship No 7511.

## Author contributions

E.C.U. and B.M.B. designed the project; E.C.U., C.H.V.-V., H.H.W., D.F. and R.M. performed the research; E.C.U., C.H.V.-V., T.H., H.H.W., S.G. and M.S. analysed the data; E.C.U., S.G., U.O., J.M. and B.M.B. wrote the paper with input from all authors.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-08942-3>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019