Unbiased dynamic characterization of RNA-protein interactions by OOPS

Rayner M. L. Queiroz^{1#}, Tom Smith^{1#}, Eneko Villanueva^{1#*}, Mie Monti¹, Mariavittoria Pizzinga², Maria Marti-Solano³, Dan-Mircea Mirea¹, Manasa Ramakrishna², Robert F. Harvey², Veronica Dezi², Sven Degroeve⁴, Lennart Martens⁴, Gavin H. Thomas⁵, Anne E. Willis², Kathryn S. Lilley^{1*}

[#]Authors contributed equally

¹ Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Cambridge, UK.

² MRC Toxicology Unit, University of Cambridge, Lancaster Rd, Leicester LE1 9NH, UK

³ MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK.

⁴ VIB, Ghent University Ghent, Belgium

⁵ Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK

*Correspondence should be addressed to:

Eneko Villanueva, Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. Email: ev318@cam.ac.uk

Kathryn S. Lilley, Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. Email: ksl23@cam.ac.uk

Abstract

Current methods for the identification of RNA–protein interactions require a quantity and quality of sample that hinders their application, especially for dynamic biological systems or when sample material is limiting. Here, we present a new approach to enrich RNA-Binding Proteins (RBPs): Orthogonal Organic Phase Separation (OOPS), which is compatible with downstream proteomics and RNA sequencing. OOPS enables recovery of RBPs and free protein, or protein-bound RNA and free RNA, from a single sample in an unbiased manner. By applying OOPS to human cell lines, we extract the majority of known RBPs, and importantly identify additional novel RBPs, including those from previously under-represented cellular compartments. The high yield and unbiased nature of OOPS facilitates its application in both dynamic and inaccessible systems. Thus, we have identified changes in RNA-protein interactions in mammalian cells following nocodazole cell-cycle arrest, and defined the first bacterial RNA-interactome. Overall, OOPS provides an easy-to-use and flexible technique that opens new opportunities to characterize RNA-protein interactions and explore their dynamic behaviour.

Introduction

Interactions between RNA-binding proteins (RBPs) and RNA are essential for the control of gene expression by regulating the processes of transcription, trafficking, decay and mRNA translation¹⁻⁷. In this way, RNA-RBP crosstalk controls cell homeostasis and cell fate. Several approaches have been developed to characterize this crosstalk. Studies of Protein-Bound RNA (PBR) generally require the purification, e.g. by immunoprecipitation, of a specific protein and subsequent sequencing of its RNA cargo^{8,9}. From the protein perspective, the repertoire of RBPs can be examined by UV crosslinking followed by capture of polyadenylated RNAs using oligo(dT) and the identification of the bound proteins¹⁰⁻¹². However, despite their considerable contribution to the understanding of RNA-protein crosstalk, such methods have limitations. Thus, techniques designed to study PBRs are incompatible with a system-wide analysis of RBPs and PBRs, whereas oligo(dT)-based purification requires a large amount of starting material, with high reagent costs, which complicates its application in dynamic conditions¹³. Moreover, the theoretical dependence on the presence of RNA polyA tails, makes oligo(dT)-based capture methods incompatible for organisms with little or no polyadenylation, such as prokaryotes, or for non-polyadenylated RNAs; thus introducing bias in the RBPs identified. There have been recent attempts to address the oligo(dT) limitations, but these have involved incorporation of modified nucleotides, which by itself restricts its applicability and introduces a bias in transcription-dependent nucleoside-incorporation¹⁴⁻¹⁶.

To address these limitations, we have developed a method based on Acidic Guanidinium Thiocyanate-Phenol-Chloroform (AGPC) phase partition, called Orthogonal Organic Phase Separation (OOPS). AGPC purification permits the recovery of RNA species in an unbiased manner^{17,18}. In standard conditions, when lysing cells in AGPC, two distinct phases are formed: RNA migrates to the upper aqueous phase and proteins occupy the lower organic phase. Here, we utilise UV crosslinking at 254 nm to generate RNA-protein adducts that combine the physicochemical properties of both molecules and thus migrate to the aqueous-organic interface¹⁹. Isolation of the interface allows specific recovery of either RBPs or PBRs by digesting the reciprocal component of the adduct.

Here, we demonstrate the specificity and versatility of OOPS. Separation of free and protein-bound RNA provides a way to quantify the proportion of RNA crosslinked to protein, enabling precise optimisation of UV dosage. Furthermore, we show that OOPS recovers all crosslinked-RNA (CL-RNA) and thus all crosslinked RBPs. To demonstrate the versatility of OOPS, we analyse RNA-binding changes through cell-cycle progression following nocodazole arrest. Finally, we apply OOPS to describe the first bacterium RNA-interactome, confirming that OOPS is fully applicable for organisms with limited polyadenylation.

Results

OOPS specifically recovers protein-bound RNA

Cell lysis in acidic Guanidinium-Thiocyanate-Phenol, and subsequent addition of chloroform to the sample, generates two distinct phases: an upper aqueous phase containing RNA and a lower organic phase containing proteins. We reasoned that stable RNA-protein adducts, induced by UV-crosslinking at 254 nm, should therefore be retained at the interface (First step of Figure 1a). CL-RNA was recovered from the interface by protein digestion using proteinase-K and extracting the resulting RNA from the aqueous phase of a subsequent phase separation (Figures 1a-b, online methods). RNA migration from the interface to the aqueous phase after protein digestion indicates that its presence at the interface was dependent on protein binding. Importantly, we observed a dose-dependent depletion of RNA from the aqueous phase (Figures 1b, S1a) with CL-RNA saturating at approximately 75% (Figure 1b). The size profile of CL-RNA resembles total free-RNA of a non-crosslinked sample (NC), with the aqueous phase of the CL sample containing free small RNAs (Figure S1b).

To compare total RNA from a non-crosslinked sample and CL-RNA, rRNA was depleted from both conditions and total RNA-seq carried out, using a dose range of 150-400 mJ/cm² (Figure S1c). The abundance of RNA species observed in the CL-RNA and NC-RNA samples was similar, with protein-coding RNAs predominating (Figures 1c, S1d). Crucially, the Pearson correlation between CL and NC samples is as high as the correlation among crosslinked samples (median correlations are 0.89 and 0.92,

respectively; Figures 1d, S1e), suggesting CL-RNAs fully represent the long RNAs present in a cell and thus, all long RNA species are bound by protein.

Despite the high correlation between RNA abundance in CL and NC samples, we observed an overall reduction of coverage in the 3' UTRs of mRNAs (Figures 1e) and a loss of coverage at discrete positions (Figure 1f). We hypothesized that this was due to steric hindrance of reverse transcription at sites of RNA-protein crosslinking, as protein binding to RNA occurs frequently within the 3' UTR²⁰. To test this, we applied a sliding window approach to identify the sites across the transcriptome where a significant loss of coverage was observed in CL samples (see supplementary notes). As expected, these sites were more frequent in mRNA 3' UTRs and showed a significant overlap with ENCODE eCLIP protein-binding peaks²¹, confirming they represent positions of protein binding (Figure 1g, S1f). Windows with a higher uracil content, which preferentially crosslink to proteins at 254 nm, are more likely to contain a protein-binding site, but adjacent uridines, which are liable to photo-dimerize with 254 nm UV and block reverse transcription²², have no effect (Figure S1g). This further suggests that protein binding is the predominant cause of observed differences in read coverage.

OOPS recovers RNA-binding proteins

Having established that OOPS allows efficient extraction of PBRs, we next identified the proteins crosslinked to RNA. Notably, this required less than 1% of the cells needed in the more established RBP-capture methods^{10,23} (see online methods). First, we

confirmed that the presence of proteins at the interface was dependent on UV-crosslinking. Stable Isotope Labeling by Amino acids in cell Culture (SILAC)²⁴ quantitation was used to determine the relative abundance of proteins from CL and NC U2OS cells in the same OOPS interface (Figure 2a; see online methods). Importantly, repeated phase separation improves the protein CL:NC ratio of RBPs, indicating that additional phase separations remove non-crosslinked proteins. The 3rd interface appears to be optimal, as subsequent phase separations reduce the number of identified proteins (Figures 2b, S2a, supplementary table 1). As expected, since glycosylated proteins share the physicochemical properties of RNA-protein adducts, their presence at the interface is CL-independent. In contrast, non-glycosylated proteins show a similar CL-enrichment, regardless of their GO annotation as RBPs (Figures 2b). This confirms crosslinking enriches known RBPs, and potential undiscovered RBPs, in the interface.

Excluding glycoproteins, 70% of proteins were enriched at the 3rd interface post UV-crosslinking (Figure S2b-e). A similar proportion of proteins were enriched with a lower dosage (150 mJ/cm²; Figure S2e). CL-enriched proteins showed a clear over-representation of RNA-related GO terms (Figure 2c). Remarkably, even within the CL-independent proteins, and accounting for protein abundance, there was a clear over-representation of RNA-binding GO terms (Figure 2d), suggesting that CL-enrichment alone is not sufficient to distinguish free proteins from RNA-bound proteins.

In order to establish that the presence of the proteins at the interface was RNA-dependent, we treated the interfaces with ribonucleases (RNase), and determined protein migration to the organic phase (Figure 2e; online methods, Figures 2f, S2f). Notably, the proteins which migrated to the organic phase included proteins which were CL-independent, suggesting their presence in the interface was RNA-dependent (Figure 2f, S2i). Moreover, proteins not annotated as RBPs show similar RNase sensitivity to those annotated as RBPs, suggesting these are undiscovered RBPs (Figure 2g). In contrast, glycoprotein abundance at the interface was not affected by RNase (Figure S2g). Since the presence of glycoproteins at the interface was also CL-independent (Figure 2b), we excluded them from downstream analyses. Ninety-three percent of proteins observed in the organic phase were RNase sensitive according to their reduction in interface abundances, whereas those missing from the organic phase were largely RNase insensitive (Figure 2h). As expected, 95% of proteins extracted from the organic phase showed an enrichment following RNase treatment (Figure 2i). These proteins show a clear over-representation of GO terms related to RNA binding (Figure S2i). Together, these experiments, demonstrate that RNase treatment is a necessary step to ensure extraction of *bona fide* RBPs. Importantly, these data were replicated in HEK293 cells, confirming cell line-independence (Figure S2c-h).

OOPS identifies canonical and novel RBPs in different cellular compartments

The RBPs identified by OOPS were compared to those from oligo(dT) RBP-Capture analysis. We confirmed a high degree of overlap, with 78% of proteins identified by RBP-capture in U2OS cells also identified by OOPS (Figure 3a, S3a for HEK293 cells).

Furthermore, even amongst the proteins identified by only one method, there was a significant over-representation of GO-annotated RBPs (p-value < 2.2e-16, Fisher's Exact Test). We next applied OOPS to identify RBPs from MCF10A (a cell line derived from a healthy individual), and observed a “common” RBPome of 619 proteins detected in all 3 cell lines (Figure 3b, S3b, supplementary table 2) and a minority of cell line-specific RBPs. A comparison of the 1555 of the proteins from the 3 cell lines used in this study with all previous human RBP-capture data, showed 70% identity (Figure 3c). In addition, OOPS identified 80% of the proteins isolated by polyA-independent RICK¹⁵ and CARIC¹⁴ methods (Figure S3c-d). These results indicate that OOPS is able to recover the majority of the annotated RBPome, including proteins that do not bind polyadenylated RNAs.

As expected, OOPS proteins show a clear over-representation of GO terms describing all forms of RNA-binding, including 5' and 3' UTR binding, and single and double-stranded RNA-binding (Figure 3d, S3e-f)). Interestingly, the novel RBPs identified by OOPS show an over-representation of GO terms related to mRNA transport and RNA localisation (Figure 3e, S3g). We therefore projected OOPS proteins onto our previous hyperLOPIT data²⁵, which identifies the average localisation of proteins, as an initial indication of the subcellular localisation of the RNA-bound fraction. Known RBPs are predominantly localised to the nucleus, mitochondria, cytosol and large protein complexes (e.g ribosomes; Figure 3f), whereas OOPS RBPs appear to be more broadly distributed with a greater proportion of membrane proteins and proteins of unknown localisation (Figure 3f). Since membrane proteins are generally underrepresented in

mass spectrometry experiments, we performed a crude cell fractionation to separate cellular compartments into 3 fractions: “heavy membranes” (e.g. nucleus, mitochondria), “light membranes” (e.g. ER, PM, etc.) and “cytosol” (Figure S3h) . RBPs were detected from all fractions with the membrane fractions yielding more novel membrane-RBPs (Figures 3f, S3h-l, supplementary table 3). Moreover, we confirmed that transmembrane-domain containing RBPs were more abundant in the membrane fractions (Figure S3i). Taken together, our data indicate that combining OOPS with fractionation allows recovery of novel RBPs from previously underrepresented compartments.

High-throughput identification of the RNA-binding sites confirms OOPS RBPs

To validate our novel RBPs and further characterise their RNA binding sites, we developed a method to identify the precise RNA-binding sites, inspired by RBD-map²³. Briefly, proteins from interfaces were sequentially digested to yield the peptide crosslinked to RNA (enriched by TiO₂) and the peptides adjacent to the crosslinked site (See online methods; Figure 4a). Importantly, the RNA-peptide enrichment techniques used were orthogonal to OOPS, providing independent validation as well as yielding RNA-binding sites. Overall, we identified discrete putative RNA-binding sites in 578 (46%) of OOPS U2OS proteins using the adjacent peptides. As expected, putative binding sites were more easily identified in proteins with a higher abundance in the interface, with a binding site identified for 57% of the most abundant novel RBPs (Figure 4b, supplementary table 4). Direct determination of the crosslinked peptide using the

novel Ionbot search engine (see supplementary notes), confirmed a significant overlap between the two approaches (Fig S4a).

To confirm the specificity of our approach, we focused on proteins containing an annotated RNA-recognition motif (RRM), and observed a significant overlap between the sites identified and RRM (Figure 4c). To further test the precision of our RNA-binding site assignment, we examined crystal structures of our detected RBPs in complex with RNA. As an example, the crystal structure of the glycyl-tRNA synthetase in complex with tRNA-Gly²⁶ confirms that the detected binding site contains residues less than 4 Å to the crystallized tRNA (Figure 4d). Additionally, we identified RNA-binding sites in 17 proteins in the ribosome quality control complex structure, together with a novel RBP whose binding site is also within 4 Å of the ribosomal RNA²⁷ (Fig S4b). Finally, we established that our method identifies known RNA-binding domains in GO annotated RBPs, including the canonical RRM and KH domains, and non-canonical helicase C-terminal^{28,29} and DZF²³ domains (Figure 4e). Alongside these non-canonical RNA-binding domains, we identified multiple NAD-binding domains (Figure 4e). These included two sites within the NAD-binding pocket of GAPDH³⁰, which confirmed previous RNA-binding site predictions based on *in vitro* experiments³¹ (Figure 4f).

OOPS facilitates quantification of RNA-binding in a dynamic system

To demonstrate the high yield and versatility of OOPS, we applied it to a dynamic system. U2OS cells were arrested in prometaphase (M) with nocodazole and dynamic changes in RNA-binding were determined following a short (6 h; G1) and long recovery

(23 h; de-synchronized cells; using TMT quantification (Figures 5a & b, S5a and online methods). As expected, we observed changes in the abundance of cell-cycle related proteins between M and the post-release phases (Figure S5b). Quantifying protein abundance in OOPS and total cell lysates from the same sample (Figure 5b), enabled us to detect changes in RNA-binding, which were distinct from concurrent changes in total protein abundance. Interestingly, the changes in OOPS-enriched protein abundance frequently did not correlate with variations in total protein abundance, suggesting specific RBPs bind RNA differentially across the different cell-cycle stages (Figure 5c, supplementary table 5).

To better understand the dynamics of these proteins, we used a linear model framework to identify proteins with apparent changes in RNA-binding, taking into account their total abundance (see online methods). We focused on changes occurring between arrested cells and 6 h release since these were the most synchronised time points. KEGG-pathway³² over-representation analysis indicates that many of the proteins that showed dynamic RNA-binding between these stages were involved in metabolism (Figure 5d). Interestingly, 20/23 tRNA synthetases detected showed an increase in RNA binding post nocodazole release, suggesting a coordinated increase in aminoacyl-tRNA availability (Figure 5e). In addition, we observed an over-representation of proteins involved in glycolysis (Fig5d). Importantly, the presence of glycolysis related proteins in OOPS interfaces was CL-dependent and RNase sensitive in our previous SILAC experiments (Figure S5c). Many of the proteins that function in glycolysis have been described as eukaryotic RNA-binding proteins previously^{12,33,34}. However, this is the

first demonstration of dynamic RNA-binding for these RBPs (Figure 5f), providing a potential model to study the RNA-binding function of these proteins.

OOPS reveals the first RBPome of a non-polyadenylated organism

Since OOPS is not restricted to the study of polyA-RNA binding proteins, we used this method to explore the RBPome of *E. coli* (see online methods for OOPS in bacteria). We detected 397 proteins, which represents 9% of the predicted proteome for the studied K-12 strains³⁵, and is approximately equivalent to the proportion obtained in eukaryotic cells. We recovered 89/172 GO annotated RBPs (Figure 6a, supplementary table 2) and observed that the over-represented GO terms for OOPS proteins are related with the full repertoire of RNA biotypes, including the terms "tRNA aminoacylation", "ribosome assembly" and "ribonucleoprotein complex assembly" (Figure 6b). As expected, the most enriched protein domains are related to nucleic acid binding (Figure 6c). In addition, helicase and P-loop domains, previously described in protein-RNA interactions are also present^{28,29,36}. Furthermore, if we only consider the 308 novel OOPS-RBPs, we still find a clear enrichment of RNA-associated GO-terms, the great majority of them related to tRNAs or ncRNAs (Figure 6d). However, 259 OOPS-RBPs are not annotated with an RNA-related GO term, suggesting OOPS provides a useful approach to study new RBP functions in prokaryotic systems.

Interestingly, many of the glycolytic enzymes that bind RNA in *H. sapiens*, also bind RNA in *E. coli* (Figure 6e). Enolase 1 and Pyruvate kinase, detected by previous RBP-capture studies were identified as RBPs by OOPS in *E.coli*. Furthermore, GAPDH and

293 PKG, previously described as low-confidence candidate RBPs in human by RBP-
 294 Capture, were also found as RBPs in our human and prokaryotic studies. Surprisingly,
 295 phosphoglycerate mutase, a glycolytic protein not previously identified in any human
 296 RBP-capture, was also detected as a RBP in both our eukaryotic and prokaryotic OOPS
 297 experiments. Taken together, the changes in RNA binding detected in our nocodazole
 298 experiment and the consistent identification of particular glycolytic RBPs between
 299 human and *E.coli* may indicate a conserved link between metabolism and RNA binding.

Discussion

OOPS is able to retrieve crosslinked RNAs representing the complete cellular transcriptome and recover crosslinked RBPs, and strongly agrees with data obtained from other RBP identification methods. Importantly, our approach also unveils new RBPs from underrepresented subcellular compartments. We have further identified RNA-binding sites, providing independent evidence of specific RNA-protein interactions. Finally, we have assessed the ability of OOPS to analyse dynamic systems in a straightforward manner and its independence of requirement on a poly-A tail by describing the first bacterial RBPome.

Although OOPS recovers RNAs in an unbiased manner, we observe an under-representation of small RNAs (sRNAs) in the PBR fraction. We believe the most likely explanation is that small RNAs are less frequently protein-bound³⁷. It is also possible that there is a lower probability of UV crosslinking of sRNAs to protein, as they have fewer RNA-protein interactions. Despite the apparent under-representation of sRNAs in our crosslinked interface, we consistently found sRNA-binding proteins in both human and bacteria. Indeed, in *E. coli*, OOPS recovers both canonical (Hfq) and recently discovered (ProQ) sRNA binding proteins³⁸.

OOPS is dependent on separation of macromolecules by their physicochemical properties. Under such a separation, glycoproteins and RNA-protein adducts cannot be distinguished since both glycans and RNAs are hydrophilic polymers. Our observation

that the interface abundance of the vast majority of the glycoproteins is CL-independent and RNase insensitive could indicate that they do not bind RNA. Alternatively, although very few glycoproteins have been described as RBPs, it is also possible they are true RBPs but their glycans maintain them in the interface even after RNase digestion. Despite this, it is interesting to note that 18/21 glycoproteins enriched by CL are localised to the exosome - a RNA-rich compartment³⁹⁻⁴¹ - and include 3 known RNA binding glycoproteins^{10,42}. Yet, to enable a proper interrogation of RNA-binding glycoproteins, it would be necessary to remove glycans and achieving this in a manner that does not degrade RNA is non-trivial. We therefore took a conservative approach and discounted glycoproteins from our analyses.

Crosslinking-based detection of RBPs is based on proximity of RNAs and proteins. The functional relevance of some catalogs may therefore need to be considered with caution¹⁹. Dynamic experiments, such as the one presented here, can facilitate interrogation of the biological function of the RNA-protein interactions and uncover system-wide changes in RNA-binding proteins. One of the most striking findings is the coordinated increase in RNA-binding of glycolytic enzymes following release from nocodazole arrest. Considering the previously described regulation of glycolytic protein thermal stability in response to nocodazole arrest⁴³, and the reported repression of translation by GAPDH in response to changes in glycolytic flux⁴⁴, our results suggest that an interplay between metabolism and RNA-binding may be a fundamental component of the cell cycle. GAPDH has been shown to bind to a range of RNA species including tRNAs, AU-rich elements, and TERC^{45,46}. *In vitro* competition assays and spectrometry

models suggest binding occurs within its NAD-binding crevice, but this has not been observed *in vivo*^{31,45,47}. Here, we provide the first *in vivo* evidence, which shows that the RNA-binding site of GAPDH is indeed the NAD-binding crevice.

In summary, OOPS is a low-cost method which allows the isolation of RNA-protein complexes in any organism, enabling downstream analysis of either the RNA or protein component. This simple method will make the study of RNA-protein interactions more accessible and foster a systems biology view of their function by permitting the study of their dynamic properties.

Author Contributions

KSL, AW, EV, TS and RQ conceived the study. EV, RQ, TS and KSL designed the experiments. EV optimised the initial OOPS protocol, prepared RNA-seq libraries and performed flow cytometry analysis. EV performed the SILAC, cellular subfractionation and nocodazole arrest experiments with assistance from RQ. *E. coli* experiments were performed by MM, EV and RQ. U2OS RBP-Capture was performed by MP, VD. EV and RQ performed all additional experiments including the RNA binding site experiment. RQ performed all mass spectrometry. TS performed all data analysis, with the exception of the analysis of uridine content (DM) and analysis of *E. coli* data (TS, MM). EV, TS, RQ and KSL interpreted results, with critical appraisal of findings from AW, MP, MR, RH and VD, including additional experiments (MP). GT assisted with interpretation of *E. coli* data. MM-S performed the protein-RNA structural analysis. SD and LM developed the ionbot search engine and performed ionbot searches. TS, EV, RQ, KSL and MM-S drafted the manuscript, with critical revision from AW, RH, GT, DM, MP, MR and SD.

Acknowledgments

EV, TS, RQ, RH, MP, MR, VD are supported by Medical Research Council, Grant/Award number: 5TR00; Wellcome Trust, Grant/Award numbers: 110170/Z/15/Z, 110071/Z/15/Z. MM-S is supported by the MRC (MC_U105185859) and by a FEBS Long-Term Fellowship. GT is supported by IB Catalyst grant for Project DETOX (BB/N01040X/1). We would like to thank Harriet T. Parsons for donating *E. coli* cells, Mohammed A. Elzek for culturing MCF10A cells, Tom Mulroney for helping culturing U2OS cells and Bettina Fisher for kindly sharing equipment.

References

1. García-Mauriño, S. M. *et al.* RNA Binding Protein Regulation and Cross-Talk in the Control of AU-rich mRNA Fate. *Front. Mol. Biosci.* **4**, 71 (2017).
2. Müller-Mcnicoll, M. & Neugebauer, K. M. How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *RNA Biol.* **14**, 275–287 (2013).
3. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12**, 99–110 (2011).
4. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
5. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–126 (2011).
6. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
7. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
8. Hafner, M. *et al.* PAR-CLIP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* 2–6 (2010). doi:10.3791/2034
9. Huppertz, I. *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–87 (2014).
10. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).

11. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* **46**, 674–690 (2012).
12. Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **6**, (2015).
13. Sysoev, V. O. *et al.* Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*. *Nat. Commun.* **7**, 12128 (2016).
14. Huang, R., Han, M., Meng, L. & Chen, X. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3879–E3887 (2018).
15. Bao, X. *et al.* Capturing the interactome of newly transcribed RNA. *Nat. Methods* **15**, (2018).
16. Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15779–84 (2008).
17. Chomczynski, P. Single-Step Method of RNA Isolation by Acid Guanidinium Extraction. *Anal. Biochem.* **159**, 156–159 (1987).
18. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: Twenty-something years on. *Nat. Protoc.* **1**, 581–585 (2006).
19. WAGENMAKERS, A. J. M., REINDERS, R. J. & VAN VENROOIJ, W. J. Cross-linking of mRNA to Proteins by Irradiation of Intact Cells with Ultraviolet Light. *Eur. J. Biochem.* **112**, 323–330 (1980).
20. Harvey, R. F. *et al.* *Trans* -acting translational regulatory RNA binding proteins. *Wiley Interdiscip. Rev. RNA* **9**, e1465 (2018).

21. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
22. Kladwang, W., Hum, J. & Das, R. Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. *Sci. Rep.* **2**, 1–7 (2012).
23. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **63**, 696–710 (2016).
24. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–86 (2002).
25. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* (80-.). **356**, eaal3321 (2017).
26. Qin, X. *et al.* Cocystal structures of glycyl-tRNA synthetase in complex with tRNA suggest multiple conformational states in glycylation. *J. Biol. Chem.* **289**, 20359–69 (2014).
27. Shao, S., Brown, A., Santhanam, B. & Hegde, R. S. Structure and Assembly Pathway of the Ribosome Quality Control Complex. *Mol. Cell* **57**, 433–444 (2015).
28. Yang, L., Yang, J., Huang, Y. & Liu, Z.-R. Phosphorylation of p68 RNA helicase regulates RNA binding by the C-terminal domain of the protein. *Biochem. Biophys. Res. Commun.* **314**, 622–630 (2004).
29. Ranji, A., Shkriabai, N., Kvaratskhelia, M., Musier-Forsyth, K. & Boris-Lawrie, K. Features of Double-stranded RNA-binding Domains of RNA Helicase A Are Necessary for Selective Recognition and Translation of Complex mRNAs * □ S.

(2010). doi:10.1074/jbc.M110.176339

30. White, M. R. *et al.* A dimer interface mutation in glyceraldehyde-3-phosphate dehydrogenase regulates its binding to AU-rich RNA. *J. Biol. Chem.* **290**, 1770–85 (2015).

31. Singh, R. & Green, M. Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. *Science (80-.)*. **259**, 365–368 (1993).

32. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).

33. Castello, A., Hentze, M. W. & Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol. Metab.* **26**, 746–757 (2015).

34. Liao, Y. *et al.* The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Rep.* **16**, 1456–1469 (2016).

35. Anton, B. P. & Raleigh, E. A. Complete Genome Sequence of NEB 5-alpha, a Derivative of Escherichia coli K-12 DH5α. *Genome Announc.* **4**, e01245-16 (2016).

36. Wei, R. R. & Richardson, J. P. Identification of an RNA-binding Site in the ATP binding domain of Escherichia coli Rho by H₂O₂/Fe-EDTA cleavage protection studies. *J. Biol. Chem.* **276**, 28380–7 (2001).

37. Plochowitz, A., Farrell, I., Smilansky, Z., Cooperman, B. S. & Kapanidis, A. N. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria.

Nucleic Acids Res. **45**, 926–937 (2017).

38. Smirnov, A. *et al.* Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11591–11596 (2016).
39. Wei, Z. *et al.* Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat. Commun.* **8**, 1145 (2017).
40. Batagov, A. O. & Kurochkin, I. V. Exosomes secreted by human cells transport largely mRNA fragments that are enriched in the 3'-untranslated regions. *Biol. Direct* **8**, 12 (2013).
41. Kanada, M. *et al.* Differential fates of biomolecules delivered to target cells via extracellular vesicles. *Proc. Natl. Acad. Sci.* **112**, 201418401 (2015).
42. Watanabe, A. *et al.* Raftlin Is Involved in the Nucleocapture Complex to Induce Poly(I:C)-mediated TLR3 Activation. *J. Biol. Chem.* **286**, 10702–10711 (2011).
43. Becher, I. *et al.* Pervasive Protein Thermal Stability Variation during the Cell Cycle. *Cell* 1–13 (2018). doi:10.1016/j.cell.2018.03.053
44. Chang, C.-H. *et al.* Posttranscriptional Control of T Cell Effector Function by Aerobic Glycolysis. *Cell* **153**, 1239–1251 (2013).
45. Garcin, E. D. GAPDH as a model non-canonical AU-rich RNA binding protein. *Semin. Cell Dev. Biol.* 1–12 (2018). doi:10.1016/j.semcdb.2018.03.013
46. White, M. R. & Garcin, E. D. The sweet side of RNA regulation: Glyceraldehyde-3-phosphate dehydrogenase as a noncanonical RNA-binding protein. *Wiley Interdiscip. Rev. RNA* **7**, 53–70 (2016).
47. Carmona, P., Rodríguez-Casado, A. & Molina, M. Conformational structure and binding mode of glyceraldehyde-3-phosphate dehydrogenase to tRNA studied by

494 Raman and CD spectroscopy. *Biochim. Biophys. Acta - Protein Struct. Mol.*
495 *Enzymol.* **1432**, 222–233 (1999).

Methods

Cell culture:

U-2 OS (U2OS) and MCF 10A cells were obtained from the American Type Culture Collection (ATCC). HEK-293 were kindly provided by Dr. Johanna Rees (University of Cambridge). U2OS and HEK-293 cells were cultured in McCoy's 5A and DMEM (Gibco-BRL) media respectively, supplemented with 10% fetal bovine serum (Gibco-BRL). MCF 10A were maintained in MEBM media (Lonza/Clonetics) supplemented with 10 ng/ml of cholera toxin (Sigma-Aldrich). All cells were maintained at 37 °C and 5% CO₂ and regularly tested for mycoplasma contamination with negative results.

E. coli K-12 DH5a strain (Thermo Fisher Scientific), was cultured in LB Broth (Thermo Fisher Scientific) at 37 °C. All E. coli experiments were done at stationary phase after 16 h of cell growth.

Orthogonal Organic Phase Separation in human cells:

Cells were cultured in 6 cm diameter dishes (28.2 cm²) for catalog experiments, or 10 cm diameter dishes (78.5 cm²) for dynamic experiments, until a maximum of 90% of confluence was reached, using a single dish per replica and condition. Cells were washed twice with PBS and supernatant removed by pipetting. In non-crosslinked controls, cells were immediately lysed by scrapping in Acidic Guanidinium Thiocyanate-Phenol (Trizol, Thermo Fisher Scientific), and the homogenate transferred to a new tube. In crosslinked samples, UV-crosslinking was performed on PBS-washed cells by UV-irradiation at 254 nm (CL-1000

Ultraviolet Crosslinker; UVP). Immediately after crosslinking, cells were scraped in Trizol and the homogenized lysate was transferred to a new tube and incubated at room temperature (RT) for 5 min to dissociate unstabilised RNA-protein interactions. For biphasic extraction, 200 μ L of chloroform (Fisher Scientific) were added, phases were vortexed and centrifuged for 15 min at 12,000 x g at 4 °C. The upper aqueous phase (containing non-crosslinked RNAs) was transferred to a new tube, and RNA precipitated following manufacturer instructions. The lower organic phase (containing non-crosslinked proteins) was transferred to a new tube and proteins precipitated by addition of 9 volumes of methanol (Fisher Scientific). Interface (containing the Protein-RNA adducts) was subjected to extra AGPC phase separation cycles, precipitated by addition of 9 volumes of methanol, and pelleted by centrifugation at 14,000 x g, RT for 10 min.

For RNA analyses, the precipitated interfaces were incubated for 2h at 50 °C in 30 mM Tris HCl (pH8)/10 mM EDTA and 18 Us of proteinase K (Thermo Fisher Scientific). Samples were cooled and released RNA was purified by standard phenol/chloroform extraction (Thermo Fisher Scientific) according to the manufacturer instructions.

For RNA-binding protein analyses, the precipitated interface was resuspended in 100 μ L of 100 mM TEAB, 1 mM $MgCl_2$, 1% SDS, incubated at 95 °C for 20 min, cooled down and digested with 2 μ g RNase A, T1 mix (2 mg/mL of RNase A and 5000 U/mL of RNase T1, Thermo Fisher Scientific) for 2-3 h at 37 °C. Another 2 μ g of RNase mix was added and incubated overnight at 37 °C, after which a final cycle of AGPC phase partitioning was performed and released proteins recovered from the organic phase by methanol precipitation.

Orthogonal Organic Phase Separation in bacteria:

E. coli cultures were grown overnight. 3 ml of culture was pelleted by centrifugation (5 min at 6000 g, RT) and washed twice with PBS. Cells were resuspended in PBS and crosslinked in solution at 254 nm for 525 mJ/cm². Crosslinked cells were pelleted again and supernatant removed by pipetting, leaving approximately 50 µl of PBS. 500 µl of 0.5 mm glass beads (Sigma-Aldrich) were added to each sample, mixed gently, frozen on dry ice and dried by sublimation for 2 h. Dried cells were disrupted by vortexing for 5 min, at intervals of 1 min to avoid warming the sample. 1 ml Trizol was added to each tube and samples were homogenized by vortexing. Supernatant (avoiding glass beads) was transferred to a new tube and centrifuged 5 min at 6000 x g at 4 °C. The supernatant was transferred to a new tube, leaving the unlysed cells as a pellet. Finally, OOPS was performed as described above.

RNA quantification and integrity assessment:

RNA purity was assessed by Nanodrop (Thermo Fisher Scientific). Samples with a 260/280 ratio below 1.9 or 260/230 below 2 were discarded. RNA concentration was estimated using the Qubit RNA BR (Broad-Range) Assay Kit (Thermo Fisher Scientific) in the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). RNA integrity was evaluated using the Agilent 2100 Bioanalyzer system (Agilent).

RNA sequencing:

Protein Bound RNA (PBR) and total non-crosslinked (NC) RNA were purified using OOPS or standard Trizol extraction respectively. All RNA samples were treated with turbo DNase (Thermo Fisher Scientific). Ribosomal RNA (rRNA) was depleted using RiboCop kit V1.2

(Lexogen, Greenland, NH, USA) according to manufacturer instructions, starting with 1ug of RNA. Two nanograms of rRNA-depleted NC-RNA or 8 ng of rRNA-depleted PBR were used to generate sequencing libraries using SENSE total RNA-Seq Library Prep kit (Lexogen). All libraries were sequenced in parallel on a NextSeq 500 for 75 cycles (Illumina).

RNA-Seq data processing and bioinformatics

Quality control of raw fastqs was performed using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the hg38 human genome and Ensembl 87⁴⁸ using hisat2⁴⁹ with default settings and reads with MAPQ < 10 were discarded. Transcript quantification was performed with Salmon⁵⁰ using default settings. The meta-plot of read coverage over gene model was obtained using the CGAT bam2geneprofile script with *reporter=utrprofile*⁵¹. For details of the identification of putative RNA binding sites and the overlap with eCLIP data, see supplementary notes.

Oligo(dT) RBP-capture:

RBP-Capture was performed according to²³, with the following modifications. We used 4 x 500cm² plates per condition. Oligo(dT)25 magnetic beads (NE Biolabs) were reconditioned as per manufacturer's instructions and incubated with the lysates for a second round of RBP-capture with eluates from the two rounds were pooled together

Subcellular fractionation:

U2OS cells from a single 80% confluent 500 cm² cell culture dish (Sigma-Aldrich) were detached using trypsin without EDTA (Thermo Fisher Scientific), pelleted 5 min at 250 g,

washed with PBS, resuspended in 50 ml of PBS and crosslinked in solution at 254 nm at 400 mJ/cm². Cells were pelleted again for 5 min at 250 x g, resuspended in 1 ml of lysis buffer (0.25 M sucrose, 10 mM HEPES pH 7.4) containing protease inhibitors (Roche), and lysed with a ball-bearing homogenizer (Isobiotec) on ice. Unlysed cells were removed by centrifugation at 200 x g, 5 min at 4 °C. The supernatant was transferred to a new tube and centrifuged at 1000 x g, 10 min at 4 °C with the pellet collected as 'heavy membrane fraction'. The supernatant was centrifuged again at 12.200 g with the pellet collected as the 'light membrane fraction'. The supernatant was collected as cytosolic fraction, frozen and dried by sublimation by SpeedVac (Labconco). Pellets from the heavy membranes, light membranes and cytosol were re-suspended in Trizol and RBPome and "total" proteome were extracted using OOPS.

Nocodazole synchronisation:

A single 10 cm² diameter dish (per replica and condition) of U2OS cells at 70 % of confluence was arrested in prometaphase by direct addition of 1 µg/ml of nocodazole (Sigma-Aldrich) to the cell culture media. 16-18 hours post treatment, synchronised cells were washed twice in PBS and crosslinked at 254 nm at 400 mJ/cm². Arrested cells were detached by mechanical stimulation, pelleted, solubilised in Acidic Guanidinium Thiocyanate-Phenol (Trizol) and stored at -80 °C. For post-release timepoints (6 h and 23 h post-arrest), synchronised cells were detached from the dish by mechanical stimulation, washed in PBS and re-seeded in media without nocodazole. Cells were then washed twice with PBS and crosslinked at 254 nm at 400 mJ/cm². Cell lysates were obtained by directly scraping the crosslinked cells in Acidic Guanidinium Thiocyanate-Phenol (Trizol). The total proteome was extracted from the lysate

and the RBPome was determined using OOPS (see Orthogonal Organic Phase Separation in human cells). A parallel cell dish was cultured for every time point and replicate to assess the synchronisation efficacy by flow cytometry. DNA content per cell was analysed using the Propidium Iodide Flow Cytometry Kit (Abcam) as indicated by the manufacturer. Flow cytometry results were analysed using FlowJo 8.7.

Proteomic sample preparation:

Samples were resuspended in 100 µL of 100 mM Triethylammonium bicarbonate (TEAB) (Sigma-Aldrich), reduced with 20 mM DTT (Sigma-Aldrich) at room temperature for 60 min and alkylated with 40 mM iodoacetamide (Sigma-Aldrich) at room temperature in the dark for at least 60 min. Samples were digested overnight at 37 °C with 1 µg of Trypsin (Promega) with the exception of samples for TMT labeling which were digested overnight at 37 °C with 1 µg Lys-C (Promega). Subsequently, 1 µg of modified trypsin (Promega) was added, and the samples were incubated for 3-4 h at 37 °C. Samples were then acidified with TFA (0.1% (v/v) final concentration; Sigma-Aldrich) and centrifuged at 21,000 g for 10 min, with the supernatant frozen at -80 °C until required.

For peptide clean-up and quantification, 200 µL of Poros Oligo R3 (Thermo Fisher Scientific) resin slurry (approximately 150 µL resin) was packed into Pierce™ Centrifuge Columns (Thermo Fisher Scientific) and equilibrated with 0.1% TFA. Samples were loaded, washed twice with 200 µL 0.1% TFA and eluted with 300 µL 70% acetonitrile (ACN) (adapted from⁵²). 10 µL was taken from each elution for Qubit™ protein assay (Thermo Fisher Scientific) quantitation, with the remaining sample retained for MS.

631

632 **LC-MS/MS**

633 SILAC labelling was performed according to the manufacturer's instructions by growing cells in
634 DMEM media containing light (Arg0-Lys0) or heavy (Arg10-Lys8) isotopes (SILAC Protein
635 Quantitation Kit, Thermo Fisher Scientific). SILAC and unlabeled samples generated from
636 OOPS experiments were analysed in the Orbitrap Fusion Lumos (Thermo Fisher Scientific)
637 coupled to a nanoLC Dionex Ultimate 3000 UHPLC (Thermo Fisher Scientific). Mass spectra
638 were acquired using CHarge Ordered Parallel Ion aNalysis (CHOPIN) acquisition in positive
639 ion mode as previously reported⁵³

640 TMT-10plex (Thermo Fisher Scientific) labelling from desalted peptides was performed
641 according to the manufacturer's protocol. Equal amounts of desalted peptides were labelled
642 immediately after being quantified with QubitTM protein assay (Thermo Fisher Scientific).
643 Multiplexed TMT samples were separated into 4 fractions using PierceTM High pH Reversed-
644 Phase Peptide Fractionation Kit (Thermo Fisher Scientific). TMT-10plex fractions were
645 analysed in an Orbitrap Fusion Lumos. Mass spectra were acquired in positive ion mode
646 applying data acquisition using synchronous precursor selection MS³ (SPS-MS³) acquisition
647 mode⁵⁴.

648 Samples from Oligo(dT) capture and from subcellular fractionation were analysed in an
649 Orbitrap nano-ESI Q-Exactive mass spectrometer (Thermo Fisher Scientific), coupled to a
650 nanoLC (Dionex Ultimate 3000 UHPLC).

651

652 **Peptide spectrum matching**

Raw data were viewed in Xcalibur v.2.1 (Thermo Fisher Scientific), and data processing was performed using Proteome Discoverer v2.1 (Thermo Fisher Scientific). The Raw files were submitted to a database search using Proteome Discoverer with Mascot, SequestHF and MS Amanda algorithm⁵⁵ against the *Homo sapiens* database for U2OS and MCF 10A cells or *E. coli* database, downloaded in early 2017 containing human (or *E. coli*) protein sequences from UniProt/Swiss-Prot and UniProt/ TrEMBL. Common contaminant proteins (several types of human keratins, BSA, and porcine trypsin) were added to the database, and all contaminant proteins identified were removed from the result lists prior to further analysis. The spectra identification was performed with the following parameters: MS accuracy, 10 ppm; MS/MS accuracy, 0.5 Da for samples run with Orbitrap Fusion Lumos and MS accuracy, 10 ppm; MS/MS accuracy, up to 0.2 Da for samples run with Q-Exactive mass spectrometer; up to two missed cleavage sites allowed; carbamidomethylation of cysteine (as well as TMT6plex tagging of lysine and peptide N-terminus for TMT labeled samples) as a fixed modification; and oxidation of methionine and deamidated asparagine and glutamine as variable modifications. Percolator node was used for false discovery rate estimation and only rank 1 peptide identifications of high confidence (FDR < 1 %) were accepted. A minimum of two high confidence peptides per protein was required for identification using Proteome Discoverer. TiO₂-enriched fractions intended for direct assessment of RNA-crosslinked site in proteins were searched with beta-testing version *Ionbot* search engine (see Supplementary notes) with carbamidomethylation of cysteine and oxidation of methionine respectively as a fixed and variable modifications, *Homo sapiens* Swissprot database and both *Extended Modification Search* and *RNA Binding Modification Search* enabled.

Direct assessment of RNA crosslinking site in proteins:

Starting from the precipitated OOPS interface, proteins were digested using 1 µg Lys-C (Promega, Madison, WI, USA) in 100 µL of 100 mM TEAB (Sigma-Aldrich) with 1 µL of RNaseOUT (Thermo Fisher Scientific) overnight at 37 °C. Two different approaches were used to enriched RNA-peptides:

(i) Silica-based RNA purification using the RNeasy kit (Qiagen), according with the manufacturer's instructions;

(ii) Precipitation in 80% ethanol. Two rounds of precipitations were used to further clean the sample.

RNA-peptides were re-suspended in 100 µL of 100mM Tris-HCl (pH8.0)/ 2 mM MgCl₂, sonicated for 15 min and incubated at 95 °C for 20 min. 2 µg RNase A/T1 mix (2 mg/mL of RNase A and 5000 U/mL of RNase T1) was added to cooled samples, and incubated for 4 h at 37 °C followed by a second protease digestion using 1 µg trypsin (Promega) overnight at 37 °C. Digested samples were desalted with Oligo R3 as described in the “proteomics sample preparation” section and dried on speedvac (Labconco).

Digests were re-suspended in 30-40 µL of 80% acetonitrile (ACN)/2% TFA containing 1 µg of TiO₂ beads (GL Sciences). The slurry was transferred into a p200 tip containing a C8 “plug” (3M Empore, Sigma-Aldrich) to retain the loaded TiO₂ beads and the flow-through collected. The packed TiO₂ was washed with 20 µL 80% ACN/2% TFA, then 20 µL 10% ACN/0.1% TFA and the flow-through from both retained. The TiO₂-enriched fraction was eluted from the beads with two rounds of 20 µL of Ammonia solution (1.5-1.8%), pH>10.5, and 20 µL of 50% ACN.

697

698 **Proteomics bioinformatics and data analysis**

699 Peptide-level output from Proteome Discoverer was re-processed with the
700 add_master_protein.py script (<https://github.com/TomSmithCGAT/CamProt>) to ensure uniform
701 peptide to protein assignment for all samples from a single experiment and identify peptides
702 which are likely to originate from contaminating proteins such as keratin (see supplementary
703 notes). For quantitative experiments, peptide-level quantification was obtained by summing the
704 quantification values for all peptides with the same sequence but different modifications.
705 Protein-level quantification was then obtained by taking the median peptide abundance. For
706 SILAC experiments, the ratio between treatment and control protein abundance was calculated
707 for each sample separately and aggregated to average protein ratio. For TMT experiments,
708 data analysis was performed using the *MSnbase* R package⁵⁶. Log₂-transformed protein
709 abundance was centre-median normalised within each sample. For the crude fractionation
710 experiments (*n*=5), the protein abundance was quantified by label-free quantification, averaged
711 across the replicates per fraction and normalised per protein such that the sum of abundances
712 over the 3 fractions was 1. For the U2OS RBP-Capture experiment, only proteins observed in
713 all 3 CL replicates and no NC replicates were retained. In crosslink-testing SILAC
714 experiments, only proteins present in at least 2 replicates were retained.

715

716 GO terms, Interpro protein domains and KEGG pathway annotations were obtained using the
717 R package *UniProt.ws*⁵⁷. GO terms were expanded to include all parent terms using the R
718 package *GO.db*⁵⁸. Glycoproteins were identified using the Uniprot⁵⁹ API with *categories=PTM*

and *types=CARBOHYD*. Transmembrane proteins were identified using the Uniprot API with *types=TRANSMEM*.

Data handling and statistics

Data handling was performed with R v3.4.1 using tidyverse packages and python v3.6.5.

Plotting was performed with the *ggplot2* R package⁶⁰.

Proteins observed only in CL in at least one replicate were deemed enriched. For the RNase-testing SILAC experiments, proteins only ever observed in the RNase condition at the organic phase were deemed enriched. Vis versa, those only ever observed in the control condition at the interfaces were deemed depleted. For proteins which did not meet these criteria, all peptides observed across the replicates were treated as independent observations and a Mann-Whitney-Wilcoxon Test was used to test whether the log₂ median CL:NC or RNase:Control ratio was > 0 (enriched) or < 0 (depleted), with a BH-adjusted p-values < 0.05 considered significant. Proteins with less than 6 peptides were excluded from the statistical test due to insufficient power.

GO, InterPro and KEGG over-representation analyses were conducted using the R package goseq to account for protein abundance. p-values were adjusted to account for multiple testing using the Benjamini-Hochberg⁶¹ FDR procedure. GO-terms with adjusted p-value <0.01 and at least 5 proteins were considered significantly over-represented. Over-representation values given are not adjusted for protein abundance. For U2OS and HEK-293, protein abundance was derived from⁶² taking the maximum abundance recorded across the replicates. For

MCF10A, we used an in-house deep proteomics data set. For *E.coli*, protein abundance was obtained from PaxDB⁶³.

For the nocodazole arrest/release experiment, proteins with a change in abundance or RNA binding were identified using the *lm* function in R. Specifically, to identify protein with a change in abundance between nocodazole arrest and 6 h release, total protein abundance was modelled as a function of the time point alone (abundance ~ timepoint). The p-values for the timepoint coefficients for each proteins were adjusted to account for multiple hypothesis testing according to Benjamini-Hochberg⁶¹ and proteins with an adjusted p-value < 0.01 (1 % FDR) were considered to have changed abundance. To identify proteins with a change in RNA binding between nocodazole arrest and 6 h release, protein abundance in the total proteome and OOPS samples was modelled as a function of the time point, the abundance type (total or OOPS), and the interaction between these two variables (abundance ~ timepoint + type + timepoint*type). Here, the interaction term denotes whether the abundance in OOPS and total follows the same pattern across the timepoints (coefficient is zero), indicating total abundance determines the amount of protein bound to RNA, or diverges (non-zero coefficient), indicating a change in RNA binding between the timepoints. The p-values for the interaction term were obtained and adjusted as indicated above. For the heatmap representation, protein abundances were z-score normalised within the total and OOPS samples separately. Hierarchical clustering was performed with the R *hclust* function using 1-Spearman's rho as the distance metric and average linkage.

For details of the identification of RNA binding sites see supplementary notes.

Structural Assessment of RNA-protein contacts

763 In order to look for structural information to validate our direct evidence for RNA-protein
 764 contacts, the Uniprot IDs of the detected proteins were used to retrieve all their associated
 765 PDB IDs using the Uniprot Retrieve/ID mapping tool. In parallel, we retrieved information for all
 766 structures annotated as containing protein-RNA complexes in the nucleic acid database⁶⁴.
 767 Comparison of PDB IDs common in both subsets revealed the structures of the ribosome
 768 quality control complex (PDB ID 3J92) and of a Glycyl-tRNA synthetase in complex with tRNA-
 769 Gly (PDB ID 4KR2). These structures, together with the structure of GADPH in complex with
 770 NAD (PDB ID 4WNC), were later visualized using VMD 1.9.4⁶⁵.

771

772

Methods references

48. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
49. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–60 (2015).
50. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
51. Sims, D. *et al.* CGAT: computational genomics analysis toolkit. *Bioinformatics* **30**, 1290–1 (2014).
52. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. & Roepstorff, P. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **34**, 105–116 (1999).
53. Davis, S. *et al.* Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis. *J. Proteome Res.* **16**, 1288–1299 (2017).
54. McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
55. Dorfer, V. *et al.* MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* **13**, 3679–84 (2014).

- 796 56. Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged
797 mass spectrometry data visualization, processing and quantitation. *Bioinformatics*
798 **28**, 288–289 (2012).
- 799 57. M, C. UniProt.ws: R Interface to UniProt Web Services. R package version 2.20.0.
800 (2018). doi:10.18129/B9.bioc.UniProt.ws
- 801 58. M, C. GO.db: A set of annotation maps describing the entire Gene Ontology. R
802 package version 3.6.0. (2018). doi:10.18129/B9.bioc.GO.db
- 803 59. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic*
804 *Acids Res.* **46**, 2699 (2018).
- 805 60. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New
806 York, 2009).
- 807 61. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical
808 and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef
809 Hochberg Source : Journal of the Royal Statistical Society . Series B (
810 Methodological), Vol . 57 , No . 1 Published by : J. R. Stat. Soc. Ser. B **57**, 289–
811 300 (1995).
- 812 62. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic
813 analysis of eleven common cell lines reveals ubiquitous but varying expression of
814 most proteins. *Mol. Cell. Proteomics* **11**, M111.014050 (2012).
- 815 63. Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C.
816 Version 4.0 of PaxDb: Protein abundance data, integrated across model
817 organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).
- 818 64. Coimbatore Narayanan, B. *et al.* The Nucleic Acid Database: new features and

819 capabilities. *Nucleic Acids Res.* **42**, D114-22 (2014).

820 65. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol.*

821 *Graph.* **14**, 33–8, 27–8 (1996).

822

823

Figures

Figure 1. OOPS unbiasedly recovers all PBR species.

(a) Schematic representation of the OOPS method to extract protein-bound RNA

(b) Relative proportions of free RNA (aqueous phase) and protein-bound RNA (PBR; interface)

with increasing UV dosage. Data shown as mean +/- SD of 3 independent experiments

(c) Relative proportions of reads assigned to Ensembl gene biotypes for 400 mJ/cm² CL and NC

samples

832 (d) Correlation between NC replicate 1 and 400 mJ/cm² CL replicate 1. Blue dashed lines
 833 represent a 10-fold difference
 834 (e) Meta-plot of read coverage over protein-coding gene-model
 835 (f) Read coverage across the 3' end of ATCB for CL (400 mJ/cm²) and NC replicates. Red
 836 boxes denote regions with consistently reduced coverage in CL
 837 (g) Relationship between the number of eCLIP proteins with a peak in a sliding window and the
 838 probability of the window being identified as a protein binding site
 839 Non-crosslinked=NC, Crosslinked=CL
 840

Figure 2. OOPS specifically recovers RBPs.

(a) Schematic representation of the SILAC experiment used to determine the effect of UV crosslinking on protein abundance in the interface and the effect of additional phase separation cycles to wash the interface

(b) Protein CL vs NC ratios for the 1st to 4th interfaces

(c) Top 10 molecular function GO terms over-represented in proteins enriched by CL in the 3rd interface

(d) Top 10 molecular function GO terms over-represented in proteins not enriched by CL in the 3rd interface

(e) Schematic representation of a SILAC experiment to determine protein abundance in the 3rd interface and 4th organic phase following RNase treatment

853 (f) Protein CL vs NC ratio and RNase vs control ratio in the interface. Red box denotes proteins
 854 which are not CL-enriched and not depleted by RNase. The regions around the graph in blue
 855 denote ratios which cannot be accurately estimated as they are $> 2^{10}$ or $< 2^{-10}$, or the protein
 856 was only detected in one condition

857 (g) RNase vs control ratio in the interface for GO annotated RBPs and other OOPS RBPs

858 (h) Protein RNase vs control ratio in the RNase treated interfaces for proteins identified in the
 859 4th step organic phase

860 (i) Proportion of proteins enriched in the organic phase following RNase treatment

861

Figure 3. OOPS recovers known and new RBPs, even from under-represented cell compartments.

(a) Overlap between OOPS, RBP-Capture and GO-annotated proteins for U2OS cells. Proteins were restricted to those expressed in U2OS

(b) Overlap between proteins identified with OOPS from U2OS, HEK293 and MCF10A. Proteins were restricted to those expressed in all cell lines

(c) Overlap between the union of OOPS proteins identified in the 3 cell lines, all published RBP-Capture studies, and GO annotated RBPs. Proteins were restricted to those expressed in at least one of the three OOPS cell lines

(d) Top 10 molecular function GO terms over-represented in the proteins identified in U2OS, HEK293 or MCF10A OOPS

(e) Top 10 biological process GO terms over-represented in novel RBPs identified by OOPS

(f) HyperLOPIT projections of protein localisation. Canonical subcellular localisation markers indicated in colour as shown. NA: non-marker protein. Highlighted RBPs shown as black asterisks. Light membrane-enriched fraction: Lm, Cytoplasm/Other fraction: C/o

Figure 4. Direct assessment of crosslink sites validates most OOPS RBPs.

(a) Schematic representation of the sequential digestion method used to identify the RNA-binding site. Red=peptides containing site of crosslinking. Green=peptides adjacent to the RNA-binding site peptide

(b) Proportion of OOPS proteins in which the RNA-binding site was identified. Proteins separated into GO annotated RBPs and novel RBPs, and by their abundance at the OOPS interface

(c) Distance of putative RNA-binding sites to the nearest RRM. Smaller putative RNA-binding sites are closer to RRM. Analysis restricted to proteins with an RRM

(d) Crystal structure of Glycyl-tRNA synthetase in complex with tRNA-Gly (PDB ID 4KR2). RNA is shown as transparent lime ribbon, while the Glycyl-tRNA synthetase is shown in a cyan transparent cartoon representation. The protein region detected by direct evidence is shown in an opaque representation and RNA and protein residues at 4 Å or less from each other are shown as lime and cyan sticks respectively

(e) Domains containing a putative RNA-binding site

895 (f) Crystal structure of GAPDH complexed with NAD (PDB ID 4WNC). GAPDH is shown as a
896 cyan transparent cartoon and the protein region detected by direct evidence is shown in an
897 opaque representation. Residues at 4 Å or less from NAD (yellow sticks and surface
898 representation) are shown as cyan sticks
899

Figure 5. Dynamic characterization of the RBP-ome after nocodazole arrest.

(a) Left: schematic representation of the nocodazole arrest/release experiment. Cells were analysed after 18 h nocodazole arrest and after a 6 h or 23 h release from the treatment release. Right: relative proportions of cells in G1, S and M phase for cells synchronised at each time-point (shown as the mean $\pm$ SD of 3 independent experiments)

(b) Schematic representation of protein extraction for nocodazole-arrest experiment. Total proteomes were extracted from cell lysates and RBPs were extracted following OOPS proteome method

(c) Protein abundance from total proteome and OOPS extractions. Abundance z-score normalised within each extraction type. Proteins hierarchically clustered across all samples as shown on left

912 (d) KEGG pathways over-represented in the RBPs with a significant increase in RNA-binding
 913 (e-f) Protein abundance for KEGG pathways over-represented in proteins with a significant
 914 increase in RNA-binding. Individual proteins with a significant increase are highlighted in green.
 915 (e) tRNA aminoacylation pathway. (f) Glycolysis pathway
 916

Figure 6. OOPS unravels the first bacterial RBPome.

(a) Overlap between *E.coli* OOPS RBPs and GO annotated RBPs

(b) Top 10 biological process GO terms over-represented in *E.coli* OOPS RBPs

(c) Protein domains over-represented in *E.coli* OOPS RBPs

(d) GO terms over-represented in OOPS novel *E.coli* RBPs

(e) RNA-binding capacity of glycolysis/gluconeogenesis proteins. Proteins coloured by

pathways; blue text = only glycolysis, orange text = only gluconeogenesis, black text = both,

Asterisks = dynamic RNA-binding after nocodazole arrest. GO:RBP=GO-annotated RBP.

Orange filled circle = protein observed in the dataset indicated. Dark blue fill = protein in human

RBP-Capture experiments but listed as a lower-confidence “candidate” RBP. Empty circle =

928 protein present in species but not observed in dataset. Thick black line indicates an RNA-
929 binding site was identified in the sequential digestion experiment.