nature biotechnology

SUPPLEMENTARY INFORMATION

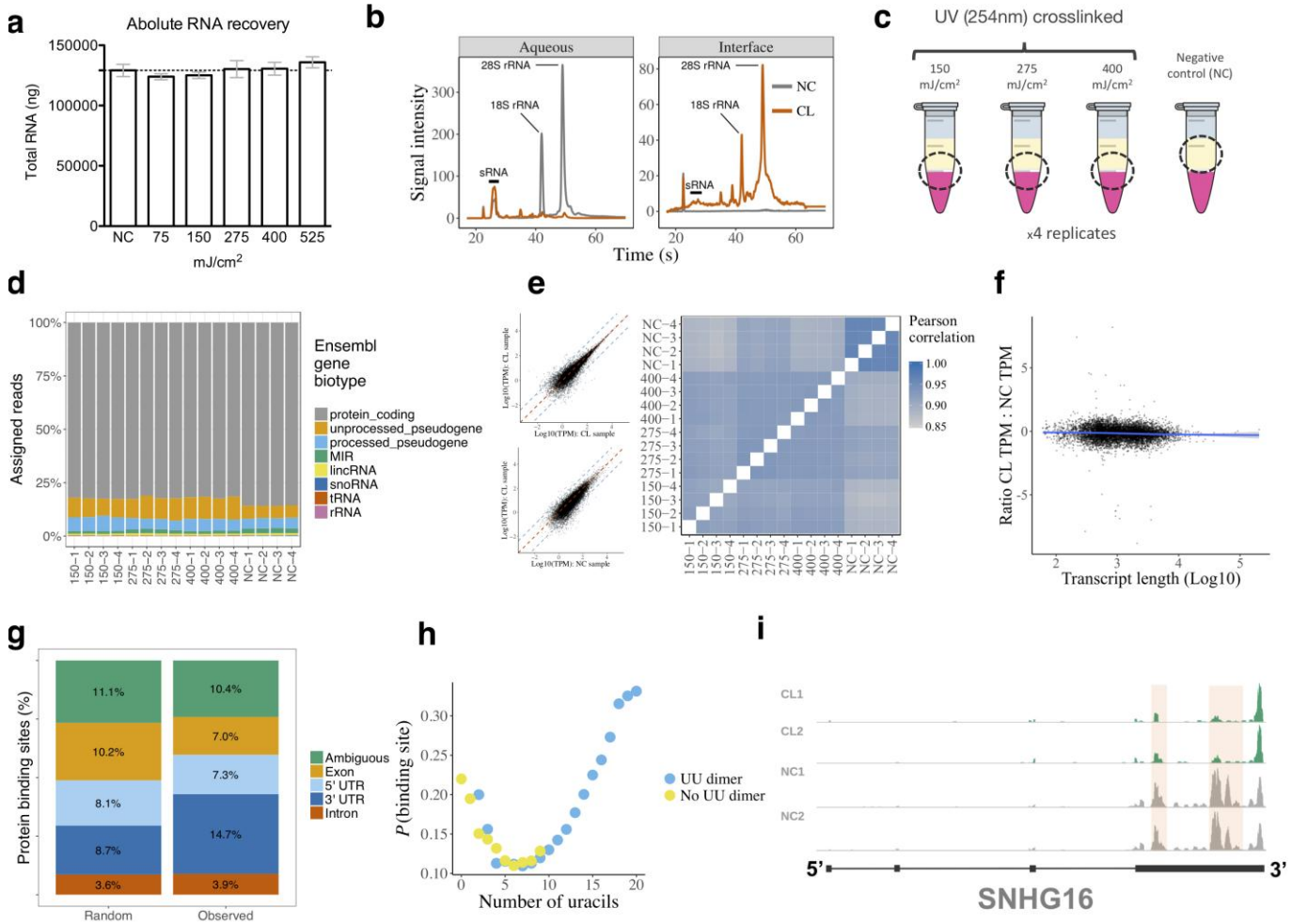
https://doi.org/10.1038/s41587-018-0001-2

In the format provided by the authors and unedited.

Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS)

Rayner M. L. Queiroz[1,5], Tom Smith[1,5], Eneko Villanueva[1,5]*, Maria Marti-Solano[2], Mie Monti[1], Mariavittoria Pizzinga[3], Dan-Mircea Mirea[1], Manasa Ramakrishna[3], Robert F. Harvey[3], Veronica Dezi[3], Gavin H. Thomas[4], Anne E. Willis[3] and Kathryn S. Lilley[1]*

[1]Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Cambridge, UK. [2]MRC Laboratory of Molecular Biology, Cambridge, UK. [3]MRC Toxicology Unit, University of Cambridge, Leicester, UK. [4]Department of Biology, University of York, York, UK. [5]These authors contributed equally: Rayner M. L. Queiroz, Tom Smith, Eneko Villanueva. *e-mail: ev318@cam.ac.uk; k.s.lilley@bioc.cam.ac.uk



Supplementary Figure 1

OOPS unbiasedly recovers all PBR species.

(a) Absolute quantification of the total recovered RNA (PBR + Free RNA). Data shown as mean +/- SD of 3 independent experiments

(b) Representative Bioanalyzer representation of the RNAs size-profile obtained from the aqueous and interfaces. Red: CL (400mJ/cm2). Grey: NC

(c) Graphical representation of the RNA-origin analysed by RNA-seq

(d) Relative proportions of reads assigned to Ensembl gene biotypes for all analysed samples

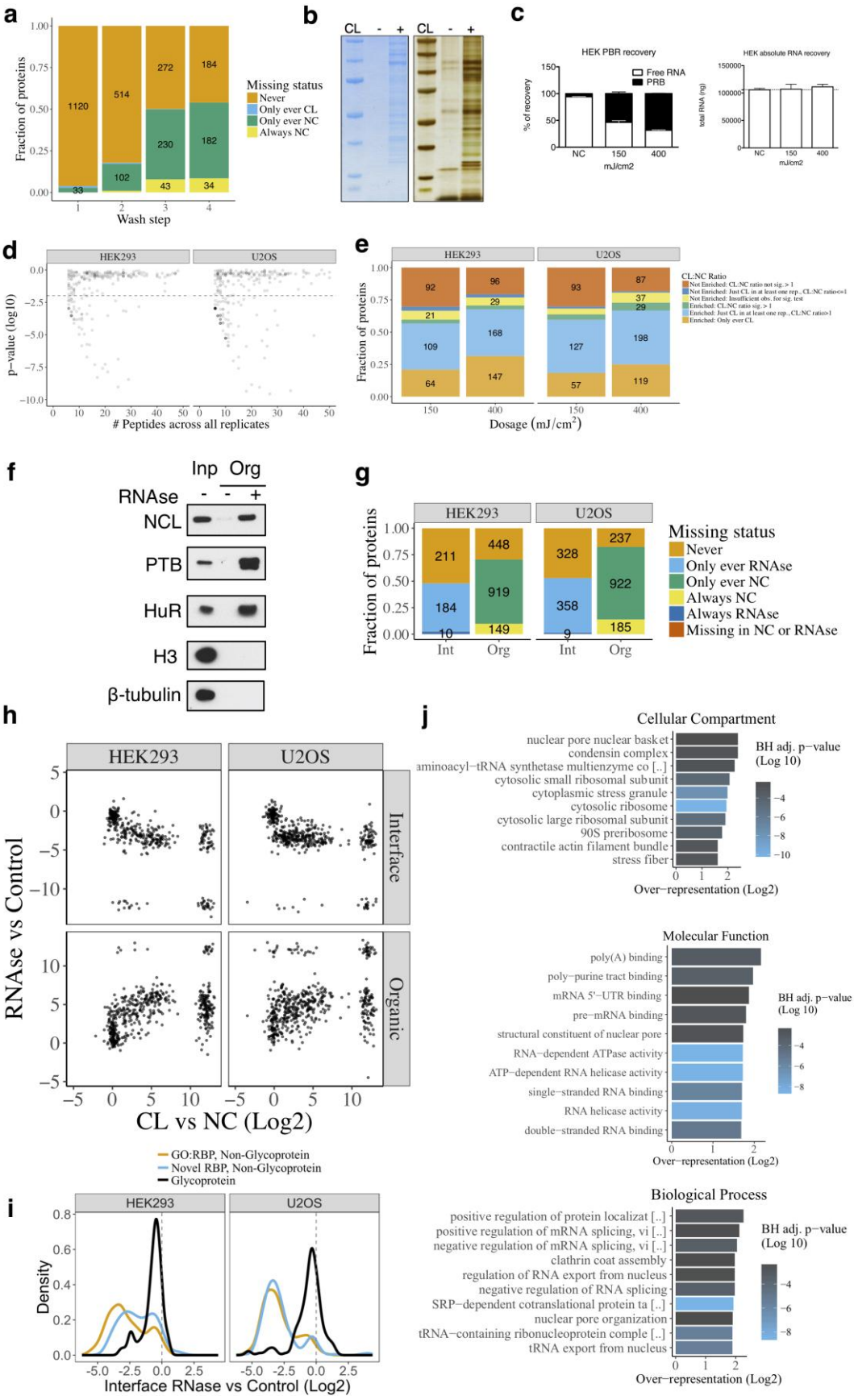
(e) Correlation between two CL replicates (upper left panel) and a NC and 400 mJ/cm2 CL replicate (upper left panel), blue dashed line represents a 10-fold difference. Pearson correlation of all analysed samples with NC (right)

(f) Correlation between the ratio of CL/NC RNA in the interface and the length of the transcript

(g) Percentage of windows deemed to be a site of protein binding. Windows split according to the mRNA feature type they overlap. Windows overlapping two none-intronic features classified as ambiguous. Random: random selection of the same number of windows, with the same distribution of read depths to obtain a null expectation based on read depth alone.

(h) Relationship between uridine content in a window and the probability of the window being deemed to contain a protein binding site

(i) Read coverage across SNHG16 for CL (400 mJ/cm2) and NC replicates. Red boxes denote regions with consistently reduced coverage in CL.


a
Fraction of proteins
Wash step
Missing status
Never
Only ever CL
Only ever NC
Always NC
1120
33
514
102
272
230
43
184
182
34
b
CL
c
HEK PBR recovery
Free RNA
PRB
% of recovery
mJ/cm2
HEK absolute RNA recovery
total RNA (ng)
d
HEK293
U2OS
p−value (log10)
Peptides across all replicates
e
HEK293
U2OS
Fraction of proteins
Dosage (mJ/cm²)
CL:NC Ratio
f
Inp
Org
RNAse
NCL
PTB
HuR
H3
β-tubulin
g
HEK293
U2OS
Fraction of proteins
Int
Org
Missing status
Never
Only ever RNAse
Only ever NC
Always NC
Always RNAse
Missing in NC or RNAse
h
HEK293
U2OS
Interface
Organic
RNAse vs Control
CL vs NC (Log2)
GO:RBP, Non-Glycoprotein
Novel RBP, Non-Glycoprotein
Glycoprotein
i
HEK293
U2OS
Density
Interface RNAse vs Control (Log2)
j
Cellular Compartment
nuclear pore nuclear basket
condensin complex
aminoacyl−tRNA synthetase multienzyme co [..]
cytosolic small ribosomal subunit
cytoplasmic stress granule
cytosolic ribosome
cytosolic large ribosomal subunit
90S preribosome
contractile actin filament bundle
stress fiber
BH adj. p−value (Log 10)
Over−representation (Log2)
Molecular Function
poly(A) binding
poly−purine tract binding
mRNA 5'−UTR binding
pre−mRNA binding
structural constituent of nuclear pore
RNA−dependent ATPase activity
ATP−dependent RNA helicase activity
single−stranded RNA binding
RNA helicase activity
double−stranded RNA binding
BH adj. p−value (Log 10)
Over−representation (Log2)
Biological Process
positive regulation of protein localizat [..]
positive regulation of mRNA splicing, vi [..]
negative regulation of mRNA splicing, vi [..]
clathrin coat assembly
regulation of RNA export from nucleus
negative regulation of RNA splicing
SRP−dependent cotranslational protein ta [..]
nuclear pore organization
tRNA−containing ribonucleoprotein comple [..]
tRNA export from nucleus
BH adj. p−value (Log 10)
Over−representation (Log2)

Supplementary Figure 2

OOPS specifically recovers RBPs.

(a) The proportion of proteins missing in at least one CL replicate, at least one NC replicate, always missing in NC replicates or never missing in any sample

(b) Representative image of an acrylamide gel showing the interface protein content of a NC and CL samples. Left: coomassie-blue staining. Right: silver staining

(c) Crosslinking efficacy in HEK-293. Left panel: Relative proportions of free RNA (aqueous phase) and protein-bound RNA (RBR; interface) with increasing UV dosage. Right panel: Absolute quantification of the total recovered RNA (PBR + Free RNA). Data shown as mean +/- SD of 3 independent experiments

(d) Relationship between the number of peptides observed for a protein across all replicates and the p-value for CL-enrichment. Many proteins have insufficiently small p-value to pass 1% FDR threshold (dashed line) because of a lack of statistical power with few observations

(e) Classification of proteins as enriched or depleted in CL vs NC experiment

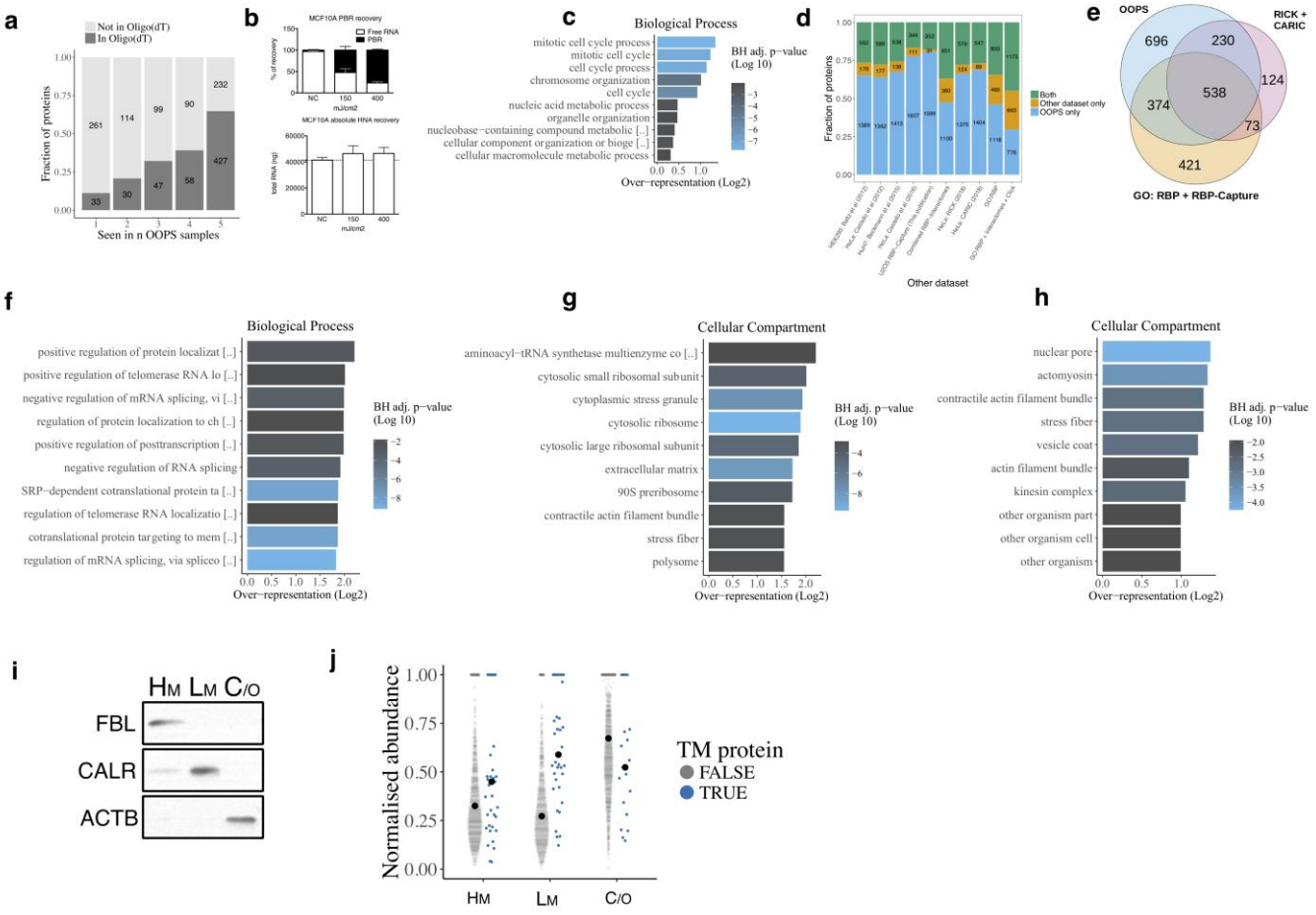
(f) Representative western blots for canonical RBPs (NCL (Nucleolin), PTB and HuR) and negative control proteins (H3 (Histone-H3) and β-tubulin)

(g) The proportion of proteins which were missing in at least one RNAse replicate, at least one control replicate, always missing in the RNAse or control replicates or never missing. Int: Interface. Org: Organic

(h) Protein CL vs NC ratio and RNAse vs control ratio in the 3rd interface and 4th organic phases

(i) RNAse vs control ratio in the 3rd interface for GO annotated RBPs, other OOPS RBPs and glycoproteins

(j) GO terms over-represented in the proteins that migrate to the organic phase upon RNAse treatment



Supplementary Figure 3

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

(a) Agreement between OOPS (n=5) and Oligo(dT) RBP-Capture in HEK 293 cells

(b) Crosslinking efficacy in MCF10A. Upper panel: Relative proportions of free RNA (aqueous phase) and protein-bound RNA (RBR; interface) with increasing UV dosage. Lower panel: Absolute quantification of the total recovered RNA (PBR + Free RNA). Data shown as mean +/- SD of 3 independent experiments

(c) Biological process GO terms over-represented in tumoral-cell specific RBPs

(d) Detailed agreement between OOPS and published human RBPomes

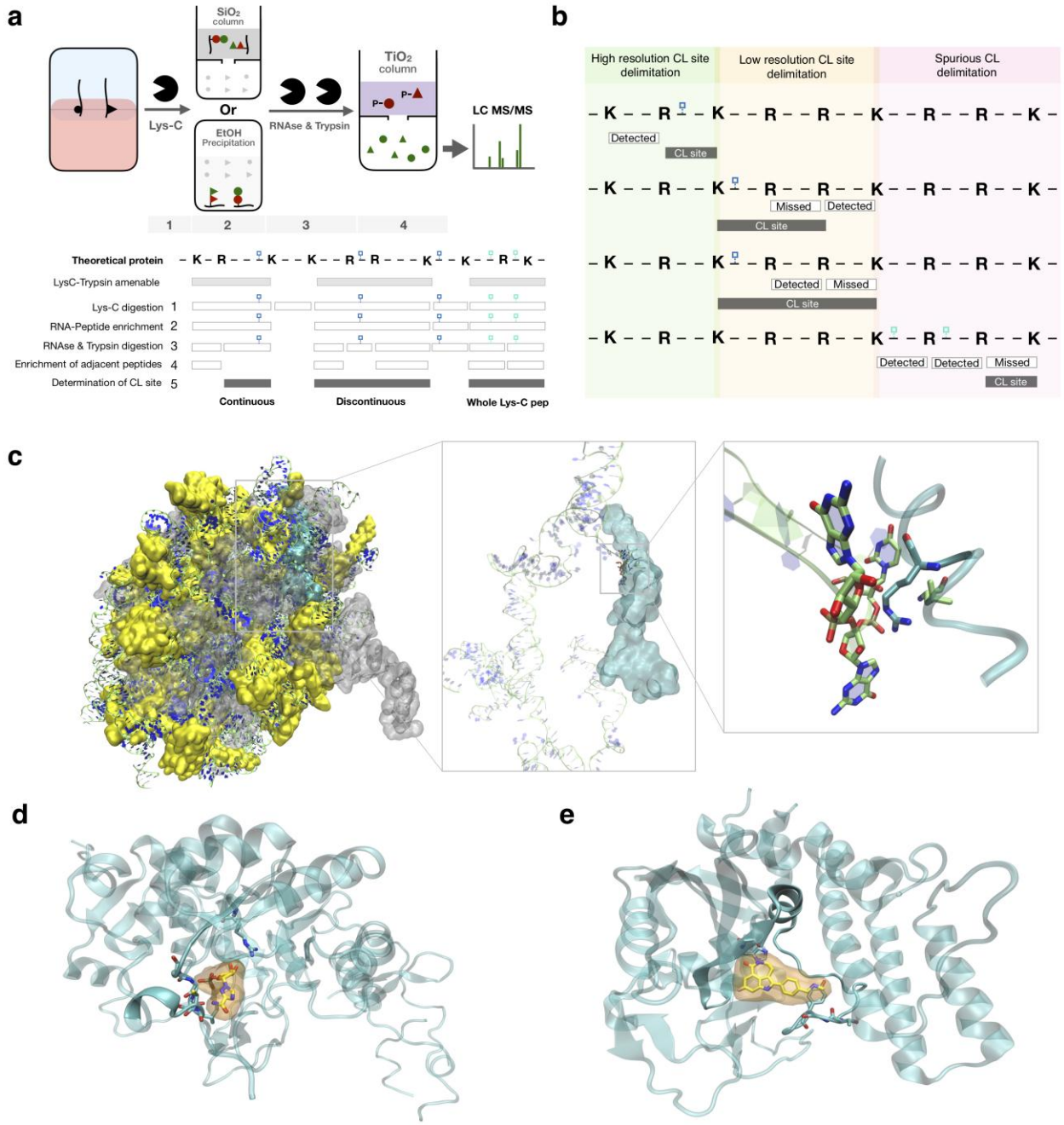
(e) Overlap between the union of OOPS proteins identified in the 3 cell lines, the union of RICK and CARIC studies, and GO annotated RBPs. Proteins were restricted to those expressed in at least one of the three OOPS cell lines

(f-g) Top 10 Biological processes and Cellular compartment GO terms over-represented in the proteins identified in U2OS, HEK293 or MCF10A OOPS

(h) Top 10 Cellular Compartments GO terms over-represented in OOPS-exclusive RBPome.

(i) Representative western blot of the cellular subfractionation. Analysed markers are: nuclear fibrillarin (FBL), endoplasmic reticulum calreticulin (CALR), and cytosolic beta actin (ACTB), in heavy membranes (HM), light membranes (LM) and cytoskeleton and others (C/o) fractions.

(j) Normalised RBP abundance in the indicated fractions. Blue: transmembrane-containing RBPs. Grey: Non transmembrane-containing RBP.



Supplementary Figure 4

Direct assessment of cross-link sites validates most OOPS RBPs.

(a) Top: Schematic representation of the sequential digestion method used to identify the RNA-binding site. RNA-protein adducts are extracted from the interface and digested with Lys-C to yield RNA-peptides which are subsequently enriched by silica affinity column or ethanol precipitation. Enriched RNA-peptides are treated with RNAses followed by trypsin digestion. Peptides containing the UV-crosslinked nucleotide/RNA are retained by a TiO2 affinity column and the unbound fraction containing the peptide sequences adjacent to RNA crosslinking site is analysed by LC-MS/MS. Red=peptides containing site of crosslinking. Green=peptides adjacent to the RNA-binding site peptide.

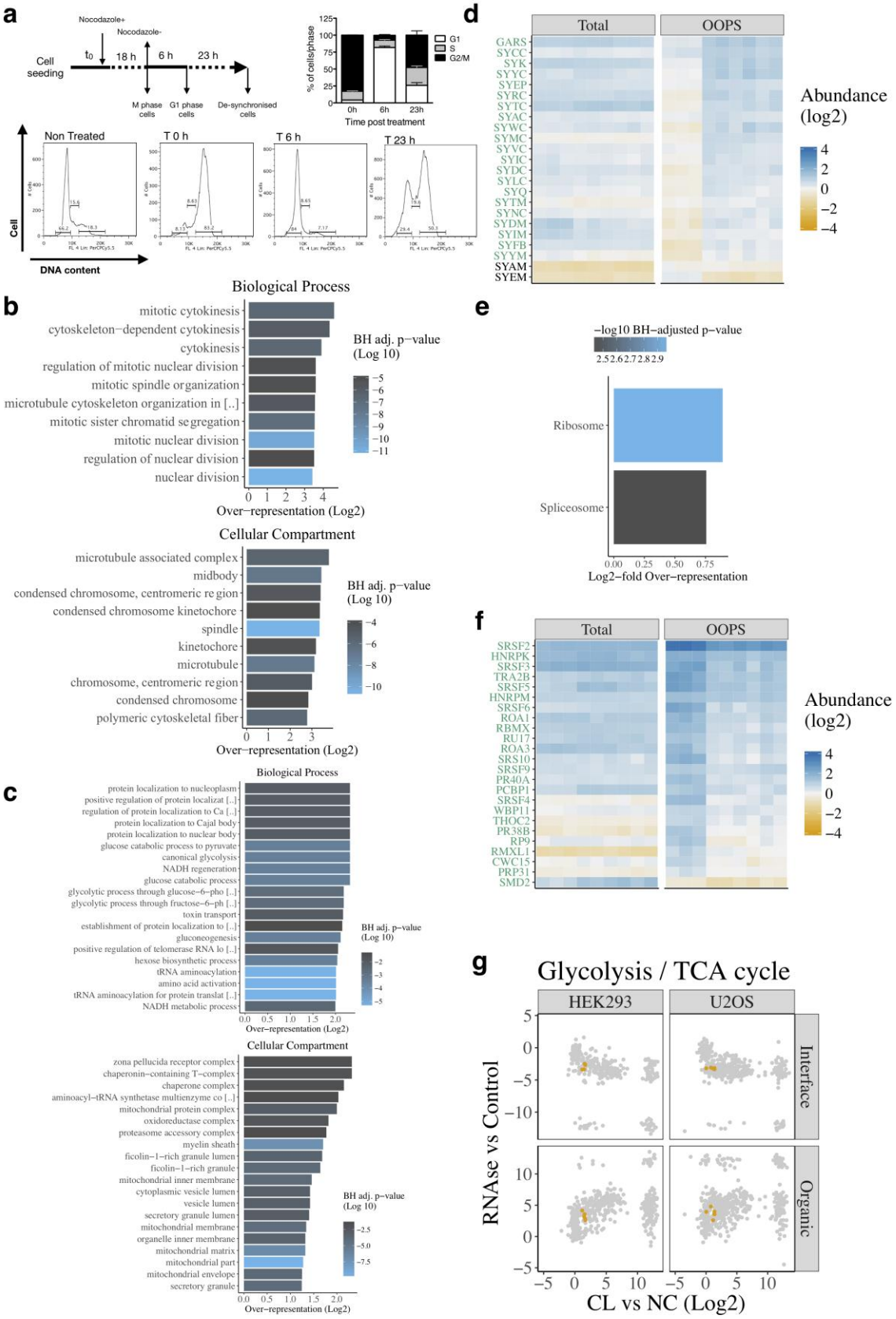
Bottom: Schematic representation of the process followed to determine crosslinking peptide. Blue squares represent a consistent

binding site. Turquoise squares represent a promiscuous binding site.

(b) Resolution and accuracy of RNA binding site detected is dependent on peptides detected.

(c) Cryo-EM structure of the ribosome quality control complex (PDB ID 3J92). Proteins are shown as grey transparent surface, while RNA is depicted as transparent lime ribbon. Proteins previously detected as RBPs are highlighted in yellow, while the newly detected protein is shown in cyan. Zoomed in structures show this new protein as transparent cyan cartoon with the interacting residue detected by direct evidence shown as cyan sticks and RNA residues at 4 Å or less from it shown as lime sticks.

(d-e) Crystal structures of (d) IMPDH2 in complex with ribavirin (PDB ID 1NF7) and (e) PARP1 in complex with rucaparib (PDB ID 1NF7). Proteins are shown as a cyan transparent cartoon and the protein region detected as RNA binding is shown in an opaque representation. Residues at 4 Å or less from inhibitor compounds (yellow sticks and surface representation) are shown as cyan sticks


a
Nocodazole+
Nocodazole-
Cell seeding
t0
18 h
6 h
23 h
M phase cells
G1 phase cells
De-synchronised cells
% of cells/phase
G1
S
G2/M
0h
6h
23h
Time post treatment
Non Treated
T 0 h
T 6 h
T 23 h
Cell
DNA content
b
Biological Process
mitotic cytokinesis
cytoskeleton−dependent cytokinesis
cytokinesis
regulation of mitotic nuclear division
mitotic spindle organization
microtubule cytoskeleton organization in [..]
mitotic sister chromatid segregation
mitotic nuclear division
regulation of nuclear division
nuclear division
BH adj. p−value (Log 10)
Over−representation (Log2)
Cellular Compartment
microtubule associated complex
midbody
condensed chromosome, centromeric region
condensed chromosome kinetochore
spindle
kinetochore
microtubule
chromosome, centromeric region
condensed chromosome
polymeric cytoskeletal fiber
BH adj. p−value (Log 10)
Over−representation (Log2)
c
Biological Process
protein localization to nucleoplasm
positive regulation of protein localizat [..]
regulation of protein localization to Ca [..]
protein localization to Cajal body
protein localization to nuclear body
glucose catabolic process to pyruvate
canonical glycolysis
NADH regeneration
glucose catabolic process
glycolytic process through glucose−6−pho [..]
glycolytic process through fructose−6−ph [..]
toxin transport
establishment of protein localization to [..]
gluconeogenesis
positive regulation of telomerase RNA lo [..]
hexose biosynthetic process
tRNA aminoacylation
amino acid activation
tRNA aminoacylation for protein translat [..]
NADH metabolic process
BH adj. p−value (Log 10)
Over−representation (Log2)
Cellular Compartment
zona pellucida receptor complex
chaperonin−containing T−complex
chaperone complex
aminoacyl−tRNA synthetase multienzyme co [..]
mitochondrial protein complex
oxidoreductase complex
proteasome accessory complex
myelin sheath
ficolin−1−rich granule lumen
ficolin−1−rich granule
mitochondrial inner membrane
cytoplasmic vesicle lumen
vesicle lumen
secretory granule lumen
mitochondrial membrane
organelle inner membrane
mitochondrial matrix
mitochondrial part
mitochondrial envelope
secretory granule
BH adj. p−value (Log 10)
Over−representation (Log2)
d
Total
OOPS
GARS
SYCC
SYK
SYYC
SYEP
SYRC
SYTC
SYAC
SYWC
SYMC
SYVC
SYIC
SYDC
SYLC
SYQ
SYTM
SYNC
SYDM
SYIM
SYFB
SYYM
SYAM
SYEM
Abundance (log2)
e
−log10 BH−adjusted p−value
Ribosome
Spliceosome
Log2−fold Over−representation
f
Total
OOPS
SRSF2
HNRPK
SRSF3
TRA2B
SRSF5
HNRPM
SRSF6
ROA1
RBMX
RU17
ROA3
SRS10
SRSF9
PR40A
PCBP1
SRSF4
WBP11
THOC2
PR38B
RP9
RMXL1
CWC15
PRP31
SMD2
Abundance (log2)
g
Glycolysis / TCA cycle
HEK293
U2OS
Interface
Organic
RNAse vs Control
CL vs NC (Log2)

Supplementary Figure 5

Characterization of the nocodazole arrest experiments.

a) Top: Schematic representation of the nocodazole arrest/release experiment. Bottom-left: representative image of flow cytometry analysis of the cell-cycle stage by DNA content assessment. Bottom-right: Relative proportions of cells in G1, S and M phase for cells synchronised at each time-point is shown as the mean +/- SD of 3 independent experiments

(b) Top 10 Biological Processes and Cellular Compartment GO terms over-represented in the total proteome of synchronised cells versus 6h post released cells

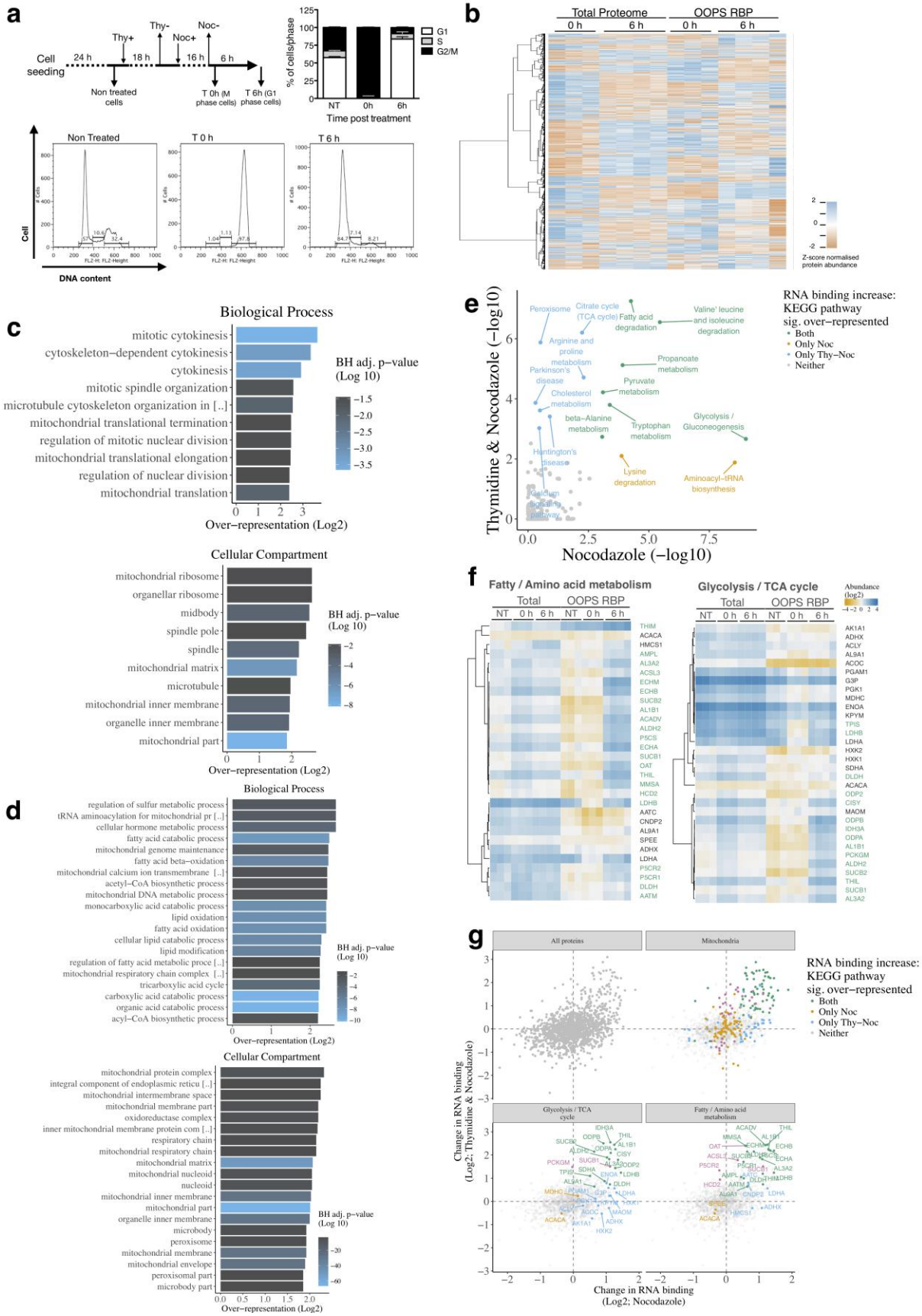
(c) Top 20 Biological Processes and Cellular Compartment GO terms over-represented in the RBPs with a significant increase in RNA-binding post-release in the nocodazole experiment

(d) Protein abundance for tRNA aminoacylation proteins with a significant increase in RNA-binding. Individual proteins with a significant increase are highlighted in green.

(e) KEGG pathways over-represented in the RBPs with a significant decrease in RNA-binding post-release

(f) Protein abundance for splicing proteins with a significant increase in RNA-binding at 0h

(g) Crosslink vs control and RNAse vs control protein abundance ratio in interface and organic phase for glycolytic and TCA cycle proteins (yellow)


a
Cell seeding
24 h
Thy+
18 h
Thy-
Noc+
16 h
Noc-
6 h
Non treated cells
T 0h (M phase cells)
T 6h (G1 phase cells)
% of cells/phase
G1
S
G2/M
NT
0h
6h
Time post treatment
Non Treated
T 0 h
T 6 h
Cell
DNA content
b
Total Proteome
OOPS RBP
0 h
6 h
Z-score normalised protein abundance
c
Biological Process
mitotic cytokinesis
cytoskeleton−dependent cytokinesis
cytokinesis
mitotic spindle organization
microtubule cytoskeleton organization in [..]
mitochondrial translational termination
regulation of mitotic nuclear division
mitochondrial translational elongation
regulation of nuclear division
mitochondrial translation
BH adj. p−value (Log 10)
Over−representation (Log2)
Cellular Compartment
mitochondrial ribosome
organellar ribosome
midbody
spindle pole
spindle
mitochondrial matrix
microtubule
mitochondrial inner membrane
organelle inner membrane
mitochondrial part
d
Biological Process
regulation of sulfur metabolic process
tRNA aminoacylation for mitochondrial pr [..]
cellular hormone metabolic process
fatty acid catabolic process
mitochondrial genome maintenance
fatty acid beta−oxidation
mitochondrial calcium ion transmembrane [..]
acetyl−CoA biosynthetic process
mitochondrial DNA metabolic process
monocarboxylic acid catabolic process
lipid oxidation
fatty acid oxidation
cellular lipid catabolic process
lipid modification
regulation of fatty acid metabolic proce [..]
mitochondrial respiratory chain complex [..]
tricarboxylic acid cycle
carboxylic acid catabolic process
organic acid catabolic process
acyl−CoA biosynthetic process
BH adj. p−value (Log 10)
Over−representation (Log2)
Cellular Compartment
mitochondrial protein complex
integral component of endoplasmic reticu [..]
mitochondrial intermembrane space
mitochondrial membrane part
oxidoreductase complex
inner mitochondrial membrane protein com [..]
respiratory chain
mitochondrial respiratory chain
mitochondrial matrix
mitochondrial nucleoid
nucleoid
mitochondrial inner membrane
mitochondrial part
organelle inner membrane
microbody
peroxisome
mitochondrial membrane
mitochondrial envelope
peroxisomal part
microbody part
BH adj. p−value (Log 10)
Over−representation (Log2)
e
Thymidine & Nocodazole (−log10)
Nocodazole (−log10)
Peroxisome
Citrate cycle (TCA cycle)
Fatty acid degradation
Valine' leucine and isoleucine degradation
Arginine and proline metabolism
Propanoate metabolism
Parkinson's disease
Pyruvate metabolism
Cholesterol metabolism
beta−Alanine metabolism
Tryptophan metabolism
Glycolysis / Gluconeogenesis
Huntington's disease
Lysine degradation
Aminoacyl−tRNA biosynthesis
Calcium signaling pathway
RNA binding increase: KEGG pathway sig. over−represented
Both
Only Noc
Only Thy−Noc
Neither
f
Fatty / Amino acid metabolism
Glycolysis / TCA cycle
Total
OOPS RBP
NT
0 h
6 h
Abundance (log2)
g
All proteins
Mitochondria
Glycolysis / TCA cycle
Fatty / Amino acid metabolism
Change in RNA binding (Log2; Thymidine & Nocodazole)
Change in RNA binding (Log2; Nocodazole)
RNA binding increase: KEGG pathway sig. over−represented
Both
Only Noc
Only Thy−Noc
Neither

Supplementary Figure 6

Characterization of the RBPome dynamics in the thymidine-nocodazole arrest experiments.

(a) Top-left: Schematic representation of the thymidine-nocodazole arrest/release experiment. Top-right: Relative proportions of cells in G1, S and G2/M phase for each time-point is shown as the mean +/- SD of 4 independent experiments Bottom: representative image of flow cytometry analysis of the cell-cycle stage by DNA content assessment.

(b) Protein abundance from total proteome and OOPS extractions in thymidine-nocodazole arrested and released cells. Abundance z score normalised within each extraction type. Proteins hierarchically clustered across all samples as shown on left

(c) Top 10 Biological process and Cellular Compartment GO terms over-represented in the total proteome of thymidine-nocodazole arrested cells versus non treated cells.

(d) Top 20 Biological process and Cellular Compartment GO terms over-represented in the RBPs with a significant increase in RNA-binding post-release in the thymidine-nocodazole experiment

(e) Correlation between p-values for KEGG pathway over-representation in the two nocodazole experiments.

(f) Protein abundance for groups of overlapping KEGG pathways over-represented in proteins with a significant increase in RNA-binding. Individual proteins with a significant increase are highlighted in green.

(g) Protein abundance for groups of overlapping KEGG pathways over-represented in proteins with a significant increase in RNA-binding. Individual proteins with a significant increase in RNA binding in 6 h vs 0 h are highlighted in green.

Supplementary Note

Deduplication of SENSE total RNA-Seq

The SENSE total RNA-Seq Library Prep kit (Lexogen) adds a non-template 2-9bp to the 5' end of the read sequence. The first 9bp of the read was therefore treated as a pseudo-unique molecular identifier (pUMI) for the purposes of de-duplication to help avoid over-deduplication. Error-aware read deduplication using pUMIs was performed with UMI-tools[1].

Identification of protein binding sites

In order to identify regions of the transcriptome with decreased read coverage, a two-step process was performed. Firstly, for each gene in each pair of samples, the mean absolute difference in read coverage was calculated across all nucleotides. A Wilcoxon rank-sum test was applied to identify genes with a significant difference between intra-condition (NC and CL) and inter-condition comparisons (Benjamini-Hochberg[2] adjust p-value < 0.01). For these genes, 40 bp sliding windows (20 bp overlap) were tiled across the complete gene model (including introns) and the number of reads overlapping each window computed with *bedtools coverage*[3]*.* Windows with a significantly decreased coverage (putative protein binding site) in the four 400 mJ/cm^2 replicates vs four NC samples were identified using DESeq2[4] (Benjamini-Hochberg[2] adjusted p-value < 0.01). DESeq2 performs an initial filtering step to remove all input features (in this case windows) with insufficient read counts.

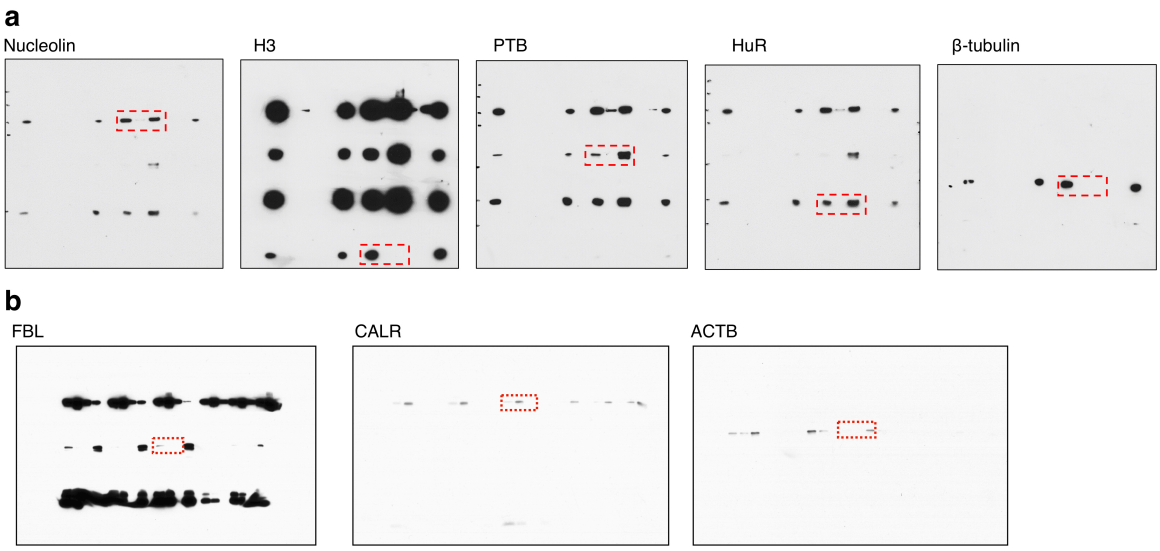
Overlap with eCLIP data

ENCODE eCLIP bed files[5] were downloaded from https://www.encodeproject.org on 02/01/2018 and filtered according to the ENCODE stringent filter: -log10(p-value) ≥ 5 and log2(fold-enrichment) ≥ 3. For all windows retained for statistical testing by DESeq2, the number of U2OS OOPS RBPs with an eCLIP peak within the window was computed. A relationship was observed between the number of eCLIP peaks in a window and *P*(putative binding site). However, since read coverage was a potential confounding factor, we grouped windows based on the number of eCLIP peaks and for each group computed the null expectation for *P*(putative binding site) by randomly selecting a set of windows with the same distribution of read coverage, with 100 iterations (denoted as random shuffle in Figure 1g). To identify the relationship between uracil content and (putative binding site), we computed the number of uracils per window and separated windows based on whether they contained a UU dimer.

Antibodies:

Anti-PTB (1:1000, homemade); anti-nucleolin (1:1000, Clone 4E2, GeneTex); anti-HuR (1:1000, Clone 3A2, Santa Cruz Biotechnology); anti-histone H3 (1:1000, Bethyl Laboratories,); anti-β-tubulin (1:1000, Cell Signaling Technology). Anti-fibrillarin as nuclear marker (1:1000, clone C13C3, Cell Signaling, Leiden, Netherlands), anti-calreticulin as endoplasmic reticulum marker (1:1000, clone D3E6, Cell Signaling), and anti-beta actin as cytosolic marker (1:5000, ab8227, Abcam, Cambridge, UK).

Uncropped western blot scans:



(a) Uncropped scans of western blots presented in the Supplementary Figure S2f.
(b) Uncropped scans of western blots presented in the Supplementary Figure S3i.

SILAC labelling rationale

U-2 OS osteosarcoma cells were cultured in commercially available DMEM media for SILAC labeling (Thermo Fisher Scientific). We cultured the cells for two weeks (four passages) and optimal labelling confirmed from test samples obtained in parallel to the samples used in this study. Four biological replicates were performed in all our SILAC experiments, with the labels switched such that the treatment condition was the heavy isotope in replicates 1 & 2 and the light isotopes in replicates 3 & 4. For MS analysis of SILAC labelled samples, the CHOPIN acquisition method was used in order to allow the use of both detectors in the Orbitrap Lumos on predefined precursor types that optimizes parallel ion processing, thus ensuring higher yield. Despite having the possibility to trigger the MS^2 fragmentation in only peptides present in SILAC pairs, we reasoned that many peptides are likely to be observed with no signal in the control condition, which was indeed confirmed.

67

68 **LC-MS/MS technical details**

69 Supplementary table 6 summarises the main MS parameters in each experiment set
70 and the TMT tags for each sample in the nocodazole experiments.

71

72 Samples from RBP-capture and subcellular fractionation were analysed in an
73 Orbitrap nano-ESI Q-Exactive mass spectrometer (Thermo Fisher Scientific),
74 coupled to a nanoLC (Dionex Ultimate 3000 UHPLC). Samples were trapped on a
75 100 µm × 2 cm, C18, 5 µm, 100 trapping column (Acclaim PepMap 100) in µL-pickup
76 injection mode at 15 µL/min flow rate for 10 minutes. Samples were then loaded on a
77 Rapid Separation Liquid Chromatography, 75 µm × 50 cm nanoViper C18 3 µm 100
78 column (Acclaim, PepMap) at 50 ºC retrofitted to an EASY-Spray source with a flow
79 rate of 300 nL/min (buffer A, HPLC H2O, 0.1% formic acid; buffer B, 100% ACN,
80 0.1% formic acid; 0–15 min: at 4% buffer B, 15–90 min: linear gradient 4% to 40%
81 buffer B, 90-90.3 min: 40% to 90% buffer B, 90.3-95 min: at 90% buffer B, 95-95.3
82 min: 90% to 4% buffer B, 95.3-120 min: at 4% buffer B). Mass spectra were acquired
83 in positive ion mode applying data-dependent automatic survey MS scan and
84 tandem mass spectra (MS/MS) acquisition modes. Each MS scan in the Orbitrap
85 analyser (mass range = m/z 380 –1500, resolution = 35,000) was followed by
86 MS/MS of the 20 most intense peptides. Fragmentation was performed by high-
87 energy collision-activated dissociation (NCE = 25; resolution = 17,500), and selected
88 fragmented ions were dynamically excluded for 20 s.

89

90 SILAC and unlabeled samples generated from OOPS experiments in *E. coli* and
91 MCF10A were analysed in the Orbitrap Fusion Lumos (Thermo Fisher Scientific)

coupled to a nanoLC Dionex Ultimate 3000 UHPLC (Thermo Fisher Scientific) with the same columns, temperature and flow rates as above. Analytical chromatography was performed over 120 min (buffer A, HPLC H2O, 0.1% formic acid; buffer B, 100% ACN, 0.1% formic acid; 0–15 min: at 2% buffer B, 15–100 min: linear gradient 2% to 40% buffer B, 100-100.3 min: 40% to 90% buffer B, 100.3-110 min: at 90% buffer B, 110-110.3 min: 90% to 2% buffer B, 100.3-120 min: at 2% buffer B). Mass spectra were acquired using CHarge Ordered Parallel Ion aNalysis (CHOPIN) acquisition in positive ion mode as previously reported[6]. MS scans were acquired at a resolution of 120,000 between 400 and 1500 m/z and an AGC target of 4×10^5. MS/MS spectra were acquired in the linear ion trap (rapid scan mode) after collision-induced dissociation (CID) fragmentation at a collision energy of 35% and an AGC target of 4×10^3 for up to 250 ms, employing a maximal duty cycle of 3 s, prioritizing the most intense ions and injecting ions for all available parallelizable time. Selected precursor masses were excluded for 60 s. For precursor selection, we prioritized the least abundant signals. Doubly charged ions were scheduled for CID/ion trap analysis with the same parameters applied as above. Charge states 3–7 with precursor intensity >500 000, however, were scheduled for analysis by a fast HCD/Orbitrap scan of maximal 40 ms (17,500 resolution). The remaining charge-state 3–7 ions with intensity <500 000 were scheduled for analysis by CID/ion trap, as described above.

Samples for assessment of RNA crosslinking site were acquired in the Orbitrap Fusion Lumos with the same configurations and chromatography setup as above, but with 60 min of analytical chromatography (buffer A, HPLC H2O, 0.1% formic acid; buffer B, 100% ACN, 0.1% formic acid; 0–3 min: at 2% buffer B, 3–30 min: linear gradient 2% to 40% buffer B, 30-30.3 min: 40% to 90% buffer B, 30.3-35 min:

at 90% buffer B, 35-35.3 min: 90% to 2% buffer B, 35.3-60 min: at 2% buffer B). MS scans were acquired at a resolution of 120,000 between 380 and 1500 m/z and an AGC target of 4×10^5. MS/MS spectra were acquired using HCD and Orbitrap analyser with collision energy of 38%, AGC target of 5×10^4 for up to 86 ms (50,000 resolution), dynamic exclusion of 70 s and employing a maximal duty cycle of 3 s.

TMT-labelled fractions were analysed in an Orbitrap Fusion Lumos with the same configurations and chromatography setup as above, but with 240 min of analytical chromatography (buffer A, HPLC H2O, 0.1% formic acid; buffer B, 100% ACN, 0.1% formic acid; 0–15 min: at 2% buffer B, 15–215 min: linear gradient 2% to 40% buffer B, 215-215.3 min: 40% to 90% buffer B, 215.3-225 min: at 90% buffer B, 225-225.3 min: 90% to 2% buffer B, 225.3-240 min: at 2% buffer B). Mass spectra were acquired in positive ion mode applying data acquisition using synchronous precursor selection MS3 (SPS-MS3) acquisition mode[7]. Each MS scan in the Orbitrap analyzer (mass range = m/z 400 –1500, resolution = 120,000). The most intense ions over a threshold of 5×10^4 were selected for collision induced dissociation (CID)-MS2 fragmentation, with an AGC target and maximum accumulation time of 1×10^4 and 60 ms. Mass filtering was performed by the quadrupole with 0.7 m/z transmission window, followed by CID fragmentation in the linear ion trap with 35% normalized collision energy. SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–1,200 m/z range and were set to preclude selection of the precursor ion and TMT ion series. AGC targets and maximum accumulation times were set to 5×10^4 and 86 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 65% normalized collision energy and were analysed in the Orbitrap with nominal resolution of 50,000. The number of

SPS-MS3 spectra acquired between full scans was restricted to a duty cycle of 3 s. Selected fragmented ions were dynamically excluded for 60 s.

TMT reporter values were assessed through Proteome Discoverer using *Most Confident Centroid* method for peak integration and integration tolerance of 20 ppm (0.0024 Da). Reporter ion intensities were adjusted to correct for the isotopic impurities of the different TMT reagents (following manufacturer specifications). Sample labels for each TMT tag are presented in supplementary table 6.

Peptide to protein assignment

Peptide to protein assignment was performed using an approximate minimum parsimonious approach. Swiss-Prot proteins identified (including non Master-proteins) were first ranked by their number of peptides identified. The peptides from the top ranked proteins deemed to be assigned to this protein and removed from all remaining proteins. The remaining proteins were then re-ranked based on the updated peptide assignments. This was repeated until all peptides were accounted, with ties settled by assigning the peptide to multiple proteins. Peptides which did not receive a unique master protein identification were discarded.

Identification of RNA binding sites

Putative RNA binding sites were identified using the trypsin peptides detected (Figure S4a). To work back from the detected peptides to the sites of RNA crosslinking, a catalog of theoretical peptides was generated by *in-silico* protein digestion using Lys-C and trypsin. Peptides less than 7 or greater than 36 amino acids were discarded (these thresholds were derived from the 1^{st} and 99^{th} percentile

for the size of trypsin peptides detected in the other MS experiments). Lys-C peptides which did not yield at least two trypsin peptides were discarded since identification of RNA binding sites is dependent on detection of the adjacent peptides without RNA crosslinking but from the same Lys-C peptide. Detected peptides were then matched to this *in-silico* generated catalogue. Peptides observed in all 3 ethanol replicates, or at least 1 silica replicate were retained. The putative site of RNA crosslinking was then identified as the portion of the Lys-C peptide which was not covered by an identified trypsin peptide. Where this region was discontinuous, e.g only the middle portion of a Lys-C peptides was covered by a detected trypsin peptides, the binding site was identified as the region from the first amino acid which was not covered by a trypsin peptide to the last non-covered amino acid, including all amino acids in between (Figure S4a). Where all Lys-C peptides amino acids were covered, the entire peptide was considered a putative RNA binding site. Putative binding sites > 30 amino acids were discarded from downstream analyses.

Supplementary references

1. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27,** 491–499 (2017).
2. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 Published by : *J. R. Stat. Soc. Ser. B* **57,** 289–300 (1995).
3. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).
4. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
5. Van Nostrand, E. L. *et al.* A Large-Scale Binding and Functional Map of Human RNA Binding Proteins. *bioRxiv* (2017).
6. 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).
7. 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).