System-wide identification of RNA-binding proteins by interactome capture

Alfredo Castello, Rastislav Horos, Claudia Strein, Bernd Fischer, Katrin Eichelbaum, Lars M Steinmetz, Jeroen Krijgsveld & Matthias W Hentze

European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. Correspondence should be addressed to M.W.H. (hentze@embl.de).

Published online 14 February 2013; doi:10.1038/nprot.2013.020

Owing to their preeminent biological functions, the repertoire of expressed RNA-binding proteins (RBPs) and their activity states are highly informative about cellular systems. We have developed a novel and unbiased technique, called interactome capture, for identifying the active RBPs of cultured cells. By making use of *in vivo* UV cross-linking of RBPs to polyadenylated RNAs, covalently bound proteins are captured with oligo(dT) magnetic beads. After stringent washes, the mRNA interactome is determined by quantitative mass spectrometry (MS). The protocol takes 3 working days for analysis of single proteins by western blotting, and about 2 weeks for the determination of complete cellular mRNA interactomes by MS. The most important advantage of interactome capture over other *in vitro* and *in silico* approaches is that only RBPs bound to RNA in a physiological environment are identified. When applied to HeLa cells, interactome capture revealed hundreds of novel RBPs. Interactome capture can also be broadly used to compare different biological states, including metabolic stress, cell cycle, differentiation, development or the response to drugs.

INTRODUCTION

RBPs associate with RNAs from synthesis to decay, forming dynamic complexes called ribonucleoproteins. RBPs control RNA fate and thus have central roles in gene expression. The implementation of technologies such as immunoprecipitation combined with microarrays or next-generation sequencing have allowed the deeper study of RNA networks controlled by individual RBPs[1,2]. However, the scope of proteins involved in RNA biology is still unclear and seems to have been previously underestimated, as judged by the mRNA interactome of HeLa and HEK293 cells[3,4].

A number of approaches have been previously used to study the mRNA interactome. Genome-wide protoarrays and fluorescent RNA probes were used in two different studies to systematically identify yeast RBPs[5,6]. Immobilized RNA probes were also used as bait to capture specific RBPs *in vitro*, followed by quantitative MS[7]. However, these approaches do not discriminate nonphysiological RNA-protein interactions, facilitated by the biochemical properties of the respective polypeptides, from those taking place in living cells. *In silico* algorithms have been used to identify candidate RBPs, searching for RNA-binding domains or RNA-related enzymatic activities[8] in the cellular proteome. These approaches identified additional proteins as potential RBPs that share similar structural and functional features with previously known RNA binders. However, unconventional RBPs cannot be identified by these analyses, as illustrated by recently published mRNA interactome data sets[3,4].

We developed interactome capture to circumvent these technical limitations, combining UV cross-linking and oligo(dT) capture to pull down proteins bound to RNA in living cells. We applied interactome capture in order to successfully determine the first mRNA interactome of a human cell line[4].

Overview of the protocol

To covalently couple RBPs to RNAs *in vivo*, cell monolayers are irradiated with UV light at 254 nm to cross-link the photoreactive nucleotide bases with amino acids such as Phe, Trp, Tyr, Cys and Lys (conventional UV cross-linking, cCL)[9,10]. In parallel, we also applied the photoactivatable-ribonucleoside-enhanced cross-linking (PAR-CL) protocol, which uses the photoactivatable nucleotide 4-thiouridine (4SU)[11]. This nucleotide analog is taken up by cells and incorporated into nascent RNAs. UV light irradiation at 365 nm of cell monolayers induces efficient protein-RNA cross-linking. After cell lysis, polyadenylated RNAs are captured using oligo(dT) magnetic beads. After stringent washes, RBPs bound to poly(A)$^+$ RNA are released by RNase treatment and identified by MS (**Fig. 1**).

Advantages of the method

Interactome capture has notable advantages over previous RBP identification methods:

(1) UV irradiation generates highly reactive, short-lived states of the nucleotide bases within the RNA, inducing covalent bond formation only with amino acids in direct contact (zero distance)[12]. Neither cCL (254 nm) nor PAR-CL (365 nm) promote protein-protein cross-linking, because the peptide bond absorbs at 230 nm and tryptophan absorbs at 280 nm[12,13]. Thus, only proteins and RNAs in close contact cross-link efficiently to each other upon UV irradiation.

(2) UV light is directly applied to monolayers of living cells, thus 'freezing' physiological *in vivo* protein-RNA interactions.

(3) Nucleic acid hybridization (poly(A) tail—oligo(dT)) is stable in the presence of ionic detergents (0.5% (wt/vol) lithium dodecyl sulfate, LiDS) and in high-salt buffers (500 mM lithium chloride (LiCl)), thus allowing efficient removal of polypeptides associated with the RNA template noncovalently or via protein-protein interactions.

(4) Interactome capture is compatible with quantitative proteomics for the determination of high-confidence mRNA interactomes.

(5) Interactome capture can be applied to study RBP composition and dynamics *in vivo* in distinct biological systems and under different experimental conditions.

Limitations of the approach

Interactome capture will fail to detect RBPs that are (i) not bound to poly(A)$^+$ RNAs; (ii) not expressed in the (cellular) model under

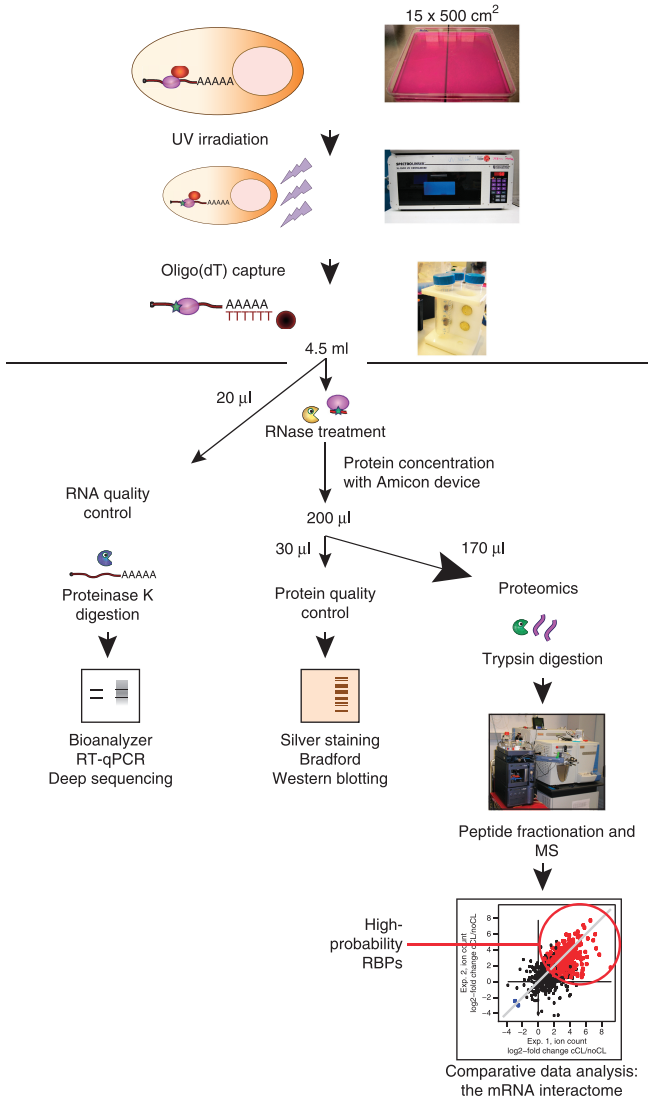

Figure 1 | Schematic representation of the interactome capture pipeline. mRNA-protein interactions are preserved by performing UV cross-linking (cCL and PAR-CL) with living cells. Poly(A) RNA-protein complexes are captured by pull-down with oligo(dT) magnetic beads and stringently washed. Eluates are processed with proteinase K for RNA quality control, RNases for protein quality control, and RNases and trypsin for MS. Comparative proteomic data analysis defines 'high-confidence' mRNA interactomes.

study; (iii) not active under the conditions of the experiment; or (iv) not cross-linked efficiently upon UV light irradiation. UV cross-linking has limited efficiency, and it therefore requires sufficient quantities of starting material for analysis (for HeLa cells we used 2.8×10^8 cells).

Applications and future uses of the method

Interactome capture can be applied to test the *in vivo* RNA-binding activity of a protein of interest using small-scale experiments and western blotting and to determine comprehensive cellular mRNA interactomes by MS.

Interactome capture was initially applied to proliferating HeLa cells[4]. It can be now applied to other cell types (e.g., stem cells) and adapted to tissues and unicellular (e.g., *Saccharomyces cerevisiae*) and multicellular (e.g., *Caenorhabditis elegans*) organisms[14–16]. Furthermore, we envisage a major future use of interactome capture as a means of monitoring the dynamic responses of mRNA interactomes to different environmental situations, stimuli, treatments (e.g., starvation, hypoxia, interleukins, drugs) or developmental and differentiation stages. Differential comparison of mRNA interactomes has not been experimentally established and requires further development, especially regarding data analysis.

Experimental design

Controls. Include a nonirradiated cell (noCL) control for cCL, because proteins identified in this sample will represent contaminants captured by oligo(dT) magnetic beads in an UV cross-linking–independent manner. Similarly, include a 4SU-treated, nonirradiated cell (4SU noCL) control for PAR-CL samples. Because magnetic beads are relatively inert and nonspecific binding is typically low, we primarily recommend the noCL control for proteomics.

Adaptation. To determine optimal conditions adapted to the specific experimental aim at hand, scale down the preparative stage of the interactome capture protocol by up to 15-fold. In particular, cell lysis and homogenization (Steps 4–6 of the PROCEDURE) may require system-specific optimization. For such experiments, we recommend monitoring both RNAs (mRNA and rRNA) and proteins quantitatively and qualitatively (**Figs. 1** and **2**).

UV cross-linking. We found 0.15 J cm^{-2} to be the optimal dosage of 254 nm (cCL) and 365 nm (PAR-CL) UV light for adherent HeLa, Huh-7 and HEK293 cells. In principle, the lowest irradiation dosage that promotes efficient RBP cross-linking should be determined, and therefore the ideal dosage for other cell lines or biological systems may require optimization.

4SU incorporation. A concentration of 100 µM was found to be the optimal concentration of 4SU for PAR-CL with HeLa[4], Huh-7 (R.H. and A.C., unpublished data) and HEK293 (ref. 3) cells. However, the optimal concentration could vary for other cell types. Confluent cells or cells with long generation times may require higher 4SU concentration or longer 4SU incubation for sufficient incorporation of the nucleotide analog into RNAs. 4SU incorporation can be determined as previously described[11,17].

Cell lysis. Appropriate cell lysis is one of the most crucial aspects of interactome capture, and in some instances it may require optimization. The lysis buffer used typically yields highly viscous lysates. If the lysate is too viscous, contaminants can be trapped in the bead pellet, affecting the purity of the mRNA isolation. When the cell lysate is still very dense after homogenization, we suggest increasing lysis buffer volume and increasing the number of passages of the lysate through the narrow needle for homogenization (Step 6 of the PROCEDURE).

Bead recycling. To limit the use of costly reagents, poly(A) RNA capture is performed in three successive cycles (Steps 8–16 of the PROCEDURE), reusing the same beads (Step 14 of the PROCEDURE). Follow the manufacturer's recommendations for oligo(dT) bead recycling. If the amount of beads used does not efficiently deplete the sample of cellular mRNAs (measured by RT-quantitative PCR) after three rounds of isolation,

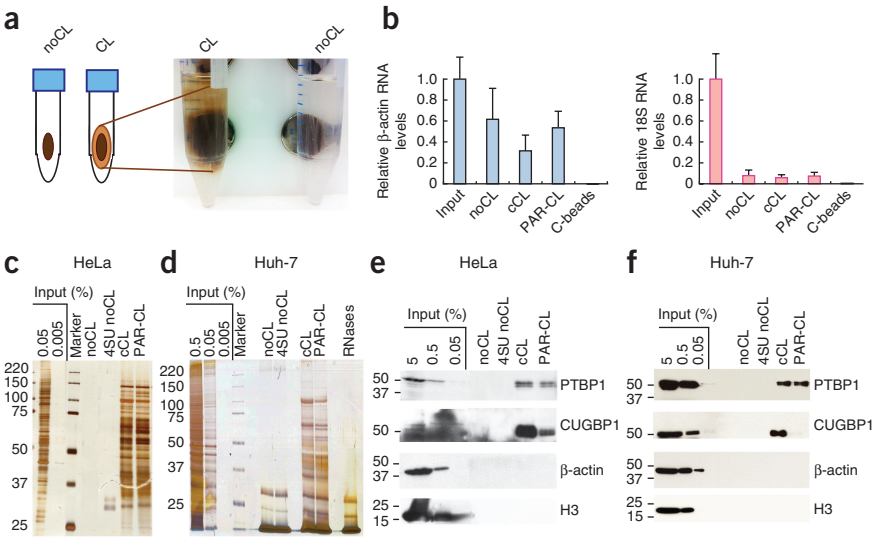

Figure 2 | Interactome capture as a selective protocol for capturing RBPs in HeLa and Huh-7 cells. After applying either cCL or PAR-CL, poly(A) RNAs were selected by oligo(dT) capture. As negative controls, nonirradiated cells incubated either with (4SU noCL) or without (noCL) 4-thiouridine were used. (**a**) A bead halo can typically be detected in samples from UV light–exposed cells around the pelleted beads during oligo(dT) capture. (**b**) Levels of 18S rRNA and β-actin in eluates were evaluated by RT-qPCR[4] (n = 4; ± s.d). (**c**,**d**) Eluted proteins from HeLa (**c**) and Huh-7 (**d**) cells were detected by SDS-PAGE and silver staining. (**e**,**f**) HeLa (**e**) and Huh-7 (**f**) eluates were analyzed by western blotting against PTBP1, CUGBP1, β-actin and histone H3.

we recommend increasing the amount of beads or the number of isolation rounds in subsequent experiments.

Proteomics. We performed proteomic analysis on a liquid chromatography–tandem MS (LC-MS/MS) platform, coupling an LC system to a mass spectrometer with MS/MS capability, such as a hybrid time-of-flight, Orbitrap or Fourier transform instrument (e.g., similar to Xu *et al.*[18]). To maximize the depth of protein identification, sample complexity can be reduced by peptide fractionation of trypsin-digested samples using isoelectric focusing (samples were analyzed in 12 independent fractions for the determination of the HeLa mRNA interactome[4], although fewer fractions may be sufficient depending on the application). Peptide fractions are then analyzed by high-resolution nano-LC-MS/MS. A key aspect is the assignment of proteins that specifically bind to mRNA as opposed to proteins identified in negative controls. In this protocol, we apply label-free quantification based on peptide counts, followed by statistical analysis (Steps 32–34 of the PROCEDURE). However, the interactome capture protocol is compatible with protein quantification methods using stable isotopic labeling (e.g., reductive methylation[19], iTRAQ[20,21], SILAC[22] and so on).

HPLC and MS parameters. Typically, a linear gradient of 5–25% solvent B over 4 h yields maximal HPLC separation and optimal sampling depth by in-line mass spectrometric detection. In principle, the gradient of solvent B should be set in accordance with the complexity of the sample. When combined with peptide fractionation up front, a shorter gradient per fraction will suffice (e.g., 2 h). Mass spectrometric settings (ion times, dynamic exclusion, collision energy) are dependent on the instrument that is used. They should be tuned to maximize the number of effective peptide identifications, but typically do not require specific adjustments for this protocol when an MS platform has been established for routine protein identification.

Importantly, all samples (including negative controls) need to be analyzed under the same LC-MS/MS conditions. Raw MS data should be processed via standard tools (e.g., MaxQuant[23] or vendor-specific software) to generate a file that can be subjected to peptide and protein identification using a database search algorithm (e.g., Mascot, Sequest, Andromeda[24]), interrogating a species-specific protein database that is appropriate for the sample (e.g., mouse, human).

Statistical data analysis. It is recommended that both samples and controls be measured in triplicate for biologically independent repeat experiments. For both label-free and labeled approaches, peptides are identified and summarized in protein groups (e.g., using Mascot software). Ion count measures for each peptide can be extracted by various software tools (e.g., MaxQuant or Scaffold (Proteomesoftware)). The ion count ratios between samples and controls (i.e., non–cross-linked samples) are tested against zero by a moderated t test. The same approach may be used to compare RNA binding between two or more experimental conditions; however, statistical methods may need further development and additional control samples or measurements may be required.

MATERIALS

REAGENTS

- Adherent HeLa (American type culture collection (ATCC), cat. no. CCL-2) or Huh-7 cells (kindly provided by M. Muckenthaler, University of Heidelberg). Maintain the cells at 37 °C in a humidified atmosphere containing 5% CO_2.
- DMEM with 4.5 mg l^{-1} D-glucose (Sigma-Aldrich)
- Heat-inactivated FBS (Gibco)
- Glutamine (Gibco, cat. no. G7513)
- Penicillin-streptomycin (Sigma-Aldrich, cat. no. P0781)
- PBS tablet (Sigma-Aldrich, cat. no. P4417)
- DTT (Biomol, cat. no. 04020.100)
- Complete EDTA-free proteinase inhibitor cocktail (Roche, cat. no. 11873580001)
- Oligo(dT_{25}) magentic beads (New England Biolabs, cat no. S1419S)
- Carboxy magnetic beads (control beads; New England Biolabs, cat. no. NEB900)
- 4-Thiouridine (4SU; Sigma-Aldrich, cat. no. T4509)
- Lithium dodecyl sulfate (LiDS; Sigma-Aldrich, cat. no. L9781)

- Lithium chloride (LiCl; Sigma-Aldrich, cat. no. 62476)
- Ammonium bicarbonate (NH_4HCO_3; Sigma-Aldrich, A6141)
- Iodoacetamide (IAA; Bio-Rad, cat. no. 163-2109)
- DTT (Bio-Rad, cat. no. 161-0611)
- Endoproteinase Lys-C (Wako, cat. no. 129-02541)
- Trypsin Gold, MS grade (Promega, cat. no. V5280)
- Trifluoroacetic acid (TFA; Biosolve, 2023413)
- Amicon Ultra centrifugal filters (50 ml, 10 kDa cutoff; Millipore, cat. no. UFC901024)
- Amicon Ultra centrifugal filters (0.5 ml, 3 kDa cutoff; Millipore, cat. no. UFC500396)
- Sep-Pak cartridges (Vac 1cc (50 mg) tC18; Waters, cat. no. WAT054960)
- RNase T1 from *Aspergillus oryzae* (Sigma-Aldrich, cat. no. R1003)
- RNase A from bovine pancreas (Sigma-Aldrich, cat. no. R4642)
- RNeasy kit (Qiagen, cat no. 74104)
- SuperScript II reverse transcriptase (Invitrogen, cat. no. 18064-014)
- SYBR Green PCR master mix (Applied Biosystems, cat. no. 4309155)
- NaCl in water, 0.5 M
- Endoproteinase Lys-C, stock: 1 $\mu g\ \mu l^{-1}$
- Trypsin, stock: 1 $\mu g\ \mu l^{-1}$
- Tris-HCl
- NP-40
- SDS
- $CaCl_2$

EQUIPMENT

- Humidified 37 °C, 5% CO_2 incubator
- Sterile hood
- Square dishes, 500 mm^2 (Nunclon, cat. no. 166508)
- GP Millipore express plus 500 ml funnel (Millipore, cat. no. SCGPT05RE)
- Cross-linking devices: cCL = 254 nm bulbs; PAR-CL = 365 nm bulbs
- Spectrolinker UV cross-linkers (Spectroline)
- Needle (27G, 3/4-inch; no. 20, 0.4 mm × 19 mm; BD Microlance)
- Sterile syringe (5 ml; BD Plastipak).
- Magnetic separation rack, 50 ml (NEB, S1507S). For analytical experiments, 12-tube (2 ml) Magnetic separation rack (S1509S, NEB or DynaMag-2, 123.21D, Invitrogen).
- RNA 6000 Pico Bioanalyzer chip and reagents (Agilent Technologies, 5067-1513 and 5067-151)
- 7500 Real-time PCR system (Applied Biosystems)
- Refrigerated benchtop centrifuge (temperature 20 °C)
- nano-HPLC system (e.g., Proxeon or nanoAcquity UPLC system (Waters) equipped with an appropriate C18 reversed-phase column, operating at 100–200 nl min^{-1})
- LTQ Orbitrap Velos (Thermo Fisher Scientific)
- Software for protein identification via database searching, e.g., Mascot (MatrixScience), Sequest (Thermo) or Andromeda[24] (via MaxQuant[23])
- R software (http://www.r-project.org/)
- Bioconductor software (http://www.bioconductor.org/)
- Nanodrop spectrophotometer
- Eppendorf 5415R centrifuge
- Falcon tubes

REAGENT SETUP

▲ **CRITICAL** Autoclave the buffers before the addition of LiDS and DTT to inactivate contaminants that may interfere with interactome capture (e.g., proteases, RNases). After cooling, complete the sterilized buffers by the addition of LiDS and DTT, and then filter them.

Lysis buffer Mix 20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 0.5% LiDS (wt/vol, stock 10%), 1 mM EDTA and 5 mM DTT. This buffer can be stored for up to 3 months at 4 °C.

Buffer 1 Mix 20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 0.1% LiDS (wt/vol), 1 mM EDTA and 5 mM DTT. This buffer can be stored for up to 3 months at 4 °C.

Buffer 2 Mix 20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 1 mM EDTA and 5 mM DTT. This buffer can be stored for up to 3 months at 4 °C.

Buffer 3 Mix 20 mM Tris-HCl (pH 7.5), 200 mM LiCl, 1 mM EDTA and 5 mM DTT. This buffer can be stored for up to 3 months at 4 °C.

Elution buffer Mix 20 mM Tris-HCl (pH 7.5) and 1 mM EDTA. This buffer can be stored for up to 1 year at 4 °C. We recommended making aliquots of this buffer in order to reduce potential contamination due to handling.

Buffer 4 Mix 10 mM Tris-HCl (pH 7.5) and 50 mM NaCl. This buffer can be stored for up to 1 year at 4 °C. We recommended making aliquots of this buffer in order to reduce potential contamination due to handling.

RNase buffer, 10× Mix 100 mM Tris-HCl (pH 7.5), 1.5 M NaCl, 0.5% (vol/vol) NP-40 and 5 mM DTT. This buffer can be stored for up to 1 year at 4 °C.

Proteinase K buffer, 5× Mix 50 mM Tris-HCl (pH 7.5), 750 mM NaCl, 1% (wt/vol) SDS, 50 mM EDTA, 2.5 mM DTT and 25 mM $CaCl_2$. This buffer can be stored for up to 1 year at room temperature (22 – 25 °C).

Buffer UA Buffer UA is 8 M urea in 0.1 M Tris-HCl (pH 8.5).

▲ **CRITICAL** Freshly prepare the buffer and use it within 1 d.

Buffer UB Buffer UB is 8 M urea in 0.1 M Tris-HCl (pH 8.0).

▲ **CRITICAL** Freshly prepare the buffer and use it within 1 d.

IAA solution IAA solution is 0.05 M IAA in buffer UA. ▲ **CRITICAL** Freshly prepare the solution and use it within 1 d.

HPLC solvent A HPLC solvent A is 0.1% (vol/vol) formic acid

HPLC solvent B HPLC solvent B is 0.1% (vol/vol) formic acid in acetonitrile (ACN)

PROCEDURE

▲ **CRITICAL** Timing information provided for Steps 1–16 is for a 'large-scale' experiment (~2.8 × 10^8 cells); the time required can be reduced by scaling down the experiment or by the participation of more than one person in the process.

Cell culture ● TIMING overnight

1| Culture HeLa cells at 40% confluence in 15 × 500 cm^2 dishes for each experimental and control sample overnight (7,500 cm^2 of total growth area) in normal medium (DMEM and 10% (vol/vol) FCS) for cCL, or use the same medium supplemented with 100 μM 4SU for PAR-CL. Typically, these cultures reach 80–90% confluence before UV cross-linking (~19 × 10^6 cells per dish).

Preparation for UV cross-linking ● TIMING ~30 min

2| Wash cells twice with 30 ml of PBS at room temperature.

▲ **CRITICAL STEP** DMEM must be completely removed. Repeat the washing step until the PBS remains colorless. 4SU-labeled control cells should be handled in a darkened room to avoid cross-linking mediated by natural light.

UV cross-linking ● TIMING ~1 h

3| Remove the PBS and place the culture dishes without their lids on ice (immediately before cross-linking) at a distance of ~10–15 cm from the UV source. Irradiate with 0.15 J cm^{-2} (~1 min) at 254-nm UV light for cCL or at 365-nm UV light for

4SU-labeled cells (PAR-CL). Add 15 ml of ice-cold PBS to each dish and place the treated dishes on ice while the process is repeated with the rest of the dishes.
▲ CRITICAL STEP It is important to process the cells quickly from Steps 3–5; therefore, if you are dealing with a large number of dishes, proceed with small sets (about 4–6 dishes at a time) while leaving the remaining dishes in culture.

Before lysis ● TIMING ~30 min
4| Scrape the cells with a rubber policeman into the 15 ml of ice-cold PBS added in Step 3, and then centrifuge the cells in a conical tube (400*g* for 3 min at 4 °C). Remove and discard the supernatant.

Lysis ● TIMING ~10 min
5| Lyse the cells in 105 ml of ice-cold lysis buffer and resuspend the pellet by pipetting up and down.

Homogenization ● TIMING 1–2 h
6| Pass the sample three times through a syringe with a narrow needle (gauge 0.4 mm diameter) to homogenize it. Incubate the lysate for 10 min at 4 °C.
? TROUBLESHOOTING

7| Aliquot each sample into three 50-ml tubes (35 ml per tube) to facilitate sample handling.
▲ CRITICAL STEP All volumes (magnetic beads, buffers, etc.) from Steps 7–13 have already been divided by three and therefore represent 'volume per tube'.
▲ CRITICAL STEP The three tubes (per condition) can be processed simultaneously or successively, even on different days, before being pooled again at Step 16. If the second option is implemented, proceed in parallel with experimental and control samples (one noCL, one cCL and/or one 4SU noCL, one PAR-CL).
▲ CRITICAL STEP Avoid repeated freeze-thaw cycles to prevent RNA degradation.
■ PAUSE POINT The samples can be frozen at −80° C for up to 1 week.

Oligo(dT) capture ● TIMING 1 d
8| Add 3 ml of oligo$(dT)_{25}$ magnetic beads per tube, previously equilibrated in three volumes of lysis buffer, to the lysate and incubate them for 1 h at 4° C with gentle rotation. Place the tubes on a magnet at 4 °C and wait until the beads are completely captured (this process may take up to 30 min). Recover the supernatant and keep it in a new tube at 4 °C for use in the other two cycles of oligo(dT) capture (Step 14).

9| Resuspend the beads in 35 ml of ice-cold lysis buffer. Incubate the beads for 5 min at 4 °C with gentle rotation and pellet the beads with the help of a magnet. Remove and discard the supernatant.

10| Add 35 ml of ice-cold buffer 1, resuspend the beads and wash them for 5 min at 4 °C with gentle rotation. Pellet the beads with the magnet and discard the supernatant. Repeat this step once. Note that when the mRNA-protein complexes are efficiently isolated a halo surrounding the bead pellet will be noticeable in the experimental (cross-linked) sample but not in control samples, especially in large-scale experiments (**Fig. 2a**).

11| Add 35 ml of ice-cold buffer 2, resuspend the beads and wash them for 5 min at 4 °C with gentle rotation. Pellet the beads with the magnet and discard the supernatant. Repeat this step once.

12| Add 35 ml of ice-cold buffer 3, resuspend the bead pellet and wash it for 5 min at 4 °C with gentle rotation. Pellet the beads with the magnet and discard the supernatant. Repeat this step once.

13| Elute the mRNA-protein complexes with 500 µl of elution buffer for 3 min at 55 °C. Determine the RNA concentration using a Nanodrop device.
? TROUBLESHOOTING
■ PAUSE POINT The samples can be stored at −80° C for up to 1 week.

14| Repeat Steps 8–13 twice more to deplete samples of poly(A) RNAs using recycled oligo(dT) beads. Follow the manufacturer's recommendations for recycling the beads.
▲ CRITICAL STEP Do not reuse the beads from one 'condition' for other conditions (e.g., beads used in cCL cannot be used for control samples) in order to avoid cross-contamination.

15| Pool the eluates from the three successive cycles of poly(A)$^+$ RNA capture (final volume 1.5 ml).
■ **PAUSE POINT** The samples can be stored at −80 °C for up to 1 week.

16| Pool all the eluates from the same sample (see Step 7 of the PROCEDURE; final volume, 4.5 ml).

RNA analysis ● TIMING 1 d
17| Take 20 µl of each pooled eluate (from Step 16) and add 5 µl of 5× proteinase K buffer and 1 µg of proteinase K. Incubate the mixture for 1 h at 50 °C and isolate the RNA with the RNeasy kit (Qiagen) or TRIzol (Invitrogen). Use these protein-free samples for RNA analyses (**Fig. 1**). We typically assess RNA quality by RT-qPCR with primers against rRNAs and mRNAs and/or Bioanalyzer chips. RNA sequencing can be also used to define in depth the RNA populations present in pull-downs[4].
? TROUBLESHOOTING

RNA digestion and protein concentration ● TIMING ~3 h
18| Add 500 µl of 10× RNase buffer and ~200 U of RNase T1 and RNase A to the remainder of the pooled eluate from Step 16 (4,480 µl). Incubate the mixture for 1 h at 37 °C. This step can be performed using other RNases, with the exception of those incompatible with EDTA (e.g., micrococcal nuclease).

19| Transfer the RNase-treated eluate to an Amicon Ultra 10 3-kDa cutoff device (a 50-ml device for large-scale experiments, and a 2-ml device for small scale experiments). Top up the Amicon device with buffer 4 and centrifuge at 4000*g* for 45 min at 4 °C.

20| Discard the flow-through and top up the Amicon device again with buffer 4. Centrifuge at 4,000*g* for 45 min at 4 °C.

21| Recover the sample from the Amicon device (~200 µl).
■ **PAUSE POINT** The samples can be stored indefinitely at −20 °C.

Protein quality control analysis ● TIMING 2 d
22| Before MS, take 30 µl of the 'final' sample from Step 21 for the following analyses: use 5 µl of sample for total protein determination by Bradford analysis; use 15 µl of sample for protein quality control using silver staining; and use 10 µl for western blotting (e.g., CUGBP, Elav-like family member 1 (CUGBP; Santa Cruz Biotechnology, SC-20003) and polypyrimidine tract binding protein 1 (PTBP1; Sigma-Aldrich, clone 3H8, WH0005725M1) as positive controls; and β-actin (Sigma-Aldrich, cat no. A5441), α-tubulin (Sigma-Aldrich, cat. no. T8203), histone (H)3 (Abcam, cat. no. ab1791) and histone H4 (Abcam, cat. no. ab10158) as negative controls). Use the remaining sample from Step 21 for MS. A large-scale interactome capture experiment typically yields 200–300 µg of protein (10–20 µg for a small-scale (1.9×10^7 cells) experiment).
? TROUBLESHOOTING
■ **PAUSE POINT** The samples can be frozen indefinitely at −20 °C.

Sample preparation before MS ● TIMING 2 d
▲ **CRITICAL** The samples are buffer-exchanged, concentrated, reduced, alkylated and digested using Amicon Ultra centrifugal filters (0.5 ml, 3-kDa cutoff) according to the FASP protocol[25] as indicated below:

23| Adjust the samples to 0.1 M DTT by adding the appropriate amount of 1 M DTT. Heat the mixture to 95 °C for 5 min.

24| Add 200 µl of buffer UA to the sample, and then mix and transfer the mixture to the Amicon centrifugal filter. Concentrate the sample by centrifugation at maximum speed (16,000*g* in an Eppendorf 5415R centrifuge) for ~30 min at 20 °C.

25| Add 100 µl of IAA solution to the samples; mix the samples in a thermomixer at 600 r.p.m. for 1 min, and then incubate them for 5 min without mixing. Transfer the centrifugal filter to the centrifuge and concentrate the proteins at maximum speed for 30 min at 20 °C.

26| For buffer exchange, add 100 µl of buffer UB and mix it properly. Concentrate the sample by centrifugation at maximum speed (16,000*g* in an Eppendorf 5415R centrifuge) for ~30 min at 20 °C. Repeat this step two more times.

Protein digestion, peptide labeling and fractionation ● TIMING 5 d
27| Add 0.5 µg of endoproteinase Lys-C in 40 µl of buffer UB to the Amicon filter and mix the sample in the thermomixer at 600 r.p.m. for 1 min. Incubate the sample in the Amicon filter at room temperature overnight.

28| Add 0.5 µg of trypsin in 120 µl of 50 mM NH_4HCO_3 and incubate the mixture at room temperature for 4 h.

29| Collect the peptides by centrifugation at maximum speed (16,000*g* in an Eppendorf 5415R centrifuge) for ~30 min at 20 °C and wash the filter unit with 50 µl of 0.5 M NaCl. Acidify the combined sample by adding 10% (vol/vol) TFA and desalt using Sep-Pak cartridges (Vac 1cc (50 mg) tC18) as described elsewhere[26]. Optionally, stable isotope labeling can be performed at this point (e.g., by reductive methylation according to the protocol of Boersema *et al.*[19]). This has been successfully tested in Huh-7 cells (R.H., unpublished data). Peptide fractionation can also be optionally performed to reduce the complexity of the sample before MS, e.g., by isoelectric focusing[27] or strong cation exchange (SCX) chromatography[26].
■ **PAUSE POINT** The samples can be frozen indefinitely at −20 °C until analysis by LC-MS/MS.

Protein identification by LC-MS/MS ● TIMING 1 d
30| Inject an appropriate amount of digested protein to optimally load the LC column (typically ~1 µg). Apply a gradient in HPLC solvent B that is appropriate for the complexity of the sample (see Experimental design).
▲ **CRITICAL STEP** Analyze all the samples (including negative control) using the same MS parameters.

31| Compute a false discovery rate (FDR) of 1% for both peptide and protein identification.

Protein quantification and data analysis ● TIMING ~1 week
32| Extract ion count measures for each peptide with MaxQuant software[28]. Summarize the Ion count ratios between sample and control to protein ion count ratios per replicate.

33| Test ion count ratios against the hypothesis that ratios are zero using three or more replicates by a moderated *t* test implemented in the R/bioconductor package limma[29,30].

34| Correct *P* values by the method of Benjamini-Hochberg, controlling for FDR. We apply an FDR of 1%.

? TROUBLESHOOTING
Troubleshooting advice can be found in **Table 1**

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
6	Viscous lysate	Chromatin precipitation	Increase the volume of lysis buffer
			Repeat the homogenization process
			Lyse the sample in conventional NP40 lysis buffer and perform a DNAse treatment before addition of LiDS lysis buffer
13	Inefficient elution	RNA retention on oligo(dT) beads	Add RNase T and RNase A to the elution buffer and incubate for 10 min at 37°C before applying the normal elution protocol
17, 22	Low RNA and protein yield in eluates	Inefficient mRNA capture	Increase the starting material (number of cell equivalents)
	Low RNA and protein yield in eluates	RNA is degraded upon UV irradiation	Optimize UV dosage
17	Low RNA recovery	Low performance of oligo(dT) beads	Increase the amount of oligo(dT) beads or/and isolation rounds (typically we perform three cycles)
22	Low protein recovery	Inefficient cross-linking	Replace UV bulbs, especially if you see a black halo. Optimize UV irradiation dosage
	Low protein recovery	Cells are lysed during scraping	For cell types sensitive to scraping, lysis can be performed by adding the lysis buffer directly to the plate
	Low protein recovery in PAR-CL sample	Inefficient incorporation of 4SU into RNAs	Increase 4SU concentration; Increase incubation time with 4SU

● TIMING

Note that the time required to carry out Steps 1–16 can be reduced by scaling down the experiment or by having more than one person involved in the experiment.

Step 1, cell culture: overnight
Step 2, preparation for UV cross-linking: ~30 min
Step 3, UV cross-linking: ~1 h
Step 4, before lysis: ~30 min
Step 5, lysis: ~10 min
Steps 6 and 7, homogenization: 1–2 h
Steps 8–16, oligo(dT) capture: 1 d
Step 17, RNA analysis: 1 d
Steps 18–21, RNA digestion and protein concentration: ~3 h
Step 22, protein quality control analysis: 2 d
Steps 23–26, sample preparation before MS: 2 d
Steps 27–29, protein digestion, peptide labeling and fractionation: 5 d
Steps 30 and 31, protein identification by LC-MS/MS: 1 d
Steps 32–34, protein quantification and data analysis: ~1 week

ANTICIPATED RESULTS

Interactome capture should enrich mRNAs over rRNAs (which are the most abundant RNA species in the cell), and this can be validated by different techniques, including RT-qPCR, Bioanalyzer and next-generation sequencing as illustrated in **Figure 2b** and in Castello *et al.*[4]. Gel electrophoresis combined with silver staining should be applied to analyze the protein pattern of experimental and control samples. Both cCL and PAR-CL samples should yield protein patterns that notably differ from that of the (diluted) whole-cell lysate (**Fig. 2c**). Conversely, the negative control (nonirradiated) samples are expected to be devoid of contaminating proteins (**Fig. 2c**). The presence of proteins in control lanes (with the exception of RNases used in Step 18 of the PROCEDURE) reflects the need for troubleshooting (**Table 1**). Although both cross-linking protocols yield similar protein patterns, some differences are expected to be observed, representing RBPs favored by either of the two UV-cross-linking protocols. Reduced protein content in PAR-CL samples may reflect insufficient 4SU incorporation into nascent RNAs, which can be determined as previously described[11,17] (**Table 1**). Interactome capture leads to similar protein patterns in cross-linked samples from different mammalian cell lines, e.g., HeLa[4], Huh-7 (**Fig. 2c**,**d**) and HEK293[3] cells. This similarity suggests that the most abundant RBPs, which are preferentially detected by silver staining, are commonly expressed and active in these different cell types. Therefore, the protein pattern observed in HeLa cells can be used as a benchmark for other mRNA interactomes from at least other mammalian sources.

Interactome capture strongly enriches for well-known RBPs, such as PTBP1 (**Fig. 2e**,**f**). The absence of well-known RBPs in cross-linked samples may reflect inefficient UV irradiation or loss of RNA during the lysis or isolation process (**Table 1**). Although both cross-linking methods capture CUG triplet repeat RNA-binding protein 1 (CELF1 or CUGBP1), cCL performs better than PAR-CL for this protein in both HeLa and Huh-7 cells (**Fig. 2e**,**f**). Conversely, the Y box–binding protein 1 (YBX1) cross-links more effectively with PAR-CL[4]. These results show that the choice of the UV-cross-linking method can affect the outcome of interactome capture, and that the two protocols are complementary for the generation of comprehensive RBP atlases. Highly abundant proteins that do not bind RNA, such as α-actin and the DNA-binding histone 3, should not be detected in cross-linked samples by western blotting (**Fig. 2e**,**f**). The presence of these proteins in eluates indicates the need for troubleshooting (**Table 1**).

Interactome capture can determine whether a protein of interest displays poly(A) RNA–binding activity *in vivo*. Serine hydroxymethyltransferase 2 (SHMT2) is an enzyme of intermediary metabolism that was recently found also to bind RNA[4]. Interactome capture was applied to ~1.9 × 10^7 HeLa cells to determine whether SHMT2 interacts with poly(A) RNA in living cells. As expected for a bona fide RBP, SHMT2 is enriched in cCL and PAR-CL samples and undetectable in controls (nonirradiated cells) (**Fig. 3a**). Conversely, both control and cross-linked samples are devoid of α-tubulin and the metabolic enzyme transkelotase, which was not identified in the HeLa mRNA interactome (**Fig. 3b**).

A large-scale interactome capture experiment (2.8 × 10^8 cells, 200–300 μg of total protein) may yield more than

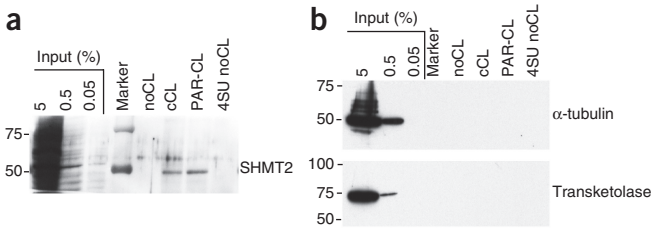

Figure 3 | Interactome capture applied to determine *in vivo* RNA-binding activities of candidate proteins of interest. cCL or PAR-CL protocols were applied to HeLa cells (~1.9 × 10^7). Poly(A) RNA in cross-linked and non–cross-linked samples was captured with 500 μl of oligo(dT) using three rounds of isolation. (**a**,**b**) Eluted proteins were analyzed by western blotting with antibodies against SHMT2 (**a**) and α-tubulin and transkelotase (**b**).

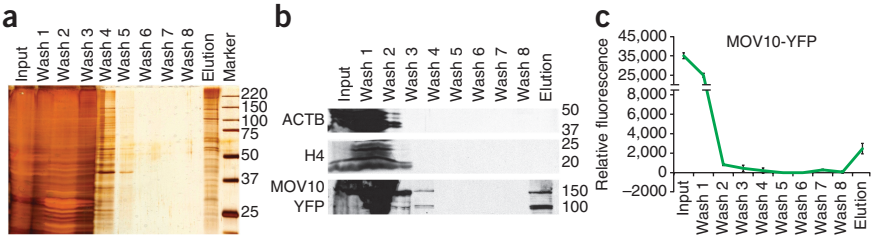

Figure 4 | Monitoring protein content in input, wash and eluate samples of a cellular interactome capture experiment. (**a**–**c**) Proteins in input, wash and eluate samples of an interactome capture experiment with the HeLa MOV10-YFP-expressing cell line[4] (induction: 16 h with tetracycline) were examined by silver staining (**a**), western blotting against β-actin (ACTB), H3 and YFP (**b**) and YFP fluorescence measurement using a Sapphire II fluorometer device (**c**).

1,000 proteins identified by MS and more than 700 different proteins statistically enriched in cross-linked fractions[3,4]. By contrast, only around 500 different proteins were identified in cross-linked fractions when interactome capture was applied to a lower quantity of starting material (1.9 × 10^7 HeLa cells, 10–20 µg of total protein) (A.C., unpublished data). It is important to perform quality control analyses as indicated above with each large-scale interactome capture sample before MS. mRNA interactomes can be benchmarked with Gene Ontology (GO) and InterPro domain analyses, in which gene ontology terms related to RNA metabolism and RNA-binding domain should be prevalent[4].

To further determine the selectivity and sensitivity of interactome capture, we analyzed proteins in input, wash supernatant and eluate samples. For this experiment, we made use of a HeLa stable cell line expressing the RBP MOV10 fused to the yellow fluorescent protein (YFP)[4]. Silver staining reveals that the protein patterns of supernatants 1–4 are very similar to those of the whole-cell lysate, whereas no proteins are detectable in supernatants 6–8 (**Fig. 4a**). As shown in **Figure 2c**,**d**, the pattern of proteins eluted from the oligo(dT) beads is distinct from input samples and represents the HeLa mRNA interactome (**Fig. 4a**). As expected, β-actin and H3 are present in the first three fractions, but absent in the following supernatants and in the eluate, demonstrating that the interactome capture protocol effectively removes proteins that do not bind RNA (**Fig. 4b**). We also found MOV10-YFP in the first four supernatants, probably representing the non–cross-linked fraction of this protein, and it was undetectable in the following wash supernatants (**Fig. 4b**,**c**). In contrast to the negative controls, ~6% of the initial MOV10-YFP was recovered in the eluate (**Fig. 4b**,**c**), which correlates well with the UV cross-linking efficiency reported for most RBPs (**Fig. 2e**,**f**)[4]. Collectively, these results demonstrate the high degree of specificity of interactome capture. As exemplified here with MOV10, interactome capture can be applied to cells expressing RBP fused to fluorescent tags to monitor its RNA-binding activity under different experimental conditions by simple fluorescence measurement.

The activity of many RBPs is regulated post-transcriptionally in response to altered biological states (metabolism, stress, signaling and so on). As only the active fraction of RBPs can be cross-linked to RNAs, comparative quantitative analysis of interactomes is envisaged to generate informative activity landscape maps, shedding light on post-transcriptional response networks in a system-wide manner.

ACKNOWLEDGMENTS We thank M. Landthaler (Max Delbrück Center for Molecular Medicine) for generously sharing his expertise on PAR-CL. We are grateful to R. Pepperkok (EMBL) for plasmids and M. Gromeier (Duke University Medical Center) for the HeLa Flip-in TRex cell line. We thank M. Muckenthaler (University of Heidelberg) for Huh-7 cells. We acknowledge A. Perez and the EMBL Flow Cytometry Core Facility for FACS experiments, and EMBL Gene and Proteomics Core Facilities for support throughout this work. We also thank Chromotek for expert technical advice and support. A.C. is the beneficiary of a Marie Curie postdoctoral fellowship (FP7). M.W.H. acknowledges support by the European Research Council (ERC) Advanced Grant ERC-2011-ADG_20110310 and the Virtual Liver Network of the German Ministry for Science and Education.

AUTHOR CONTRIBUTIONS A.C., J.K., L.M.S. and M.W.H. contributed to the conception and design of the project. A.C., R.H., C.S., B.F. and K.E. carried out the experimental work. A.C. and M.W.H. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nprot.2013.020. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. & Tuschl, T. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev. RNA* **3**, 159–177 (2012).
2. Konig, J., Zarnack, K., Luscombe, N.M. & Ule, J. Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.* **13**, 77–83 (2011).
3. Baltz, A.G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
4. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
5. Scherrer, T., Mittal, N., Janga, S.C. & Gerber, A.P. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS ONE* **5**, e15499 (2010).
6. Tsvetanova, N.G., Klass, D.M., Salzman, J. & Brown, P.O. Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS ONE* **5**, e12671 (2010).
7. Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc. Natl Acad. Sci. USA* **106**, 10626–10631 (2009).
8. Anantharaman, V., Koonin, E.V. & Aravind, L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30**, 1427–1464 (2002).
9. Brimacombe, R., Stiege, W., Kyriatsoulis, A. & Maly, P. Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol.* **164**, 287–309 (1988).
10. Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R. & von Hippel, P.H. Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. *J. Biol. Chem.* **261**, 3512–3518 (1986).
11. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
12. Pashev, I.G., Dimitrov, S.I. & Angelov, D. Cross-linking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* **16**, 323–326 (1991).

13. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2**, 261–267 (2005).
14. Creamer, T.J. *et al.* Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* **7**, e1002329 (2011).
15. Jamonnak, N. *et al.* Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* **17**, 2011–2025 (2011).
16. Jungkamp, A.C. *et al. In vivo* and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell* **44**, 828–840 (2011).
17. Hafner, M. *et al.* PAR-CliP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* **41**, e2034 published online; doi:10.3791/2034 (2010).
18. Xu, T., Wong, C.C., Kashina, A. & Yates, J.R. Identification of N-terminally arginylated proteins and peptides by mass spectrometry. *Nat. Protoc.* **4**, 325–332 (2009).
19. Boersema, P.J., Aye, T.T., van Veen, T.A., Heck, A.J. & Mohammed, S. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* **8**, 4624–4632 (2008).
20. Choe, L. *et al.* 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **7**, 3651–3660 (2007).
21. Ross, P.L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* **3**, 1154–1169 (2004).
22. Ong, S.E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1**, 2650–2660 (2006).
23. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
24. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
25. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
26. Villen, J. & Gygi, S.P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630–1638 (2008).
27. Krijgsveld, J., Gauci, S., Dormeyer, W. & Heck, A.J. In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J. Proteome Res.* **5**, 1721–1730 (2006).
28. Cox, J. *et al.* A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* **4**, 698–705 (2009).
29. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3** published online; doi:10.2202/1544-6115 (2004).
30. Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. & Huber, W. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (Springer,, 2005).