HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2019 March 20.

Published in final edited form as:

J Proteome Res. 2018 September 07; 17(9): 3022–3038. doi:10.1021/acs.jproteome.8b00189.

HyPR-MS for Multiplexed Discovery of MALAT1, NEAT1, and NORAD lncRNA Protein Interactomes

Michele Spiniello[†,#], **Rachel A. Knoener**[†,#], **Maisie I. Steinbrink**[†,‡], **Bing Yang**[§], **Anthony J. Cesnik**[†], **Katherine E. Buxton**[†], **Mark Scalf**[†], **David F. Jarrard**[‡,§,∥], and **Lloyd M. Smith**[*,†,⊥]

[†]Department of Chemistry, University of Wisconsin Madison, Madison, Wisconsin 53706, United States

[‡]Molecular and Environmental Toxicology, University of Wisconsin Madison, Madison, Wisconsin 53706, United States

[§]Department of Urology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin 53705, United States

[∥]Carbone Comprehensive Cancer Center, University of Wisconsin Madison, Madison, Wisconsin 53792, United States

[⊥]Genome Center of Wisconsin, University of Wisconsin Madison, Madison, Wisconsin 53706, United States

Abstract

RNA–protein interactions are integral to the regulation of gene expression. RNAs have diverse functions and the protein interactomes of individual RNAs vary temporally, spatially, and with physiological context. These factors make the global acquisition of individual RNA–protein interactomes an essential endeavor. Although techniques have been reported for discovery of the

[*]**Corresponding Author** smith@chem.wisc.edu. Tel: 608-263-2594.

[#]These authors contributed equally to this work.

Author Contributions

HyPR-MS development and optimization was initially performed by R.A.K with additional development and optimization con-ducted by M.Sp. and M.I.S. Study design and target selection was initiated by R.A.K. and K.E.B. with assistance from A.J.C. and an additional target selected by M.Sp. and M.I.S. Cell cultures were maintained and supplied by B.Y. Implementation of HyPR-MS in PC3 cells, RT-qPCR experiments and data analysis, protein sample preparation, and RIP-qPCR analysis were conducted by M.Sp. and M.I.S. Protein sample preparation and mass spectrometer maintenance and method development were conducted by M.Sc. R.A.K., M.Sp., and M.I.S. conducted data interpretation and manuscript writing. eCLIP validation analysis was conducted by M.Sp. and A.J.C. D.F.J. and L.M.S. provided support and advice during development and implementation.

Notes

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00189.

HyPR-MS protocol; HyPR-MS advances the capabilities of in vivo RNA capture technologies; determination of capture efficiencies and cell numbers required for protein detection; capture oligonucleotide (CO) and release oligonucleotide (RO) sequences and amounts used; primer and probe sequences for qPCR assays; qPCR amplification parameters; capture efficiencies (CE) at each qPCR region for each target lncRNA (PDF);

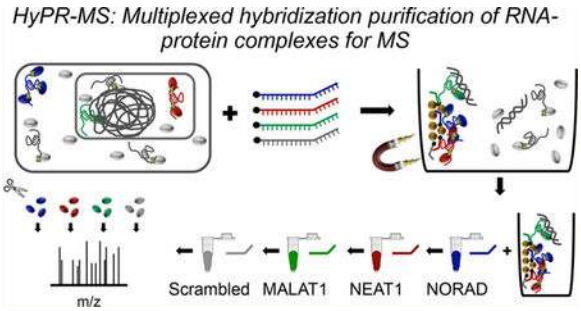
qPCR assay data statistics (Table S5); Protein Intensities and Fold Enrichments for MALAT1, NEAT1, and NORAD Protein Interactomes (Table S6); Protein intensities and fold enrichments for MALAT1, NEAT1, and NORAD protein (XLSX)

Gene ontology term enrichment analysis for MALAT1, NEAT1, and NORAD protein interactomes (Table S7); Summary of validated proteins (Table S8); eCLIP Validation data analysis (Table S9)

Note: A detailed protocol can be found in the Supporting Information (SI).

protein interactomes of specific RNAs they are largely laborious, costly, and accomplished singly in individual experiments. We developed HyPR-MS for the discovery and analysis of the protein interactomes of multiple RNAs in a single experiment while also reducing design time and improving efficiencies. Presented here is the application of HyPR-MS to simultaneously and selectively isolate the interactomes of lncRNAs MALAT1, NEAT1, and NORAD. Our analysis features the proteins that potentially contribute to both known and previously undiscovered roles of each lncRNA. This platform provides a powerful new multiplexing tool for the efficient and cost-effective elucidation of specific RNA–protein interactomes.

Graphical Abstract



Keywords

lncRNA; RNA-binding proteins; mass spectrometry; proteomics; hybridization capture; interactomes; RNA

INTRODUCTION

RNA–protein interactions are crucial to a wide variety of biological processes involved in gene expression including transcription, nuclear organization, splicing, and translation.[1] The highly coordinated interplay between RNAs and proteins is vital to proper cell physiology, and its dysregulation causes many human diseases. Knowledge of the RNA-binding proteome in normal and diseased states is thus essential to the understanding of RNA biology and for the development of efficient therapies.[2]

Several experimental strategies have been developed for the study of RNA–protein interactions.[3] Among these are RNA-centric methods, which aim to reveal proteins associated with a specific RNA. *In vivo* RNA-centric approaches, like ChIRP-MS (comprehensive identification of RNA binding proteins by mass spectrometry),[4] CHART-MS (capture hybridization analysis of RNA targets and mass spectrometry),[5] and RAP-MS (RNA antisense purification and mass spectrometry)[6] have become gold standards in the field because of their ability to reveal RNA-interacting proteins under specific spatiotemporal and biological contexts. The commonalities among these *in vivo* strategies include the exposure of cells to a protein–RNA cross-linker, capture of the target RNA by specific biotinylated capture oligonucleotides and, after elution, identification of the RNA-associated proteome by mass spectrometry.[7] These technologies have been used to identify the interacting proteomes of MALAT1, NEAT1, and Xist lncRNAs in various cell types.[4–6]

However, several unaddressed challenges with these technologies reduce their widespread application, including (a) capture oligonucleotide design limitations and (b) costly and labor-intensive procedures.[2]

Long noncoding RNAs (lncRNAs) are transcripts compris-ing 200 or more nucleotides that do not encode a protein product. Over the last two decades, lncRNAs have increasingly been studied for their role in cancer development and progression.[8] In particular, the dysregulation of lncRNAs MALAT1 and NEAT1 has been observed in a large spectrum of cancers, including prostate cancer.[9–12] Both MALAT1 and NEAT1 are abundant, ubiquitously expressed, and nuclear-localized.[5, 13,14] MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) is 8.7 kilobases (kb) in length and is an active regulator of alternative splicing through association with protein factors such as SRSF1, SC35, and SRSF3.[13] It also interacts in a tissue-specific manner with other nucleic acids and proteins such as TDP-43 and AGO2 to control various aspects of gene expression, cell growth, synapse formation, and the cell cycle.[13,15] NEAT1 (nuclear enriched abundant transcript 1; 3.7 kb in length) is a necessary component of paraspeckles, and its removal results in the disintegration of this nuclear body.[16] It is essential for paraspeckle biogenesis, interacting with known paraspeckle components such as NONO, SFPQ, and PSPC1, among other proteins.[10,14] While the function of paraspeckles is unclear, they have been shown to be involved in cell differentiation, alternative splicing, and the cellular response to stress through sequestration of transcripts and proteins.[17–20] NEAT1 also acts independently of paraspeckles, regulating microRNA levels and epigenetic gene expression by sequestering RNAs using sponge-like activity, and through direct interaction with the chromatin.[10]

NORAD (noncoding RNA activated by DNA damage), or LINC00657, is 5.3 kb in length and is a highly expressed lncRNA in many cancers and ubiquitously present in normal tissue, suggesting a relevant cellular function.[21,22] NORAD has been called the “defender of the genome” due to its role in the preservation of chromosomal stability; therefore, its dysregulation may implicate it in tumorigenesis.[23] Several manuscripts also discuss the molecular mechanism of NORAD regarding the induction of hypoxia in endothelial cells and other stress pathways.[24,25] Considering this, elucidation of the NORAD in vivo interacting proteome has high potential to be biologically important.[23] Currently, the documented primary NORAD interactors are the translational regulators Pumilio 1 and Pumilio 2 (PUM1 and PUM2).[22,24,26] NORAD can act as a molecular decoy for PUM1 and PUM2 via several repetitive binding motifs within its sequence allowing the binding of at least 15 Pumilio proteins per molecule. Several other NORAD interactors, such as XRN2, IGF2BP1/2/3, and PABPN1, have been hypothesized to interact with NORAD based on *in vitro* RNA pulldowns in a human osteosarcoma cell line (U2OS).[22] Elucidation of the *in vivo* interactome can be used to confirm these hypothesized NORAD interactors and expand the breadth of knowledge regarding its biological function.

In the present work we describe HyPR-MS (hybridization purification of RNA−protein complexes followed by mass spectrometry)[27] as a multiplexed *in vivo* RNA-centric method and apply it to discover, concurrently, the RNA−protein interactomes of three lncRNAs in the human prostate cancer cell line, PC3.[28] Briefly, HyPR-MS utilizes biotinylated capture oligonucleotides specifically complementary to RNAs of interest to capture the target RNA

–protein complexes from formaldehyde-cross-linked PC3 cell lysate. Following isolation of the capture oligonucleotide-target RNA hybrid using streptavidin-coated magnetic-beads, the RNA–protein complex is released from the beads using a sequence-specific release oligonucleotide strategy that permits selective release and isolation of multiple RNA targets from a single lysate sample. The proteins associated with the RNA target are purified, trypsin digested, and analyzed using mass spectrometry (Figure 1). Here, HyPR-MS is applied to identify the protein interactomes of three lncRNAs, MALAT1–001 (specifically, the full-length variant), NEAT1, and NORAD in PC3 cells. The functions and interacting proteins of MALAT1 and NEAT1 have been the subjects of several studies using various cell lines and therefore serve here, in part, as controls for evaluating the performance of the HyPR-MS technology. However, while many RNA–protein interactions may be common between different cell and tissue types, many may also be different due to variations in cellular contexts or in lncRNA, protein, and proteoform expression.[13,29–31] An additional feature of HyPR-MS is the ability to capture specific RNA splice variants. Only one splice variant of the 17 known variants of MALAT1[32] is targeted here, allowing for the assignment of the protein interactome to only the full length MALAT1–001 transcript (hence forward referred to as only MALAT1). Using HyPR-MS, 127, 94, and 415 interacting proteins were identified for MALAT1, NEAT1, and NORAD, respectively. Of these proteins, several have been previously identified; these include SRSF and PRPF proteins for MALAT1,[9,33] NONO and SFPQ for NEAT1,[34] and PUM1 for NORAD.[22,24] The efficacy of HyPR-MS is further demonstrated by the discovery of many novel interactors with functions relating to documented features of their associated lncRNAs, including several histone modifiers and transcriptional regulators found to interact with MALAT1 and regulators of gene expression associated with NORAD. The novel design workflow of HyPR-MS achieves high efficiency, specificity, and multiplexing capability, and the results shown here demonstrate the ability of HyPR-MS to discern the unique protein interactomes of multiple RNA species, simultaneously, from one lysate sample.

EXPERIMENTAL SECTION

Special Considerations for HyPR-MS

All solutions used for cell lysis, hybridization, capture, release, and RT-qPCR were prepared using certified RNase free components. The amount of lysate, the concentration of capture oligonucleotides, and the volume of streptavidin coated magnetic beads needed in each capture experiment for identification of proteins interacting with each lncRNA is conditional on the number of copies of the target RNA present in the cell culture. The capture oligonucleotide concentration and volume of streptavidin coated magnetic beads needed for optimal capture efficiency and specificity for each biological replicate were empirically determined. To do this, small-scale capture experiments using 5×10^5 cells were performed using increasing amounts of capture oligonucleotides and streptavidin coated beads. The amount of target RNA captured was then measured using RT-qPCR (as described below and in the SI) and the oligonucleotide concentration and bead volume producing the highest capture efficiency while maintaining a desirable capture specificity was scaled up for the large-scale capture experiments. The number of cells needed for a capture experiment for protein identification was then estimated and empirically confirmed by mass spectrometry

(Figure S1B). For the data presented here, 1×10^8 PC3 cells and 2.8 mL of streptavidin coated beads were used for each biological replicate with the capture oligonucleotide and release oligonucleotide concentrations as indicated in Table S2.

Capture Oligonucleotide Design

HyPR-MS uses Mfold software[35] to predict RNA secondary structure and inform capture oligonucleotide design. The folding temperature was set at 37 °C, the same temperature used for HyPR-MS hybridization, and nucleotide regions with the highest probability to be single stranded in the most thermodynamically favorable structures were considered. Capture oligonucleotides complementary to these potentially accessible regions were then filtered based on their likelihood to form hairpins, to self-dimerize, and to form heterodimers as determined by publicly available software (https://www.idtdna.com/calc/analyzer). Two to three capture oligonucleo-tides were chosen for each target (Table S2).

Cell Culture

PC3 prostate cancer cells (ATTC CRL-1435, RRID: CVCL_0035) were cultured in DMEM medium containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 37 °C and 5% CO_2. Cells were washed twice with phosphate buffered saline (PBS) then incubated in 1% formaldehyde for 10 min at room temperature. The formaldehyde was then quenched by incubation with 250 mM Tris-HCl buffer (pH 8.0) for 10 min at room temperature. Cells were then rinsed with cold PBS twice, pelleted by centrifugation at 1200 rpm for 5 min, and flash frozen in liquid nitrogen and stored at −80 °C. PC3 cells originated in a human male and were collected from a bone metastatic site. The cell line identity was authenticated by the supplier using short tandem repeats.

Cell Lysis

Cells were resuspended on ice in lysis buffer (469 mM LiCl, 62.5 mM Tris HCl, pH 7.5, 1.25% LiDS, 1.25% Triton X-100, 12.5 mM Ribonucleoside Vanadyl Complex (RVC), 12.5 mM DTT, 125U/mL RNasin Plus, 1.25× Halt Protease Inhibitors) to a final cell concentration of 5×10^6 cells/mL. Cells were lysed by frequent vortexing for 10 min, keeping the cells on ice between vortexes. The cell lysate was then sonicated on ice for 24 s, with 4 s of rest between each 4 s sonication interval, using a full probe on a Fisher Scientific Model 550 Sonic Dismembrator. The median chromatin fragmentation size was verified by gel electrophoresis to be approximately 6 kilobases.

Hybridization and Capture

The lysate was diluted with nuclease free water so that the final component concentrations for hybridization are as follows: 375 mM LiCl, 50 mM Tris, 1% LiDS, 1% Triton X-100, 10 mM RVC, 10 mM DTT, 100 U/mL RNasin Plus, 1X Halt Protease Inhibitors. In addition, all the lncRNA capture oligonucleotides and the scrambled oligonucleotide (Table S2) were added to the diluted lysate and the sample was incubated while nutating at 37 °C for 3 h. The predetermined volume of streptavidin coated magnetic Speedbeads (Thermo Fisher Scientific, 09981140) was washed 3 times with and resuspended in three volumes of wash buffer (375 mM LiCl, 50 mM Tris, 0.2% LiDS, 0.2% Triton X-100) prior to addition to each

hybridization mixture. The bead capture mixture was then nutated at 37 °C for 1 h. Following incubation, the beads were collected to the side of the tube using a magnet, and the remaining lysate was removed. A volume of prewarmed wash buffer that was 5× the volume of beads used for capture in the sample was used to resuspend the beads, and the beads were washed at 37 °C for 15 min. The beads were then washed with a 5× volume of release buffer (375 mM LiCL, 50 mM Tris, 0.1% LiDS, 0.1% Triton X-100) at room temperature for 5 min.

Release from Beads

The beads were resuspended in release buffer (a volume 3× that of the beads used for capture), and the release oligonucleotides for one lncRNA target were added at an amount 100× that of their complementary respective capture oligonucleotides (Table S2). The mixture was nutated at room temperature for 30 min followed by magnetic separation of the beads from the supernatant containing the released RNA−protein complexes. This supernatant was transferred to a new tube and divided into aliquots for RT-qPCR and mass spectrometric protein analysis. This process was repeated for the release of each lncRNA target so that each was released into a separate solution.

RT-qPCR

RT-qPCR was used to estimate transcript copy numbers, capture efficiencies and specificities, release efficiencies, and fold enrichments. In general, 40 μL of cell lysate and 2% of captured and released target RNA−protein complexes were used to make these measurements and calculations. Samples were incubated at 37 °C with 1 mg/mL proteinase K, 4 mM $CaCl_2$, and 0.2% LiDS. The RNA was extracted with TriReagent (Sigma, T9424) per manufacturer's protocol and was then precipitated in 75% ethanol at −20 °C overnight. The tube was then centrifuged at 20,800g to pellet the RNA, and the RNA was then washed with 75% ethanol. The RNA pellet was resuspended in 15 μL of nuclease free water, and 1 μg (up to 10 μL) was used for reverse transcription (High Capacity cDNA Reverse Transcription Kit, Applied Biosystems) per the manufacturer's protocol. The resulting cDNA sample was analyzed with sequence-specific qPCR assays (Tables S3 and S4). For quantitation, standard curves were prepared using genomic DNA (G3041, Promega) (Table S5).

eFASP

The protocol for protein preparation was adapted as follows from that described by Erde et al.[36] Each release sample was brought to 0.1% deoxycholic acid (DCA) and 8 M urea. The sample was passed through the filter in 500 μL increments by centrifugation for 10 min at 14,000g, and the eluant was discarded. Exchange buffer (400 μL; 8 M urea, 0.1% DCA) was added to the filter, and the tube was centrifuged at 14,000g for 10 min. This was repeated for a total of three exchanges. Reducing buffer (200 μL; 8 M urea, 20 mM DTT) was added to the filter, and the sample was incubated for 30 min at room temperature followed by centrifugation at 14,000g for 10 min. Alkylation buffer (200 μL; 8 M urea, 50 mM iodoacetamide, 50 mM ammonium bicarbonate) was then added to the sample, followed by incubation for 30 min at room temperature in the dark, and centrifugation at 14,000g for 10 min. Finally, the sample was exchanged with three aliquots of 400 μL of digestion buffer (1

M urea, 50 mM ammonium bicarbonate, 0.1% DCA) and resuspended in a final volume of 100 μL of digestion buffer. Trypsin was added to the sample, the filter was transferred to a fresh, passivated collection tube, and the cap was sealed with parafilm followed by incubation overnight at 37 °C for protein digestion. The filter-collection tube was centrifuged for 10 min at 14,000g. Fifty microliters of 50 mM ammonium bicarbonate was added to the filter and centrifuged at 14,000g for 10 min. This step was repeated once to ensure the collection of the entire peptide sample. The 200 μL peptide sample was then brought to 0.5% TFA followed by addition of 200 μL of ethyl acetate. The sample was vortexed for 1 min then centrifuged at 15,700g for 2 min. The top layer was aspirated and discarded, and extraction with ethyl acetate was repeated two times. The aqueous layer was then dried using a Savant SVC-100H SpeedVac Concentrator, and the sample was resuspended in 150 μL 0.1% TFA. For removal of salts from the sample, a C18 solid-phase extraction pipet tip (OMIX C18, 100uL, Agilent Technologies) was first conditioned with 70% ACN, 0.1% TFA, and then equilibrated with 0.1% TFA. The peptide sample was then loaded onto the C18 solid phase by repeated passing of the 150 μL sample over the cartridge. The OMIX pipet tip was then rinsed with 0.1% TFA 10 times followed by peptide elution in 150 μL of 70% ACN, 0.1% TFA. The samples were then dried using the SpeedVac Concentrator and reconstituted in 27 μL of 95:5 H_2O/ACN, 0.2% formic acid and analyzed in triplicate (9 μL for each injection) as described below.

Mass Spectrometry of Peptides

The samples were analyzed using an HPLC-ESI-MS/MS system consisting of a high performance liquid chromatograph (nanoAcquity, Waters) set in line with an electrospray ionization (ESI) Orbitrap mass spectrometer (QE HF, ThermoFisher Scientific). A 100 μm id × 365 μm od fused silica capillary microcolumn packed with 20 cm of 1.7 μm diameter, 130 Å pore size, C18 beads (Waters BEH), and an emitter tip pulled to approximately 1 μm using a laser puller (Sutter Instruments) was used for HPLC separation of peptides. Peptides were loaded on-column with 2% acetonitrile in 0.1% formic acid at a flow rate of 400 nL/min for 30 min. Peptides were then eluted at a flow-rate of 400 nL/min over 120 min with a gradient from 5% to 35% acetonitrile, in 0.1% formic acid. Full-mass profile scans (375−1500 *m/z*) were performed in the FT orbitrap at a resolution of 120,000 followed by 20 MS/MS HCD scans of the 20 highest intensity parent ions at 30% relative collision energy and 15,000 resolution with a mass range starting at 100 *m/z*. Dynamic exclusion was enabled with a repeat count of one over a duration of 30 s.

Mass Spectrometry Data Processing

The Orbitrap raw files were analyzed using MaxQuant (version 1.5.3.30)[37] and searched with Andromeda using the Uniprot canonical human protein database supplemented with common contaminants. Samples were searched allowing for a fragment ion mass tolerance of 20 ppm and cysteine carbamidomethylation (static) and methionine oxidation (variable). A 1% false discovery rate for both peptides and proteins was applied. Up to two missed cleavages per peptide were allowed, and at least four peptides were required for protein identification and quantitation. Protein quantitation was achieved using the resulting label-free quantitation (LFQ) intensities from the MaxQuant output. A procedure for calculating fold enrichment of proteins was modified from West et al.[5] For proteins quantified in the

lncRNA capture samples but without quantifiable intensities in the input or scrambled oligonucleotide capture samples, the zero value was replaced with the intensity value of the lowest expressed protein in that input or scrambled capture sample to assist in calculating conservative fold-enrichments for that protein in the capture sample. A protein was considered to be enriched in a capture sample if its intensity was at least 5-fold greater for MALAT1 and NEAT1 or at least 8-fold greater for NORAD than that in the scrambled capture sample or input sample. Only proteins meeting this criterion in at least two of the three biological replicates were considered enriched for a particular lncRNA capture type. Proteins that were related to keratin or biotin were eliminated due to being likely contaminants.

Interactome Data Analysis

Gene Ontology (GO) Enrichment Analysis was performed using the publicly available Gene Ontology Consortium (http://www.geneontology.org/), the Protein Analysis Through Evolutionary Relationships (PANTHER) Classifica-tion System[38] and STRING.[39] These databases describe proteins using GO annotations related to their biological processes, molecular functions, and cellular components. The interactome of each lncRNA, as determined above, was evaluated using these software for annotations statistically overrepresented in the interactome relative to a random selection of the same number of proteins from the human proteome. The resulting overrepresented annotations are listed in Tables 2 and S7.

Hierarchical Clustering and Heat Map Generation

The proteins determined to be enriched in the NORAD, NEAT1, and/or MALAT1 lncRNA captures were included in the hierarchical clustering analysis. A matrix was constructed of $\log_2$-transformed protein intensities for each protein in the nine samples (three biological replicates of each of the three lncRNA targets). Missing values were imputed, using Perseus software,[40] with random numbers near the limit of detection so that the mean of the imputed values was 1.5 standard deviations from the mean of all protein intensities in the sample. The resulting matrix was then imported into Cluster 3.0,[41] the intensities of each protein for all nine samples were adjusted to center around the mean intensity of that protein, and then the uncentered-correlation similarity metric and a centroid linkage clustering method were applied. The hierarchically clustered results were then imported into TreeView[42] for visualization and for construction of the heatmaps shown in Figure 4.

RNA-Binding Protein Immunoprecipitation Validation

Immunoprecipitaion of IGF2BP2 and NAP1L4 from cross-linked PC3 cell lysate was performed using an EZ-Magna Nuclear RIP (Cross-linked) Kit (EMD Millipore, 17–10521). Antibodies specific to each target protein (Abcam; Anti-IGF2BP2, AB128175, and Anti-NAP1L4, AB21631) and negative control antibodies, normal rabbit IgG (EMD Millipore, 12–370) and normal mouse IgG (included in the EZ-Magna Kit), were used. PC3 cells were prepared as they were for the HyPR-MS experiments, and the immunoprecipitation, elution, and RNA purification were performed as indicated by the EZ-Magna Kit protocol. RT-qPCR was performed as described above, and the ΔΔCt calculation method was used for

determination of enrichment of MALAT1, NEAT1, and NORAD lncRNAs in each immunoprecipitation. Fold enrichments >2 were considered significant.

Validation of lncRNA-Binding Proteins Using eCLIP Data

The binding proteins for NORAD, NEAT1, and MALAT1 were validated using eCLIP[43] immunoprecipitation data from the ENCODE project in either K562 or HepG2.[44] The RNAs captured in these pulldowns were quantified using the integrated software program Spritz (unpublished, https://github.com/smith-chem-wisc/Spritz) using the "quantify" command, which quantifies genes and transcripts using RSEM.[45] First, Spritz prepares reference indices for the STAR[46] alignment software and for RSEM quantification software using the human reference genome (GRCh38.81). Then, it downloads FASTQ files from Sequence Read Archive accessions if available and calculates gene abundances using RSEM. The parameters used for the "rsem-calculate-expression" program within RSEM were "--time --calc-ci --star --num-threads 24 --output-genome-bam --paired-end". This analysis was performed on an Intel Xeon CPU with 24 cores and 128 GB of RAM.

RESULTS

HyPR-MS Provides Efficient and Simultaneous Isolation of Multiple RNA Targets

While sharing the basic structure of previously developed RNA-centric strategies, HyPR-MS addresses their limitations through key differences in (a) cross-linking and solubilization, (b) capture oligonucleotide design, and (c) specific, multiplexed release of RNA−protein complexes from the beads (Tables 1 and S1). These modifications reduce the cost and labor of elucidating specific RNA-associated proteomes by improving efficiency and permitting multiplexing.

Many intertwined factors contribute to achieving the desired efficiency of purifying specific RNAs from a complex lysate while preserving *in vivo* RNA−protein interactions. Not least among them is the balance of cross-linking conditions and lysate solubilization. While *in vivo* formaldehyde cross-linking preserves a snapshot of RNA−protein and protein−protein interactions, the parameters used for cross-linking can have great effect on the number of proteins cross-linked to the RNA of interest. This in turn affects the solubility of the cross-linked molecules as well as the accessibility of the RNA to hybridization with the capture oligonucleotide. Furthermore, the extent of cross-linking influences the optimal sonication parameters that allow for the solubilization of chromatin and other large biomolecules while still maintaining low fragmentation of the target RNA−protein complexes. These parameters can ultimately affect the overall cost of the experiment; an increase in efficiency reduces the total amount of materials needed to capture sufficient protein for MS analysis. Another major contributor to the cost effectiveness of HyPR-MS is the design and implementation of the capture oligonucleotides. Methods such as RNase H assays or SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) have been shown to be effective experimental techniques for determining accessible RNA regions.[5,47] HyPR-MS, however, avoids the time and cost of such methods and employs instead a publicly available program, M-fold,[35] to predict the secondary structure of the target RNA and identify the regions most likely available for hybridization. Employing theoretical assessments of RNA secondary

structure is likely less reliable than the aforementioned experimentally determined hybridization regions. Designing several oligonucleotides for each target and determining empirically the capture efficiencies for each, however, has been shown here to be an effective strategy for capture oligonucleotide design. Compared to RNA tiling strategies[4,48] that can require more than 40 capture oligonucleotides to capture a single RNA target, this strategy provides sufficient efficiency of capture so that only 2–3 capture oligonucleotides are needed (Figure S1A), thus reducing costs and allowing for the purification of specific isoforms of an RNA target. Finally, a significant design attribute of HyPR-MS is the use of a toehold-mediated release strategy,[49,50] which allows for the capture and purification of multiple RNA species from a single cell lysate. This provides not only the experimental advantage of maintaining the same background for all RNA species captured but also greatly reduces cost and time requirements. This strategy is highlighted in Figure 1B and Table S1. Each of these steps was designed to reduce the time and financial barriers of RNA-centric strategies for identification of RNA–protein interactomes; the capture oligonucleotide design limitations, high cost of experiments, and lengthy experiment time requirements. Details on reasoning for each of these parameters are provided in Table S1 and summarized in Table 1.

HyPR-MS Provides Efficient and Specific Purification of Multiple lncRNAs

The HyPR-MS strategy was utilized to identify the protein interactomes of MALAT1, NEAT1, and NORAD in three biological replicates of PC3 cells. Each capture experiment used 1×10^8 cells and two to three capture oligonucleotides (Figure 2A and Table S2) to meet the minimum capture efficiency threshold. This threshold was calculated so that, assuming one protein molecule per lncRNA molecule, a sufficient number of copies of a given protein would be present for mass spectrometric detection (Figure S1). For each replicate, small aliquots of the purified capture samples as well as lysate and post-release bead samples were analyzed using RT-qPCR to determine capture and release efficiencies, capture specificity, and fold enrichment of each lncRNA target.

Capture Efficiency.—The capture efficiency of each lncRNA target was measured using three to four target-specific RT-qPCR assays that amplify regions dispersed along the length of the RNA target (Figure 2A and Table S3). The average percent capture efficiencies for MALAT1, NEAT1, and NORAD from the large-scale proteomics experiments were approximately 8, 20, and 28, respectively (Figure 2B). The percent capture efficiencies for all qPCR-measured regions ranged from 1 to 12 for MALAT1, 3 to 40 for NEAT1, and 17 to 35 for NORAD (Figure S2A). All of these capture efficiencies, with the exception of the 5′-NEAT1 region, are several-fold higher than their respective thresholds. This result suggests that even proteins with low-occupancy on the lncRNA could be detected by MS. Similarly, this suggests that HyPR-MS can be applied to study less abundant RNAs while still using an experimentally feasible number of cells. However, it should also be noted that the capture efficiencies measured using the different RT-qPCR assays and for the different lncRNAs are highly variable. This is attributable to several causes, as described below, though the precise contribution of each cause was not determined.

First, the variable capture efficiency observed is in part an artifact of experimental design. To achieve simultaneous capture of multiple lncRNAs while also using multiple capture

oligonucleotides for each RNA target, hybridization parameters (i.e., salt concentrations and hybridization temperature) were necessarily suboptimal for some of the designed capture oligonucleotides. This will in turn negatively impact capture efficiency for the affected capture oligonucleotides. Further-more, the amount of capture oligonucleotides used for capture was adjusted to minimize the cost of each experiment while still obtaining sufficient protein for mass spectrometric analysis of the different lncRNA–protein targets. While increasing the amount of capture oligonucleotides increases the capture efficiency, it also increases the amount of streptavidin-coated beads required to complete the capture, thus increasing the experimental cost. This increase in capture oligonucleotide amounts oftentimes also has a negative impact on capture specificity. For this reason, the concentration of each capture oligonucleotide added to the capture experiment was determined to ensure that the capture efficiency (CE) would surpass the calculated CE threshold, but not necessarily to maximize the CE (see Table S2 for concentrations used).

The capture efficiency measurement for the different RT-qPCR assays within the same lncRNA target is also variable. This is likely, in part, to be caused by the lysate sonication step conducted prior to hybridization. Lysate sonication is used to solubilize the cell contents; however, it also could fragment long RNA transcripts. The sonication parameters for HyPR-MS were established to minimize this fragmentation while still achieving solubility. Residual RNase activity in the lysate may also contribute to RNA fragmentation. The use of multiple capture oligonucleotides to hybridize to various regions of the RNA allows for the capture of the resulting fragments, though potentially at different efficiencies.

Capture efficiency for a given capture oligonucleotide will also be affected by inherent differences in secondary structure or protein occupancy at each hybridization site. The local secondary structures of a given RNA could also affect the efficiency of reverse transcription and ultimately qPCR amplification.

Finally, the RT-qPCR measurements made for calculating capture efficiency are affected by each sample composition. For example, the aliquot of lysate used to determine the amount of each lncRNA transcript present in the lysate and available for capture (i.e., the denominator in the capture efficiency calculation) is a complex sample containing all the RNA transcripts present in the cell, with relatively low concentrations of the lncRNA targets. In contrast, the capture sample (i.e., the numerator in the capture efficiency calculation) is a simplified sample consisting of primarily the target lncRNA and very low levels of the various other cellular RNAs. Both the complexity of the sample and the concentration of the RNA in the sample affect the efficiency of reverse transcription. Furthermore, this effect can be different for different regions of the RNA. This complication is likely a contributor to the variable capture efficiency calculations for the different RT-qPCR assays within a given lncRNA target. All of the above-mentioned factors should be considered while assessing the capture efficiencies presented here.

Ultimately, HyPR-MS is a technology designed for cost- and time-efficient discovery of novel RNA interactors. Many complex factors affect the figures-of-merit for RNA-centric technologies, including HyPR-MS. Thus, future implementers of HyPR-MS may need to

adjust parameters to meet their experimental needs. The protocol described here provides a baseline for establishing such parameters.

Capture Specificity.—The capture specificity of HyPR-MS is a measure of the capacity to capture the target lncRNA while avoiding the capture of nontarget RNA and DNA molecules. Optimization of this parameter is crucial to the accurate identification of an RNA's interacting proteome. Capture specificity depends not only on the unique complementarity of the capture oligonucleotide (CO) to the target RNA but also on various other HyPR-MS conditions. These conditions were optimized in concert to provide selectivity of lncRNA capture relative to nontarget RNA and DNA. Capture specificity for each target was calculated using RT-qPCR to compare the amount of the target RNA in its corresponding capture sample to that in the capture sample of a different RNA target. The capture oligonucleotide specific to each lncRNA target captures at least 10-fold more of that lncRNA than does the CO specific to a different lncRNA (Figure 2C). Capture specificity was also demonstrated by conducting a capture experiment using a scrambled capture oligonucleotide. The scrambled CO was designed so that it is not complementary to any sequence in the PC3 transcriptome or genome and has a melting temperature similar to those of the targeted COs (Table S2). The amount of each lncRNA target in its corresponding capture sample relative to that in the scrambled capture sample is at least 10-fold higher for MALAT1, NEAT1, and NORAD (Figure 2D). These results confirm the specificity of HyPR-MS and the use of the scrambled oligonucleotide capture sample as a negative control for evaluation of the lncRNA protein interactomes. As a broad measure of the quality and power of the HyPR-MS method, fold enrichment for each target is calculated compared to GAPDH mRNA (an abundant housekeeping gene). To do this, the ratio of target-to-GAPDH RNA in the captured sample over target-to-GAPDH in lysate is found using RT-qPCR. The fold enrichment of each target is greater than 100-fold relative to GAPDH for all three target lncRNAs (Figure 2E). Finally, we ensured that the lncRNA capture samples were devoid of any DNA with the same sequence. Data in Figure S2B show that all capture samples contain insignificant amounts of the analogous DNA sequences, and thus, the levels of any DNA-associated proteins are inconsequential. Taken together, these data demonstrate the specificity of each capture, ensuring the uniqueness of the associated interacting proteomes, and meet or surpass the metrics of other RNA capture strategies.[5]

Release Efficiency.—Capture efficiency and capture specificity rely in part on the specific and complete release of each individual lncRNA target from its hybrid interaction with the capture oligonucleotide–bead complex. This is accomplished by using a sequence-specific toehold release oligonucleotide to displace the capture oligonucleotide, thus selectively releasing only the target RNA into solution. The release efficiency is measured using RT-qPCR analysis of the RNA released from the beads and of the RNA remaining on the beads following release of all targets. The percent release efficiencies from each capture oligonucleotide range from 40 to 80% (Figure 2F). The effect of release order on capture specificity and efficiency was previously evaluated and shown to not substantially impact these measurements.[49]

HyPR-MS Reveals the Protein Interactomes of MALAT1, NEAT1 and NORAD lncRNAs

The proteins isolated from the capture samples, as well as an aliquot of lysate used as the input material for each capture, were analyzed by mass spectrometry. The resulting spectra for three biological replicates were searched and quantified using the MaxQuant platform.[51] Statistical analysis as described in the Experimental Section determined the proteins enriched in each capture sample relative to the scrambled oligonucleotide control and input samples. This analysis yielded 127, 94, and 415 proteins in the MALAT1, NEAT1, and NORAD interactomes, respectively (Table S6). The interactome proteins for each lncRNA target were evaluated using STRING,[39] open-source software for identifying protein−protein interactions and calculating GO (gene ontology) term enrichment, and the PANTHER (Protein Analysis Through Evolutionary Relationships) classification system.[38] The enriched GO terms for each of the three interactomes are summarized in Tables 2 and S7.

Each lncRNA-Protein Interactome is Enriched for Distinct Gene Ontology Terms

Each lncRNA interactome is highly enriched for RNA-binding proteins (p-values between 2×10^{-19} and 6×10^{-49}) giving evidence that HyPR-MS identifies proteins interacting with RNAs (Figure 3 and Table 2). Also significant is that the enriched proteins in each interactome are known to have cellular localizations associated with the known locations of the lncRNAs themselves. Both MALAT1 and NEAT1 lncRNAs have been shown to localize to the nucleus,[33,52] and here, both have interactomes significantly enriched for proteins with nuclear localization but not for proteins with cytoplasmic localization (Figure 3 and Table 2). Conversely, NORAD has been shown to localize primarily to the cytoplasm, and the NORAD interactome shown here is significantly enriched for cytoplasmic proteins; however, they are also enriched for proteins with nuclear localization (Figure 3 and Table 2). These nuclear proteins could be indicative of a novel nuclear function for NORAD or could represent general RNA processing functions required by all RNAs.

The MALAT1 interactome is enriched for components of the spliceosomal complex, while both MALAT1 and NEAT1 are enriched for proteins associated with nuclear bodies called nuclear speckles and paraspeckles (Table 2). These nuclear bodies are not sites of active transcription but are thought to be locations for the assembly and storage of splicing machinery.[53] Since NEAT1 and MALAT1 have been shown to have a role in the regulation of alternative splicing in different cell lines[13,17,33] and NEAT1 is a primary structural component of paraspeckles, these enrichments provide further support for the ability of HyPR-MS to correctly reveal specific RNA-interacting proteins.

The MALAT1 and NORAD interactomes are both enriched for proteins associated with nuclear export of RNA, although the proteins that fall into this category differ for each lncRNA. For example, the NORAD interactome is enriched for several proteins of the nuclear pore complex (NPC), which is the conduit for RNA export through the nuclear membrane. This is not surprising as NORAD is localized to the cytoplasm and therefore must exit the nucleus. Interestingly, XPO1 (or CRM1) was also enriched in the NORAD capture samples. This protein is required for the nucleocytoplasmic export of only a subset of proteins and RNAs, and its overexpression has been implicated in poor prognosis of many cancer types.[54] In contrast, the annotated nuclear export proteins enriched in the MALAT1

interactome include proteins of the TREX complex and its associated proteins including SRSFs. The TREX complex links mRNA processing steps from transcription, to splicing, to nuclear export, and its depletion in cells can cause genome instability leading to cancer.[55] These TREX and SRSF proteins may be functioning to cotranscriptionally splice and export MALAT1 into the cytoplasm or could be sequestered with MALAT1 in the nucleus in a nuclear body.

All three lncRNA interactomes in the PC3 cell line are enriched for extracellular vesicle proteins (Table 2). Extracellular vesicles are involved in cell-to-cell communication, and it has been shown that their lncRNA content can be altered by the tumor microenvironment.[56] Both MALAT1 and NEAT1 have previously been discovered in extracellular vesicles produced by PC3 cells.[57] In addition, MALAT1 is known to be enriched in extracellular vesicles originating from cervical carcinoma and breast cancer cells,[58] and NEAT1 is associated with extracellular vesicles in human liver cancer (HepG2) and cholangiocarcinoma (Mz-ChA-1) cell lines.[59] These data are in accordance with the enrichment of extracellular vesicle proteins in the MALAT1 and NEAT1 interactomes. Interestingly, the NORAD interactome is more enriched for extracellular vesicles (p-value 2×10^{-19}) than are the interactomes for MALAT1 or NEAT1 (p-values: 8×10^{-6} and 3×10^{-7}, respectively). To our knowledge, NORAD has not been detected in extracellular vesicles produced by PC3 cells, and this finding suggests that it could be a significant component of such vesicles.

Heatmap Generation of Hierarchically Clustered Interactomes Shows Protein Enrichment Differences in lncRNA Captures

A clustering algorithm and heatmap visualization strategy was employed to interpret the lncRNA protein interactomes. A matrix of $\log_2$-transformed, mean-centered protein intensities for each lncRNA biological replicate was compiled for all proteins enriched in at least one species of lncRNA capture. A hierarchical clustering algorithm[41] grouped the proteins based on the similarities between each protein's intensity profile across the three lncRNA capture species, and TreeView software[42] was used to visualize these similarities and differences (see Experimental Section). Figure 4A is the resulting heatmap; the nine columns represent the biological replicates for the three lncRNA captures, and each row shows how the intensity of a protein compares to the mean intensity for all capture samples and biological replicates; a brighter red or yellow data-point indicates a greater difference from the mean intensity value. This strategy allows us to not only use the scrambled capture and input samples to determine which proteins are enriched in the lncRNA captures but also provides an additional layer of analysis that compares the lncRNA interactomes to one another.

HyPR-MS Identifies Known and Novel lncRNA Interactors

The clustering algorithm subdivides the heatmap in Figure 4A into sections consisting of the proteins that are elevated in an individual lncRNA capture sample relative to the other capture samples. For example, the heatmap shows the clustering of over 40 proteins that are more abundant in the MALAT1 and NEAT1 captures than in the NORAD captures, over 70 proteins and 250 proteins that are more abundant in the MALAT1 and NORAD captures,

respectively, relative to the opposing lncRNA captures, and over 50 proteins that are approximately equally abundant in all three lncRNA interactomes. Examples of annotated clusters demonstrating these trends are shown in Figure 4B–E.

NEAT1 is a core component of paraspeckles; satisfyingly, the GO term enrichment analysis for NEAT1 shows enrichment for paraspeckle protein components. Figure 4B shows a cluster of proteins with abundances elevated in NEAT1 and MALAT1 captures that includes six paraspeckle components. This cluster also includes seven additional proteins that share a similar intensity profile with the paraspeckle components. STRING analysis of the proteins in this cluster shows that nine of the 13 proteins in the cluster are known to interact (Figure 4F). The remaining proteins in the cluster are transcriptional regulators or have known direct interaction with transcriptional regulators thereby making them potential candidates for further analysis for their role in lncRNA mechanisms.

Figure 4C shows clusters containing proteins with greater abundance in the MALAT1 capture relative to the NEAT1 and NORAD captures, and Figure 4G shows the known interactions of the proteins in these clusters according to STRING analysis. Twenty proteins enriched in the MALAT1 capture samples have been previously shown to interact with MALAT1[5,13] with 15 of those grouped into the clusters shown in Figure 4C. Also among these clusters and enriched in the MALAT1 interactome are many nuclear speckle and splicing factors, including SR-splicing factors that contain RS-domains rich in serine (S) and arginine (R) residues. This is consistent with one proposed mode of action for MALAT1 that suggests MALAT1 regulates alternative splicing of endogenous RNAs by localizing SR splicing factors to nuclear speckles.[33] Notably, SR protein function is known to be affected by their phosphorylation states;[33] this makes the proteins with kinase or phosphorylation regulatory activities (CDK11, PRPF4B, GRB2, TARDBP) clustered among the MALAT1 interactors intriguing candidates for further functional analysis.

The clusters in Figure 4C also include many proteins involved in transcriptional regulation. Notable among these are histone modifiers, including components of the HUSH complex (PPHLN1 and FAM208A) known for gene silencing through methylation of H3K9me3, CHTOP, an associate of the methylosome complex that methylates H4R3, and HIST1H1C, which trimethylates H3K27 (Figure 4C,G). These findings, in addition to other transcriptional regulators identified by HyPR-MS (Figure 4G), are congruent with previous works that demonstrated that lncRNAs in general,[60] and MALAT1 specifically,[5] localize to a subset of genomic sites and recruit proteins that affect gene expression.

The functions and modes of action of nuclear lncRNAs such as MALAT1 contrast with those of cytoplasmic lncRNAs like NORAD. Cytoplasmic lncRNAs are less well understood; however, they are thought to affect the activity or expression of the RNAs or proteins with which they interact. Two recent studies specifically investigated the interaction of NORAD, a lncRNA whose expression is induced upon DNA damage,[24] with PUM1 and PUM2, proteins that bind a specific subset of mRNAs and promote their degradation.[22,24] Both studies found that the abundant NORAD lncRNA acts as a molecular decoy capable of binding many molecules of these PUM proteins. These interactions in turn prevent the PUM proteins from negatively regulating the expression of their usual RNA targets. Satisfyingly,

HyPR-MS identified PUM1 among the >400 NORAD interactors. Lee et al. also conducted NORAD RNA knockdown experiments to determine the effects of decreased levels of NORAD on gene expression. While the genes of 193 PUM protein targets were affected, the expression of over 1000 other genes were also affected,[24] suggesting that NORAD has additional modes of action aside from PUM protein sequestration.

Because PUM1 was identified by HyPR-MS and binds to RNAs to regulate the expression of specific genes, we searched our data for additional regulators of gene expression through RNA binding. Figure 4D shows three clusters containing proteins with elevated abundances in the NORAD captures and highlights the transcription and translation regulators among these clusters. Among these are proteins with various known modes of action on specific subsets of RNA. For example, protein EIF3C is not only required for initiation of translation, it has also been found to bind to certain RNA stem-loop configurations to regulate specifically RNAs involved in cell growth control.[61] YTHDF3 binds to N^6-methyladenosine-modified RNA and has been found to work with YTHDF1 to promote translation or with YTHDF2 to promote RNA decay.[62]

The aforementioned NORAD-RNA knockdown study conducted by Lee et al. also showed that the absence of NORAD resulted in chromosomal instability and abnormal chromosome numbers in the cells.[24] Interestingly, GO term analysis of the NORAD HyPR-MS results showed enrichment of proteins associated with the response to DNA damage, the mitotic cell cycle, and the minichromosome maintenance complex (MCM) (Table 2). Figure 4D highlights many proteins identified to function in response to DNA damage correlating with the previous finding that NORAD expression itself is induced upon DNA damage. Among these are five components of the MCM, which is involved in the regulation of DNA replication, ensuring its occurrence only once per cell.[24] NORAD has previously been implicated in the control of DNA replication through interaction with Pumilio proteins; the observation of its interaction with the MCM lends additional support to its role in this process. Also striking are the numerous proteins associated with mitosis and chromosomal regulation identified in the NORAD interactome (Figure 4D).

Validation of lncRNA–Protein Interactomes

As is described above, HyPR-MS has revealed specific proteins whose functions correlate with previously established functions of their associated lncRNAs. The validity of many of the lncRNA–protein interactions identified by HyPR-MS was further established using three strategies. First, the interactomes presented here for MALAT1 and NEAT1 were compared to previous studies interrogating these same lncRNAs. Second, the HyPR-MS interactomes were cross-referenced with publicly available eCLIP data to determine if this orthogonal technique could confirm the interactions. Third, RIP-qPCR experiments were performed to confirm the lncRNAs interacting with IGF2BP2 and NAP1L4 in PC3 cells.

RNA-Centric Studies Validate lncRNA Protein Inter-actors Identified by HyPR-MS.—Various methods have been used in previous studies for interrogating the protein interactomes of MALAT1, NEAT1, and NORAD. The results from HyPR-MS were cross-referenced with such studies revealing many proteins that have been previously determined

to interact with each lncRNA. CHART-MS, an RNA-centric *in vivo* approach, was used to study MALAT1 and NEAT1 protein interactors in MCF-7 cells.[5] The resulting protein lists from CHART-MS and HyPR-MS have 27 and 17 proteins in common for MALAT1 and NEAT1, respectively (Table S6). RNA-centric, *in vitro* methods have also been used to identify potential interactors of MALAT1 and NORAD lncRNAs.[22,24,63] These methods use synthetic RNA molecules to isolate binding proteins from cell lysates. Cross-reference of HyPR-MS data with the data from Chen et al. and Lee et al. validated 22 proteins and 18 proteins as interactors with MALAT1 and NORAD, respectively (Table S6). Though the *in vitro* experiments referenced here did not evaluate in vivo interactions as HyPR-MS does, these shared protein interactors do confirm that the proteins are capable of binding to the MALAT1 or NORAD transcripts.

Protein-Centric Techniques Validate lncRNA Interactors Identified by HyPR-MS.—Data acquired using eCLIP (enhanced cross-linking and immunoprecipitation), an orthogonal technique to HyPR-MS, were used to validate lncRNA protein interactors. The ENCODE project provides publicly available eCLIP data[44] from K562 and/or HepG2 cells for 21, 11, and 22 of the proteins identified by HyPR-MS to be interacting with NORAD, NEAT1, and MALAT1, respectively. Each eCLIP experiment consisted of two biological replicates for the target protein and one control experiment. The RNAs sequenced in these experiments were quantified as described in the Experimental Section. The results of this experiment were considered validating for a particular RNA−protein interaction if the 95% credibility interval (CI) for a lncRNA of interest was higher than the 95% CI for the control in both biological replicates. Of all 54 unique lncRNA−protein interactions, 49 passed validation (91%) in at least one cell line (Figure 5 and Table S9).

Novel lncRNA Interactors IGF2BP2 and NAP1L4 are Confirmed by RIP-qPCR.—IGF2BP2 was enriched in all three of the lncRNA captures presented here with a higher relative enrichment in the NORAD samples (Figure 4E). IGF2BP2 binds to and recruits target RNAs to cytoplasmic mRNPs, thus affecting the rate of translation and decay of the RNAs.[64] We selected IGF2BP2 as a target for validation using RNA-immunoprecipitation followed by qPCR analysis (RIP-qPCR). Figure 4I shows that, following immunoprecipitation of IGF2BP2, all three lncRNAs were enriched relative to the negative controls, immunoprecipitations using an IgG antibody. Since MALAT1 and NEAT1 are nuclear localized and NORAD is cytoplasmically localized, the modes of action for their interactions with IGF2BP2 could be different.

NAP1L4 was found by HyPR-MS to be only enriched in NORAD capture samples. NAP1L4 is described as a histone chaperone for its ability to bind core and linker histones and transfer them from the cytoplasm to the nucleus where they are assembled into nucleosomes.[65] NAP1L4 localization and function are regulated by its phosphorylation state; the phosphorylated protein localizes in the cytoplasm in complex with histones during the G1 phase of the cell cycle, while its dephosphorylated form initiates its transport into the nucleus at the start of S phase. At this phase, NAP1L4 is involved in chromatin assembly, a process coupled to DNA replication and DNA repair.[66] This is particularly interesting in relation to the finding of chromosomal abnormalities as a result of NORAD knockdown in

cells.[24] We validated the interaction of NAP1L4 with NORAD using RIP-qPCR. Figure 4H shows the enrichment of NORAD transcripts upon immunoprecipitation of NAP1L4 compared to the negative control, but not the enrichment of MALAT1 or NEAT1. Considering NORAD is already described as a molecular decoy for the PUM1 and PUM2 proteins,[24] it is possible that it may have a similar role with regards to NAP1L4 or be involved in the regulation of the phosphorylation state of NAP1L4 as MALAT1 is for SR proteins. Future studies are necessary to test these hypotheses.

Limitations.—HyPR-MS is the only current strategy that can analyze the interacting proteomes of multiple RNA targets from the same cell culture. However, the extent to which the multiplex capabilities of HyPR-MS can be further expanded past the 3-fold multiplex capability shown in this study is not yet clear. The use of stable-isotope labeling of peptide samples prior to mass spectrometric analysis could also improve upon this technology. Using labeling strategies such as iTRAQ[67] or TMT[68] for the capture samples prior to analysis would likely provide superior relative quantitation of the proteins in each sample. HyPR-MS has been previously used to isolate HIV mRNA from infected Jurkat cells;[27] however, it is yet to be determined if other species of RNAs such as rRNA or tRNA are accessible for hybridization purification. Additionally, the main challenge for the widespread adoption of *in vivo* RNA centric methods is the study of less abundant RNAs while still using a practical number of cells to meet mass spectrometer sensitivity constraints for complex biological mixtures. Our results suggest that HyPR-MS may be applicable to low copy number RNAs, but further studies are needed to determine its lower detection limit.

DISCUSSION

HyPR-MS was designed in response to several unaddressed challenges that have limited the widespread application of *in vivo* RNA-centric methods for interactome discovery. While maintaining the basic structure of previous strategies, we reworked each step to reduce the time and cost of experiments and to empower multiplexing for concurrent analysis of several RNA species.

Compared with other strategies, HyPR-MS exhibits expedient and reliable oligonucleotide design via utilization of the publicly available M-fold software. The use of a multioligonucleotide capture strategy provides a high capture efficiency along the full length of each RNA target while still providing the flexibility to investigate the specific RNA variants. Less stringent cross-linking treatments and solubilization steps reduce target RNA fragmentation and improve capture efficiency. Finally, the toehold-mediated capture and release strategy permits the study of several targets within the same cell culture preparation, thereby reducing the cost and time requirements and minimizing background and sample variability.

We utilized HyPR-MS to investigate the protein interactomes of MALAT1, NEAT1, and NORAD in the PC3 cell line. The functions and many interacting proteins of MALAT1 and NEAT1 have been previously studied in other cell lines, and these previous studies serve in part to provide markers for establishing the efficacy of the HyPR-MS method. In addition, analysis of MALAT1 and NEAT1 protein interactomes in PC3 cells presented here provides

novel discoveries for further understanding of their function. The identification of the *in vivo* NORAD interactome, a less extensively studied lncRNA, adds additional depth to the pool of knowledge surrounding the function of this lncRNA.

The centerpiece of the HyPR-MS strategy, the toehold-mediated release, allows for the isolation of multiple RNA targets and for extensive cost and time reductions but also provides the means for robust downstream analysis. Because the different RNA targets are isolated from one cell culture, the background for each sample is nearly identical. Additionally, because one can directly compare the interactomes of different RNAs, it can be ascertained which proteins are unique to or highly enriched in a particular RNA capture sample. We demonstrate one strategy for assessing these data by using clustering and heatmap visualization of proteins determined to be enriched in one or more lncRNA isolation. The strategy of using the scrambled capture oligonucleotide samples and lysate input samples for determination of protein enrichment was substantiated by validation of IGF2BP2 and NAP1L4, proteins that were assessed to interact with all three lncRNA species or only the NORAD lncRNA, respectively. The RIP-qPCR assays confirmed these specific conclusions and also validated that the data analysis strategy used was sound.

We have demonstrated that HyPR-MS is a powerful tool for identifying RNA interacting proteomes by reducing the technological complexity, cost, and time of the procedure and establishing its quality and flexibility. Future applications for HyPR-MS include comparisons between the interacting proteomes of individual RNA targets under different physiological conditions, the study of less abundant RNAs, and the identification of post-translational modifications of RNA interactomes for a deeper understanding of RNA biology in physiological and pathological conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Special acknowledgement goes to Dr. Michael Shortreed and Dr. Brian Frey for early discussions of experimental design and input in data analysis. Work for this manuscript was funded by NIH-NCI grant R01CA193481.

REFERENCES

(1). Khalil AM; Rinn JL RNA-protein interactions in human health and disease. Semin. Cell Dev. Biol 2011, 22 (4), 359–65. [PubMed: 21333748]

(2). Marchese D; de Groot NS; Lorenzo Gotor N; Livi CM; Tartaglia GG Advances in the characterization of RNA-binding proteins. Wiley Interdiscip Rev. RNA 2016, 7 (6), 793–810. [PubMed: 27503141]

(3). McHugh CA; Russell P; Guttman M Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol 2014, 15 (1), 203. [PubMed: 24467948]

(4). Chu C; Zhang QC; da Rocha ST; Flynn RA; Bharadwaj M; Calabrese JM; Magnuson T; Heard E; Chang HY Systematic discovery of Xist RNA binding proteins. Cell 2015, 161 (2), 404–16. [PubMed: 25843628]

(5). West JA; Davis CP; Sunwoo H; Simon MD; Sadreyev RI; Wang PI; Tolstorukov MY; Kingston RE The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol. Cell 2014, 55 (5), 791–802. [PubMed: 25155612]

(6). McHugh CA; Chen CK; Chow A; Surka CF; Tran C; McDonel P; Pandya-Jones A; Blanco M; Burghard C; Moradian A; Sweredoski MJ; Shishkin AA; Su J; Lander ES; Hess S; Plath K; Guttman M The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 2015, 521 (7551), 232–6. [PubMed: 25915022]

(7). Simon MD Insight into lncRNA biology using hybridization capture analyses. Biochim. Biophys. Acta, Gene Regul. Mech 2016, 1859 (1), 121–127.

(8). Martens-Uzunova ES; Bottcher R; Croce CM; Jenster G; Visakorpi T; Calin GA Long noncoding RNA in prostate, bladder, and kidney cancer. Eur. Urol 2014, 65 (6), 1140–51. [PubMed: 24373479]

(9). Gutschner T; Hammerle M; Eissmann M; Hsu J; Kim Y; Hung G; Revenko A; Arun G; Stentrup M; Gross M; Zornig M; MacLeod AR; Spector DL; Diederichs S The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res 2013, 73 (3), 1180–9. [PubMed: 23243023]

(10). Lo PK; Wolfson B; Zhou Q Cellular, physiological and pathological aspects of the long non-coding RNA NEAT1. Front. Biol. (Beijing, China) 2016, 11 (6), 413–426.

(11). Wilusz JE Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochim. Biophys. Acta, Gene Regul. Mech 2016, 1859 (1), 128–38.

(12). Yoshimoto R; Mayeda A; Yoshida M; Nakagawa S MALAT1 long non-coding RNA in cancer. Biochim. Biophys. Acta, Gene Regul. Mech 2016, 1859 (1), 192–9.

(13). Gutschner T; Hammerle M; Diederichs S MALAT1 – a paradigm for long noncoding RNA function in cancer. J. Mol. Med. (Heidelberg, Ger.) 2013, 91 (7), 791–801.

(14). Zhang Y; Yang L; Chen LL Life without A tail: New formats of long noncoding RNAs. Int. J. Biochem. Cell Biol 2014, 54, 338–349. [PubMed: 24513732]

(15). Wu Y; Huang C; Meng X; Li J Long Noncoding RNA MALAT1: Insights into its Biogenesis and Implications in Human Disease. Curr. Pharm. Des 2015, 21 (34), 5017–28. [PubMed: 26205289]

(16). Chujo T; Yamazaki T; Hirose T Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. Biochim. Biophys. Acta, Gene Regul. Mech 2016, 1859 (1), 139–46.

(17). Cooper DR; Carter G; Li P; Patel R; Watson JE; Patel NA Long Non-Coding RNA NEAT1 Associates with SRp40 to Temporally Regulate PPARgamma2 Splicing during Adipogenesis in 3T3-L1 Cells. Genes 2014, 5 (4), 1050–63. [PubMed: 25437750]

(18). Fox AH; Lamond AI Paraspeckles. Cold Spring Harbor Perspect. Biol 2010, 2 (7), a000687.

(19). Naganuma T; Nakagawa S; Tanigawa A; Sasaki YF; Goshima N; Hirose T Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J 2012, 31 (20), 4020–34. [PubMed: 22960638]

(20). Yamazaki T; Hirose T The building process of the functional paraspeckle with long non-coding RNAs. Front. Biosci., Elite Ed 2015, 7, 1–41. [PubMed: 25553361]

(21). Ashouri A; Sayin VI; Van den Eynden J; Singh SX; Papagiannakopoulos T; Larsson E Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events. Nat. Commun 2016, 7, 13197. [PubMed: 28959951]

(22). Tichon A; Gil N; Lubelsky Y; Havkin Solomon T; Lemze D; Itzkovitz S; Stern-Ginossar N; Ulitsky I A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. Nat. Commun 2016, 7, 12209. [PubMed: 27406171]

(23). Ventura A NORAD: Defender of the Genome. Trends Genet 2016, 32 (7), 390–2. [PubMed: 27157388]

(24). Lee S; Kopp F; Chang TC; Sataluri A; Chen B; Sivakumar S; Yu H; Xie Y; Mendell JT Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell 2016, 164 (1–2), 69–80. [PubMed: 26724866]

(25). Michalik KM; You X; Manavski Y; Doddaballapur A; Zornig M; Braun T; John D; Ponomareva Y; Chen W; Uchida S; Boon RA; Dimmeler S Long noncoding RNA MALAT1 regulates

endothelial cell function and vessel growth. Circ. Res 2014, 114 (9), 1389–97. [PubMed: 24602777]

(26). Mak W; Fang C; Holden T; Dratver MB; Lin H An Important Role of Pumilio 1 in Regulating the Development of the Mammalian Female Germline. Biol. Reprod 2016, 94 (6), 134. [PubMed: 27170441]

(27). Knoener RA; Becker JT; Scalf M; Sherer NM; Smith LM Elucidating the in vivo interactome of HIV-1 RNA by hybridization capture and mass spectrometry. Sci. Rep 2017, 7 (1), 16965. [PubMed: 29208937]

(28). Kaighn ME; Narayan KS; Ohnuki Y; Lechner JF; Jones LW Establishment and characterization of a human prostatic carcinoma cell line (PC-3). Invest Urol 1979, 17 (1), 16–23. [PubMed: 447482]

(29). Smith LM; Kelleher NL Proteoform: a single term describing protein complexity. Nat. Methods 2013, 10 (3), 186–7. [PubMed: 23443629]

(30). Yeger-Lotem E; Sharan R Human protein interaction networks across tissues and diseases. Front. Genet 2015, 6, 257. [PubMed: 26347769]

(31). Zhu J; Chen G; Zhu S; Li S; Wen Z; Bin L; Zheng Y; Shi L Identification of Tissue-Specific Protein-Coding and Non-coding Transcripts across 14 Human Tissues Using RNA-seq. Sci. Rep 2016, 6, 28400. [PubMed: 27329541]

(32). Aken BL; Ayling S; Barrell D; Clarke L; Curwen V; Fairley S; Fernandez Banet J; Billis K; Garcia Giron C; Hourlier T; Howe K; Kahari A; Kokocinski F; Martin FJ; Murphy DN; Nag R; Ruffier M; Schuster M; Tang YA; Vogel JH; White S; Zadissa A; Flicek P; Searle SM The Ensembl gene annotation system. Database 2016, 2016, baw093.

(33). Tripathi V; Ellis JD; Shen Z; Song DY; Pan Q; Watt AT; Freier SM; Bennett CF; Sharma A; Bubulya PA; Blencowe BJ; Prasanth SG; Prasanth KV The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell 2010, 39 (6), 925–38. [PubMed: 20797886]

(34). Sasaki YT; Ideue T; Sano M; Mituyama T; Hirose T MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc. Natl. Acad. Sci. U. S. A 2009, 106 (8), 2525–30. [PubMed: 19188602]

(35). Zuker M Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003, 31 (13), 3406–15. [PubMed: 12824337]

(36). Erde J; Loo RR; Loo JA Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. J. Proteome Res 2014, 13 (4), 1885–95. [PubMed: 24552128]

(37). Cox J; Mann M MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol 2008, 26 (12), 1367–72. [PubMed: 19029910]

(38). Mi H; Huang X; Muruganujan A; Tang H; Mills C; Kang D; Thomas PD PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 2017, 45 (D1), D183–D189. [PubMed: 27899595]

(39). Szklarczyk D; Morris JH; Cook H; Kuhn M; Wyder S; Simonovic M; Santos A; Doncheva NT; Roth A; Bork P; Jensen LJ; von Mering C The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017, 45 (D1), D362–D368. [PubMed: 27924014]

(40). Tyanova S; Temu T; Sinitcyn P; Carlson A; Hein MY; Geiger T; Mann M; Cox J The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods 2016, 13 (9), 731–40. [PubMed: 27348712]

(41). de Hoon MJ; Imoto S; Nolan J; Miyano S Open source clustering software. Bioinformatics 2004, 20 (9), 1453–4. [PubMed: 14871861]

(42). Saldanha AJ Java Treeview–extensible visualization of microarray data. Bioinformatics 2004, 20 (17), 3246–8. [PubMed: 15180930]

(43). Van Nostrand EL; Pratt GA; Shishkin AA; Gelboin-Burkhart C; Fang MY; Sundararaman B; Blue SM; Nguyen TB; Surka C; Elkins K; Stanton R; Rigo F; Guttman M; Yeo GW Robust

transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 2016, 13 (6), 508–14. [PubMed: 27018577]

(44). Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012, 489 (7414), 57–74. [PubMed: 22955616]

(45). Li B; Dewey CN RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf 2011, 12, 323.

(46). Dobin A; Davis CA; Schlesinger F; Drenkow J; Zaleski C; Jha S; Batut P; Chaisson M; Gingeras TR STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013, 29 (1), 15–21. [PubMed: 23104886]

(47). Spitale RC; Flynn RA; Torre EA; Kool ET; Chang HY RNA structural analysis by evolving SHAPE chemistry. Wiley Interdiscip Rev. RNA 2014, 5 (6), 867–81. [PubMed: 25132067]

(48). Engreitz JM; Pandya-Jones A; McDonel P; Shishkin A; Sirokman K; Surka C; Kadri S; Xing J; Goren A; Lander ES; Plath K; Guttman M The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 2013, 341 (6147), 1237973. [PubMed: 23828888]

(49). Dai Y; Kennedy-Darling J; Shortreed MR; Scalf M; Gasch AP; Smith LM Multiplexed Sequence-Specific Capture of Chromatin and Mass Spectrometric Discovery of Associated Proteins. Anal. Chem 2017, 89 (15), 7841–7846. [PubMed: 28654248]

(50). Kennedy-Darling J; Holden MT; Shortreed MR; Smith LM Multiplexed programmable release of captured DNA. Chem-BioChem 2014, 15 (16), 2353–6.

(51). Cox J; Hein MY; Luber CA; Paron I; Nagaraj N; Mann M Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics 2014, 13 (9), 2513–26. [PubMed: 24942700]

(52). Clemson CM; Hutchinson JN; Sara SA; Ensminger AW; Fox AH; Chess A; Lawrence JB An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol. Cell 2009, 33 (6), 717–26. [PubMed: 19217333]

(53). Hung T; Chang HY Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol 2010, 7 (5), 582–5. [PubMed: 20930520]

(54). Ishizawa J; Kojima K; Hail N, Jr.; Tabe Y; Andreeff M Expression, function, and targeting of the nuclear exporter chromosome region maintenance 1 (CRM1) protein. Pharmacol. Ther 2015, 153, 25–35. [PubMed: 26048327]

(55). Heath CG; Viphakone N; Wilson SA The role of TREX in gene expression and disease. Biochem. J 2016, 473 (19), 2911–35. [PubMed: 27679854]

(56). Sato-Kuwabara Y; Melo SA; Soares FA; Calin GA The fusion of two worlds: non-coding RNAs and extracellular vesicles– diagnostic and therapeutic implications (Review). Int. J. Oncol 2015, 46 (1), 17–27. [PubMed: 25338714]

(57). Ahadi A; Khoury S; Losseva M; Tran N A comparative analysis of lncRNAs in prostate cancer exosomes and their parental cell lines. Genomics Data 2016, 9, 7–9. [PubMed: 27330995]

(58). Qian Z; Shen Q; Yang X; Qiu Y; Zhang W The Role of Extracellular Vesicles: An Epigenetic View of the Cancer Micro-environment. BioMed Res. Int 2015, 2015, 649161. [PubMed: 26582468]

(59). Takahashi K; Yan IK; Wood J; Haga H; Patel T Involvement of extracellular vesicle long noncoding RNA (linc-VLDLR) in tumor cell responses to chemotherapy. Mol. Cancer Res 2014, 12 (10), 1377–87. [PubMed: 24874432]

(60). Bonasio R; Shiekhattar R Regulation of transcription by long noncoding RNAs. Annu. Rev. Genet 2014, 48, 433–55. [PubMed: 25251851]

(61). Lee ASY; Kranzusch PJ; Cate JHD eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. Nature 2015, 522 (7554), 111–U292. [PubMed: 25849773]

(62). Shi H; Wang X; Lu Z; Zhao BS; Ma H; Hsu PJ; Liu C; He C YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. Cell Res 2017, 27 (3), 315–328. [PubMed: 28106072]

(63). Chen RB; Liu Y; Zhuang H; Yang BC; Hei KW; Xiao MM; Hou CY; Gao HJ; Zhang XR; Jia CX; Li LJ; Li YM; Zhang N Quantitative proteomics reveals that long non-coding RNA

MALAT1 interacts with DBC1 to regulate p53 acetylation. Nucleic Acids Res 2017, 45 (17), 9947–9959. [PubMed: 28973437]

(64). Nielsen J; Christiansen J; Lykke-Andersen J; Johnsen AH; Wewer UM; Nielsen FC A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. Mol. Cell. Biol 1999, 19 (2), 1262–70. [PubMed: 9891060]

(65). Rodriguez P; Munroe D; Prawitt D; Chu LL; Bric E; Kim J; Reid LH; Davies C; Nakagama H; Loebbert R; Winterpacht A; Petruzzi MJ; Higgins MJ; Nowak N; Evans G; Shows T; Weissman BE; Zabel B; Housman DE; Pelletier J Functional characterization of human nucleosome assembly protein-2 (NAP1L4) suggests a role as a histone chaperone. Genomics 1997, 44 (3), 253–65. [PubMed: 9325046]

(66). Rodriguez P; Pelletier J; Price GB; Zannis-Hadjopoulos M NAP-2: Histone chaperone function and phosphorylation state through the cell cycle. J. Mol. Biol 2000, 298 (2), 225–238. [PubMed: 10764593]

(67). Wiese S; Reidegeld KA; Meyer HE; Warscheid B Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 2007, 7 (3), 340–50. [PubMed: 17177251]

(68). Thompson A; Schafer J; Kuhn K; Kienle S; Schwarz J; Schmidt G; Neumann T; Johnstone R; Mohammed AK; Hamon C Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem 2003, 75 (8), 1895–904. [PubMed: 12713048]

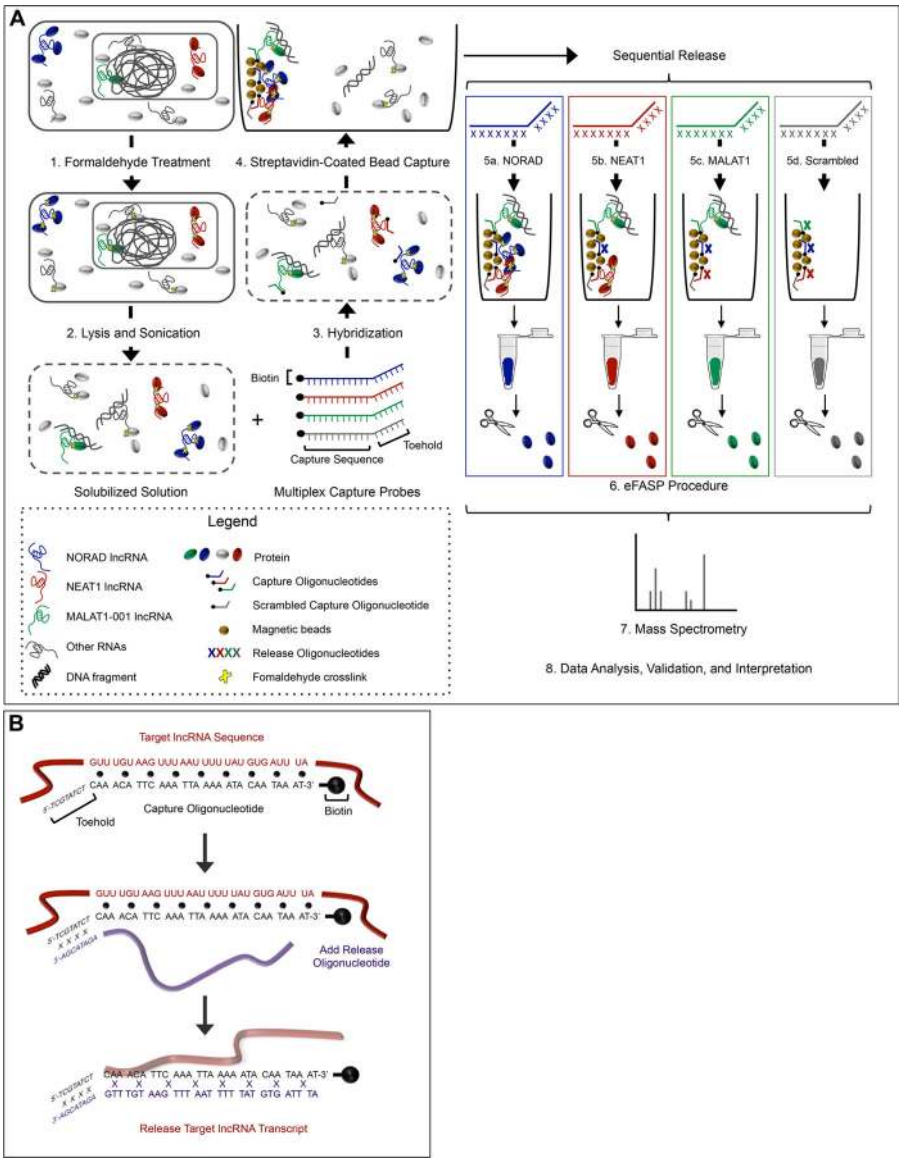


Figure 1.

Multiplexed HyPR-MS overview. (A) Formaldehyde-cross-linked cells are lysed and sonicated to solubilize the protein–nucleic acid complexes. Biotinylated capture oligonucleotides (COs) complementary to the RNA targets of interest are added and allowed to hybridize. Streptavidin-coated magnetic beads are then added to immobilize and isolate the resulting CO–RNA–protein complexes. Once the beads are washed and resuspended in release buffer, release oligonucleotides (ROs) fully complementary to the COs of a single target RNA species are added to selectively elute that RNA–protein target into solution. This release step is repeated in new release buffer for each subsequent release of the remaining RNA targets. Proteins from each release solution are then purified and analyzed by mass spectrometry. (B) A toehold-mediated capture and release strategy allows for multiplexed isolation of RNA–protein complexes. COs contain a 25–30 nucleotide (nt) sequence that is complementary to the RNA target as well as an 8 nt sequence, or toehold, that is not

complementary to the target. For release of each RNA target species from the streptavidin coated magnetic beads, an RO completely complementary to the CO sequence (25−30 nt plus toehold sequence) is added to the solution. The RO first hybridizes to the CO toehold region and then displaces the target RNA into solution by forming a thermodynamically more favorable RO−CO hybrid. This process is repeated for multiple RNA targets, allowing for their sequential release and independent analysis.

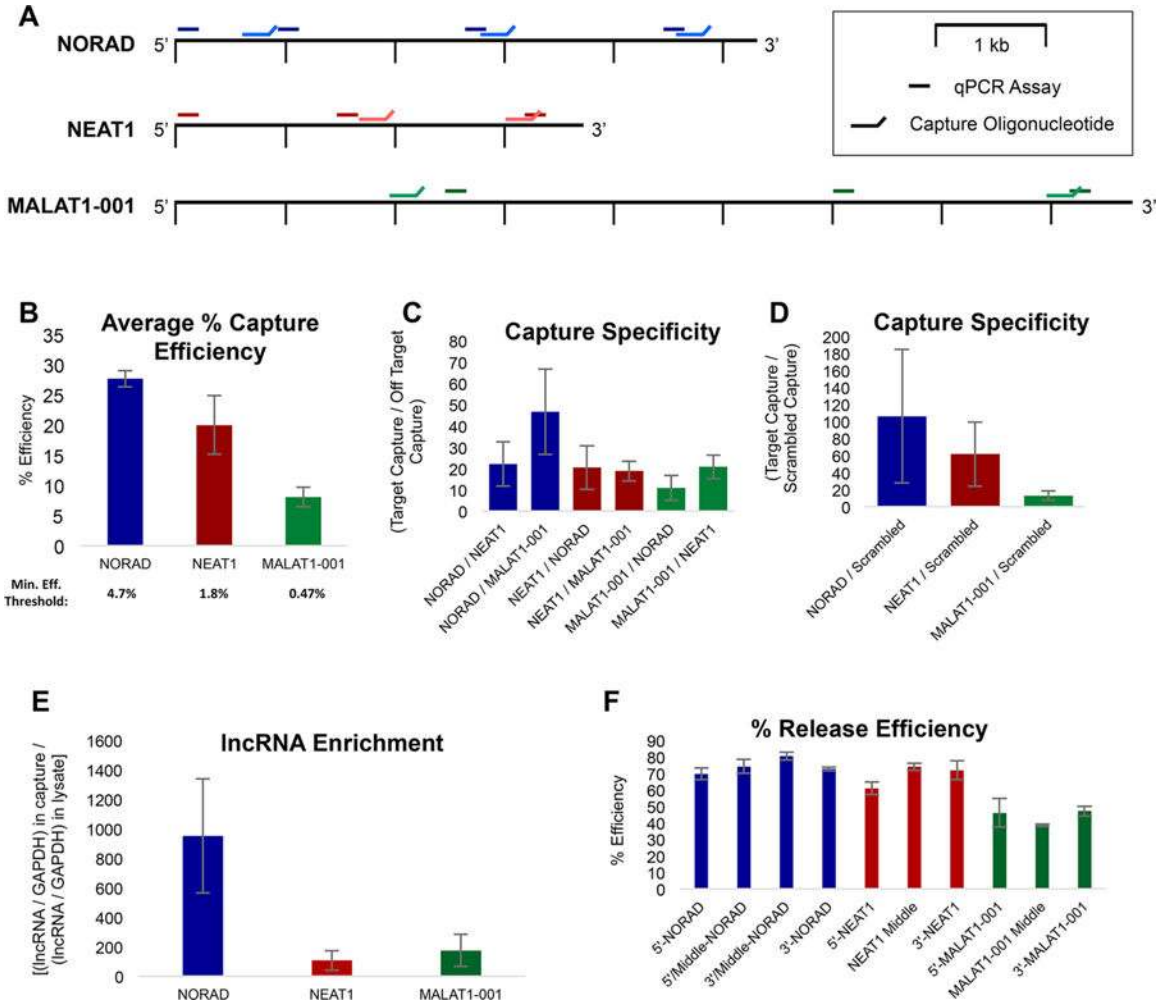


Figure 2.

HyPR-MS efficiencies and specificities. The data shown here were obtained using the appropriate RT-qPCR assays (Table S3) to measure the target lncRNA abundances in 2% of the released capture sample from the large-scale proteomics experiments. Relative quantitation was achieved for RT-qPCR calculations by using a genomic DNA calibration curve for each qPCR assay used. The data presented here was obtained from experiments using all COs for capture of each target, as shown in Table S2. (A) Approximate locations on each lncRNA for CO hybridization and for qPCR amplification products. (B) Average capture efficiency for a particular lncRNA calculated by averaging the capture efficiencies measured using each qPCR assay for that lncRNA target. Error bars represent the standard deviation of that average capture efficiency. (C–D) Capture specificity is a measure of the capacity to capture the target lncRNA while avoiding the capture of nontarget RNA molecules. Here, it is calculated by measuring the amount of a given lncRNA target captured using the complementary COs, divided by the amount of that same lncRNA captured using the COs for the alternative lncRNA targets or the scrambled CO. (E) Fold enrichment of each lncRNA target is measured by calculating the relative abundance of the target lncRNA and of a housekeeping transcript, GAPDH, in the capture sample and in the lysate sample. The lncRNA/GAPDH ratio in the capture sample is divided by the lncRNA/GAPDH ratio in

the lysate to calculate the fold enrichment of the lncRNA in the capture samples. (F) Efficiency of releasing each target from the magnetic beads with each CO-release oligonucleotide pair.

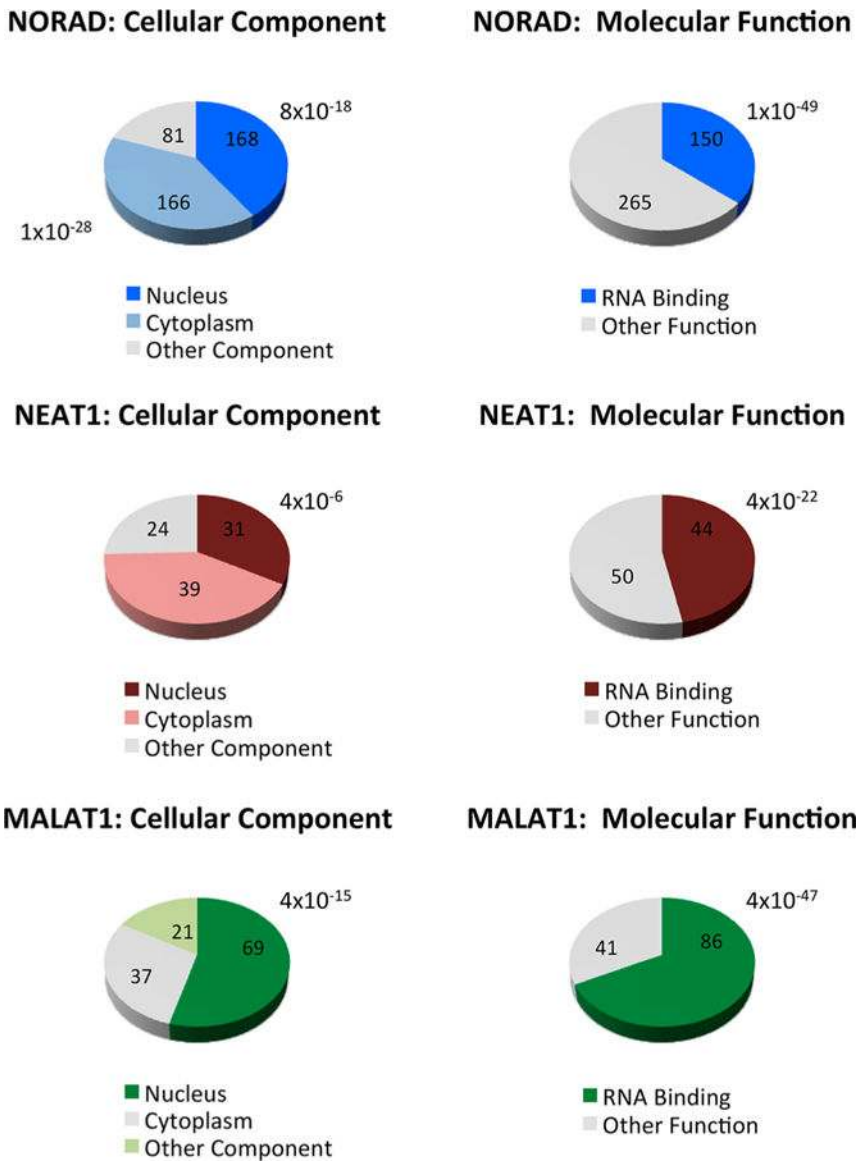


Figure 3.

Key components and functions enriched in lncRNA interactomes. Each list of proteins identified using HyPR-MS to be enriched in each lncRNA capture type was evaluated using the Gene Ontology (GO) function of Uniprot. The pie charts present the number of proteins localized to each cellular component (nucleus, cytoplasm, or other) and the number of proteins known to be RNA-binding proteins. The numbers to the outside of the pie charts are the enrichment *p*-values as determined by PANTHER[38] for the GO term identified in the chart. Proteins localized to the cytoplasm were not enriched in the NEAT1 and MALAT1 capture samples and therefore do not have a *p*-value indicated.

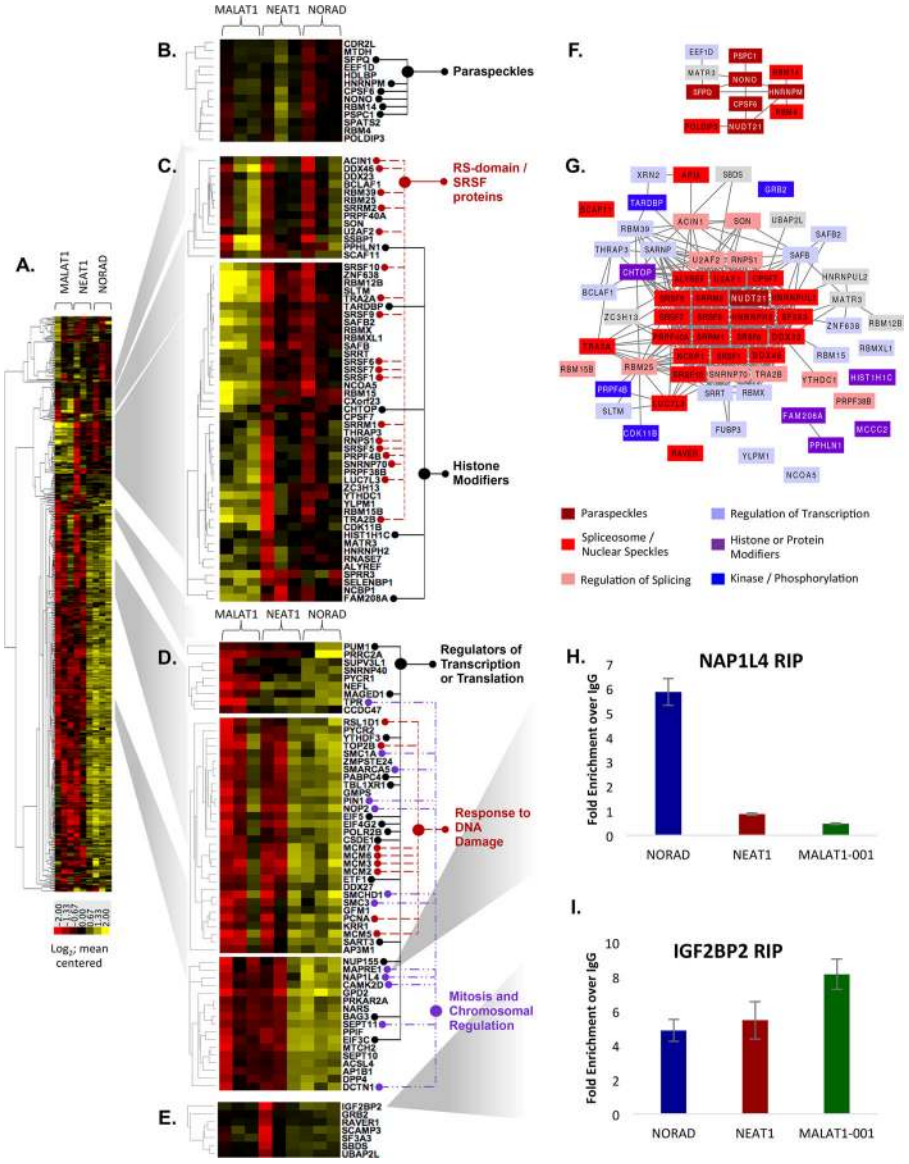


Figure 4.

HyPR-MS identifies known and novel lncRNA interactors. (A) Heat map of all proteins enriched in the capture samples of at least one lncRNA species. Yellow pixels indicate protein intensities greater than the mean intensity for that protein in all samples. Red pixels indicate protein intensities less than the mean intensity. (B–E) Clusters containing proteins with enrichment in a particular lncRNA capture type. Highlighted in these clusters are proteins with known functions related to the function of the lncRNA, they were found to interact with proteins such as paraspeckle proteins, histone modifiers, proteins with RS-domains, transcriptional and translational regulators, chromatin-associated proteins, and proteins involved in DNA damage response. (F–G) Interaction networks show the known interactions among proteins within the clusters shown in panels B–C. (H–I) lncRNA interactions with IGF2BP2 and NAP1L4 were validated using RIP-qPCR. Fold enrichments,

relative to the negative control IgG, for each lncRNA following immunoprecipitation of each protein target were determined using ΔΔCt calculations.

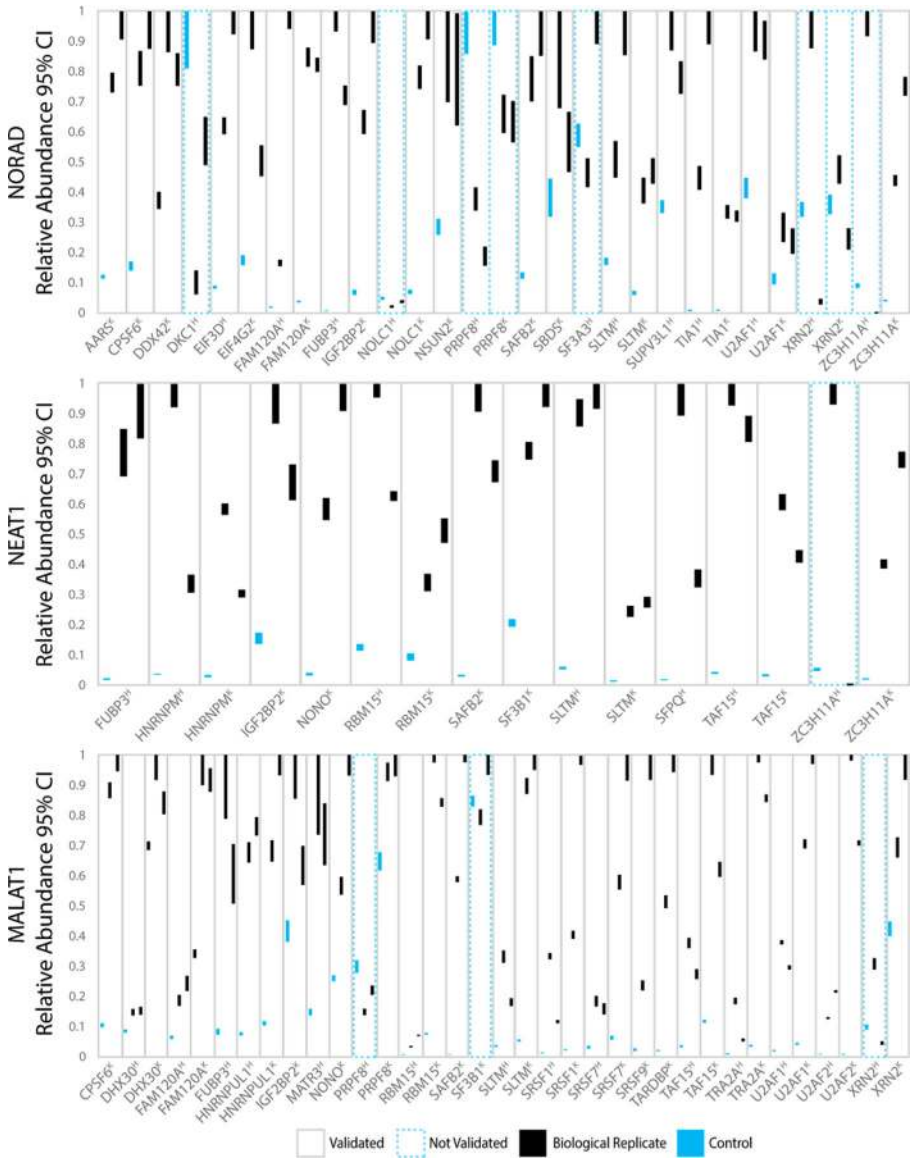


Figure 5.

Validation of lncRNA interactors with eCLIP data. Each vertical bar represents a 95% credibility interval (CI) for the abundance of the lncRNA in an eCLIP pull-down, using either K562 (^{K}) or HepG2 (^{H}) cells for two biological replicates and one control per experiment performed without cross-linking. The intervals were normalized by dividing the upper and lower bounds of the lncRNA's abundance (units in TPMs) by the maximum upper bound of the lncRNA's abundance (units in TPMs) for each target protein.

Table 1.

HyPR-MS Advances the Capabilities of in Vivo RNA Capture Technologies[a]

	CHART-MS	ChIRP-MS	RAP-MS	HyPR-MS
Capture Oligonucleotide Design	**RNase H Assay**	**Full RNA Tiling**	**Full RNA Tiling**	**Secondary Structure Prediction Software**
	A: Know available sites for hybridization, potential for isoform isolation	A: No location bias	A: No location bias	A: Decreased location bias, comparatively less costly and time-consuming, potential for isoform isolation
	L: Time-consuming, costly, location bias	L: Costly, unable to isolate isoforms	L: Costly, unable to isolate isoforms	L: Potential location bias
# of Capture Oligonucleotides Used	**1**	**43**	**142**	**2–3**
Cross-linking Conditions	**3% formaldehyde; 30 min**	**3% formaldehyde; 30 min**	**UV Cross-linking**	**1% formaldehyde; 10 min**
	A: Preserves protein–protein interactions, efficient, widely used.	A: Preserves protein–protein interactions, efficient, widely used.	A: Specific to RNA-protein interactions, irreversible	A: Preserves protein–protein interactions, efficient, widely used. Parameters selected to optimize solubilization and hybridization stages
	L: Potential fragmentation; reversal over time	L: Potential fragmentation; reversal over time	L: Low efficiency, therefore high false negatives	L: Potential fragmentation; reversal over time
Lysate Solubilization and DNA Fragment Size	Sonication; 2–10 kilobases	Sonication; 200–500 bases	Sonication; Unspecified	Sonication; About 6 kilobases
Hybridization Conditions	1.3 M urea, 800 mM salt	10% formamide, 500 mM salt	4 M urea, 500 mM salt	375 mM salt
Elution Strategy	**RNase H Digestion**	**Heat and Biotin Elution**	**RNase and DNase Digestion**	**Toehold Mediated Release**
	L: Cannot multiplex	L: Cannot multiplex	L: Cannot multiplex	A: Can multiplex to isolate multiple RNA targets from one lysate preparation. Reduces financial and time costs as well as experimental variability between samples.

[a]Highlights of the advantages (A) and limitations (L) of each technology. More detailed evaluations of these technologies, CHART-MS,[5] ChIRP-MS,[4] and RAP-MS,[48] are discussed further in the Supporting Information. See also Table S1.

Table 2.

Gene Ontology (GO) Terms Enriched in Each lncRNA Interactome[a]

			−Log$_{10}$ (*p*-value)		
Category	**Subcategory**	**GO Term**	**MALAT1**	**NEAT1**	**NORAD**
Cellular Component	Nuclear Localization	Nucleoplasm	17.5	2.8	17.0
		Nuclear Speckles	30.3	10.0	7.4
		Spliceosomal Complex	15.4	NE	3.5
		Paraspeckles	2.0	2.4	NE
		Minichromosome Maintenance	NE	NE	2.4
	Cytoplasmic Localization	Cytoplasm	NE	NE	26.1
		Extrasomal Vesicle	5.1	6.5	18.7
Molecular Function	RNA-binding	RNA-binding proteins	50.6	18.8	50.2
		RS domain binding	1.9	NE	NE
Biological Process	Transcription	DNA-templated	2.1	NE	NE
		From RNA polymerase II promoter	1.8	NE	NE
	RNA splicing	mRNA splicing via spliceosome	30.7	3.2	8.0
		regulation of alternative splicing	15.2	NE	NE
	RNA Transport	mRNA transport	8.1	NE	3.0
		RNA export from nucleus	7.5	NE	4.5
	Translation	Regulation of translation	NE	NE	8.6
	Others	Post-transcriptional regulation of gene expression	NE	NE	12.8
		Regulation of response to stress	NE	NE	2.9
		Mitotic cell cycle	NE	NE	6.7
		Regulation of RNA stability	NE	NE	2.0
		Regulation of gene silencing	NE	NE	1.8

[a]NE: no enrichment.