



Mass Spectrometry-Based Proteomics to Unveil the Non-coding RNA World

Roberto Giambruno[†], Marija Mihailovich[†] and Tiziana Bonaldi^{*}

Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Milan, Italy

OPEN ACCESS

Edited by:

Gian Gaetano Tartaglia,
Catalan Institution for Research and
Advanced Studies, Spain

Reviewed by:

Claudio Valverde,
Universidad Nacional de Quilmes
(UNQ), Argentina
Gabriele Fuchs,
University at Albany, United States

*Correspondence:

Tiziana Bonaldi
tiziana.bonaldi@ieo.it

[†]These authors share first authorship

Specialty section:

This article was submitted to
Protein and RNA Networks,
a section of the journal
Frontiers in Molecular Biosciences

Received: 26 July 2018

Accepted: 15 October 2018

Published: 08 November 2018

Citation:

Giambruno R, Mihailovich M and
Bonaldi T (2018) Mass
Spectrometry-Based Proteomics to
Unveil the Non-coding RNA World.
Front. Mol. Biosci. 5:90.
doi: 10.3389/fmolb.2018.00090

The interaction between non-coding RNAs (ncRNAs) and proteins is crucial for the stability, localization and function of the different classes of ncRNAs. Although ncRNAs, when embedded in various ribonucleoprotein (RNP) complexes, control the fundamental processes of gene expression, their biological functions and mechanisms of action are still largely unexplored. Mass Spectrometry (MS)-based proteomics has emerged as powerful tool to study the ncRNA world: on the one hand, by identifying the proteins interacting with distinct ncRNAs; on the other hand, by measuring the impact of ncRNAs on global protein levels. Here, we will first provide a concise overview on the basic principles of MS-based proteomics for systematic protein identification and quantification; then, we will recapitulate the main approaches that have been implemented for the screening of ncRNA interactors and the dissection of ncRNA-protein complex composition. Finally, we will describe examples of various proteomics strategies developed to characterize the effect of ncRNAs on gene expression, with a focus on the systematic identification of microRNA (miRNA) targets.

Keywords: ncRNA, mass spectrometry, SILAC, RNA-affinity purifications, miRNA, gene expression, proteome

INTRODUCTION

Non-coding RNAs (ncRNAs) are generally defined as transcribed, but not translated RNAs. With the exception of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs)-whose role and function have long been known- the majority of ncRNAs were considered as mere transcriptional noise until their role as key-modulators of gene expression began to be unraveled. At present, it is generally accepted that ncRNAs are central players in many biological processes, such as cell proliferation, apoptosis, differentiation and development (Beermann et al., 2016; Pasut et al., 2016; Su et al., 2016). The in-depth analysis of the mammalian transcriptome by High Throughput Sequencing (HTS) technologies revealed the existence of different types of ncRNAs including: tRNAs, rRNAs, small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), long non-coding RNAs (lncRNAs), circular RNAs (circRNAs), pseudogenes, and piwiRNAs. A detailed classification of ncRNAs was provided by P.P. Pandolfi and colleagues in Pasut et al. (2016).

Besides HTS technologies, MS-based proteomics has emerged as powerful tool to study the ncRNA world. In this review, we will first offer a concise introduction on MS-based proteomics, with emphasis on the strategies developed for protein quantitation by MS; then, focusing on studies in eukaryotic systems, we will explain how this analytical tool has been employed to address various questions in the ncRNA research field, discussing in particular the following aspects: (i) the identification and characterization of ncRNA interactomes; (ii) the dissection of protein complexes

involved in ncRNA biogenesis and function; and (iii) the measurement of the impact of ncRNAs on global gene expression (Figure 1).

Basic Principles of MS-Based Proteomics

Currently, the most common analytical strategy for large-scale protein identification in complex biological samples is shotgun MS-proteomics, whereby the identification of proteins from complex mixtures is carried out by tandem MS (MS/MS or MS₂), coupled to high performance liquid-chromatography (LC). Shotgun MS-proteomics typically is carried out using a “bottom-up” MS-approach, whereby a complex protein mixture, e.g., a whole cell extract, is first digested into peptides by a specific protease. The most frequently used enzyme is Trypsin, thanks to its selectivity in cleaving the C-terminal peptide bond of lysines and arginines and its efficiency both in-solution and in-gel conditions (Vandermarliere et al., 2013). Protease-digested peptides are then separated based on their hydrophobicity by Reversed-Phase nano-Liquid Chromatography (RP-nLC) and while eluting along a gradient of increasing concentrations of an organic buffer (typically acetonitrile, ACN) are ionized and converted into gas-phase by an Electrospray Ionization (ESI) source (Fenn et al., 1989). The volatilized peptide-ions are then accelerated through an electric and magnetic field and directly transferred into the analyzer, the core of the mass spectrometer, where peptide ions are stored and then separated based on their mass-to-charge (m/z) ratios (MS1). Hybrid mass spectrometers have been designed to combine more than one analyzer, permitting not only the measurement of the exact mass-to-charge of intact peptide ions (MS1), but to also their isolation and subsequent fragmentation to generate MS/MS (MS₂) fragmentation spectra. In particular, MS₂ spectra contain all the mass-to-charge values of the products of peptide fragmentation which provide information to extrapolate the primary sequence of peptides (peptide sequencing). The most common peptide fragmentation techniques in MS are Collision-Induced Dissociation (CID), Higher-energy Collisional Dissociation (HCD) (Olsen et al., 2007) and Electron Transfer Dissociation (ETD) (Syka et al., 2004; Brodbelt, 2016).

Several search algorithms, typically defined as search engines, have been developed to reconstruct peptide sequences starting from (MS₂); the most common are: SEQUEST (Eng et al., 1994; Yates et al., 1995), MASCOT (Perkins et al., 1999) and, more recently, Andromeda within the MaxQuant algorithm (Cox and Mann, 2008; Cox et al., 2011). All these search engines score the experimental fragmentation spectrum against the theoretical MS/MS spectrum for every peptide generated through an *in silico* digestion of the inspected proteome database with a selected protease (Paulo, 2013; Verheggen et al., 2017a). The list of candidate peptides is then filtered using a set of user-defined criteria that include the mass tolerance, the proteolytic enzyme specificity and the presence of fixed and variable modifications. The search returns a score that expresses the level of similarity between the experimental and theoretical spectra and that is therefore used as the primary parameter to discriminate correct from incorrect ID assignments. Only the best-scoring peptide matches are taken into account for the following step of

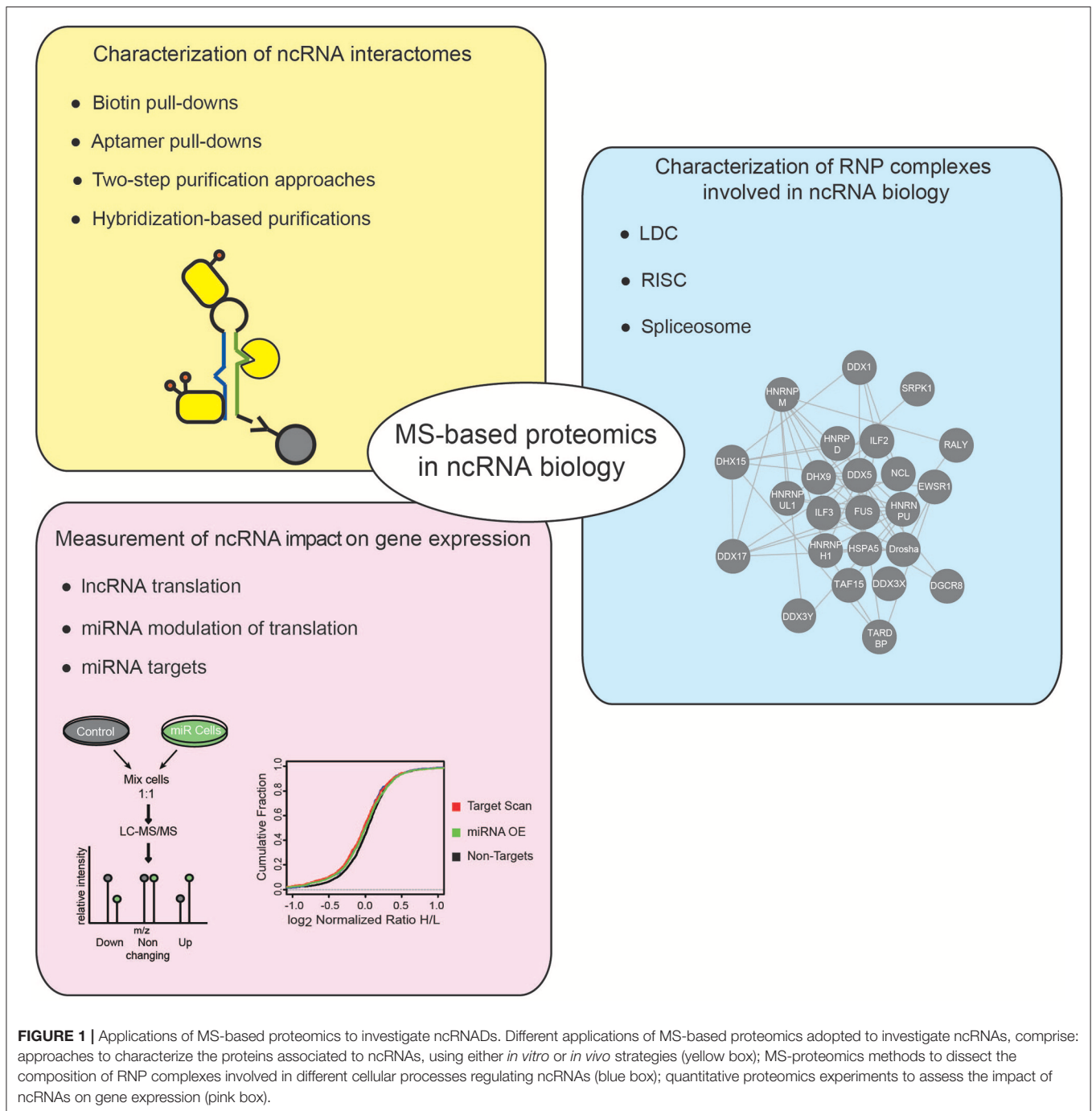
protein ID and quantification. In order to convert the score into a probability-based approach several methods have been developed, among which the target-decoy searching is currently the most common one (Elias and Gygi, 2007, 2010). Specifically, a second database (the decoy) - in which all sequences are reversed and concatenated with the original one (the target)- is used to estimate the false discovery rate (FDR) of the search, assuming that both decoy matches and false positives from the target database follow the same distribution. A defined FDR threshold is then used to filter the data and remove false positive peptide identifications, up to a fixed point (typically 1%; Bantscheff et al., 2007).

MS-based proteomics is not intrinsically quantitative, first, because the ion intensity of each peptide depends not only on its amount but also on its chemo-physical properties, directly linked to their amino acid composition; second, various external factors, such as the temperature, the presence of cross-contaminants, and the quality and stability of the nano-LC system can affect the acquisition of individual peptides within a spectrum. To overcome this limitation, two main strategies have been established to extrapolate quantitative information from MS-proteomic analyses: label free quantification (LFQ) and stable isotope-labeling approaches, summarized in Table 1 and for which a detailed review has been recently published (Lindemann et al., 2017).

LFQ strategies consist in the quantification of proteins using either intensity-based or spectral counting approaches. Intensity-based approaches use the extracted ion chromatogram (XIC) corresponding to the m/z ratios of each peptide. Under the assumption that the XIC value linearly correlates with the peptide abundance, this value can be employed to compare the quantity of the same peptide in different samples (Higgs et al., 2013). These approaches require high reproducibility in chromatography and rely on specific software that perform the chromatographic retention time (RT) re-alignment and the normalization of each peptide-intensity over the global chromatogram intensity (also defined as Total Ion Count, TIC). Spectral counting strategies, instead, measure the number of MS/MS spectra associated to each protein, which is assumed to linearly correlate with the protein abundance. The comparison of the number of spectra for each protein within set of experiments provide a relative index of its abundance in multiple samples. However, since the chemo-physical properties of each peptide can affect this linear correlation, accurate quantification is achieved only for proteins identified with a high number of spectra, while the quantification of low abundant and small proteins is less accurate (Bantscheff et al., 2007, 2012).

LFQ strategies are applicable to the quantification of a virtually-infinite number of samples, from any type of sources (cells, tissues, whole organisms). Nevertheless, these approaches require very high technical and experimental reproducibility and numerous biological replicates to reach confident protein quantification. LFQ approaches are reviewed in Bantscheff et al. (2012), Lai et al. (2013), Megger et al. (2013), and Lindemann et al. (2017).

In isotope-labeling strategies, proteins or peptides are labeled with stable (heavy) isotopes of various elements that



make them chemo-physical identical to their natural (light) counterparts, except for a specific difference in their nominal mass. Therefore, when heavy and light proteomes are mixed, each peptide is detected in the mass spectrum as a peptide-pair, whereby the two peaks are virtually identical- except for a specific mass difference (delta mass) distinguishing them- and the ratio of the intensities of the heavy and light peptides is directly proportional to the respective abundances in the samples of origin. Hence, a relative quantification is

extrapolated within the same spectrum, circumventing run-to-run variability.

Labeling of peptides with stable isotopes of C, N, H elements can be achieved *in vitro*, using a tag added covalently to the reactive side chains of amino acids, through a chemical reaction, either before or after the proteolytic cleavage. This is suitable for profiling biological samples not amenable to *in vivo* labeling, or when multiple (>3) samples must be compared. Most common methods for chemical labeling are: *Isotope-Coded Affinity Tagging*

TABLE 1 | Quantitative Mass-spectrometry-based proteomics strategies to study the ncRNA biology.

MS-strategies	Strengths	Weaknesses	Applications
Label free quantification 1) XIC 2) Spectral counting	<ul style="list-style-type: none"> • Applicability to virtually infinite number of conditions and to any biological samples • Ease of use and inexpensiveness 	<ul style="list-style-type: none"> • Requirement of high technical and experimental reproducibility • Lower accuracy in protein quantification compared to isotope-based approaches 	<ul style="list-style-type: none"> • Identification of ncRNA-protein interactions • Characterization of ncRNA-protein complexes
Chemical labeling 1) ICAT 2) iTRAQ 3) TMT	<ul style="list-style-type: none"> • Accuracy in protein quantification based on isotope labeling • Possibility of multiplexing 	<ul style="list-style-type: none"> • Expensiveness • Variable efficiency in labeling • Lower accuracy in protein quantification compared to metabolic labeling 	<ul style="list-style-type: none"> • Identification of ncRNA-protein interactions • Characterization of ncRNA-protein complexes • Impact of ncRNA on gene expression
Metabolic labeling 1) SILAC	<ul style="list-style-type: none"> • Accuracy in protein quantification based on isotope labeling • Applicability to <i>in vivo</i> studies • Minimization of quantitation error due to sample preparation • Compatibility with complex purification procedures 	<ul style="list-style-type: none"> • Limited possibility of multiplexing • Expensiveness in large-scale studies 	<ul style="list-style-type: none"> • Identification of ncRNA-protein interactions • Characterization of ncRNA-protein complexes • Impact of ncRNA on gene expression

(ICAT) (Gygi et al., 1999), *Isobaric Tags for Relative and Absolute Quantitation* (iTRAQ) (Ross et al., 2004), and *Tandem Mass Tag* (TMT) (Thompson et al., 2003). ICAT was the first chemical labeling method adopted for quantitative proteomics analyses and was specifically designed for tagging cysteines with an isotopic linker region bearing a defined delta mass and a biotin module for peptide-affinity enrichment. The main limitation of ICAT lies in the fact that only cysteine-containing proteins are labeled and quantified, which reduces significantly the number of peptides profiled, and thus the depth and accuracy of proteome quantification. In iTRAQ, instead, the peptide N-terminus and the amino-groups of the lysine side chains are labeled with an isobaric tag that allows the quantification upon fragmentation in MS/MS spectra. The iTRAQ tag includes an amino-reactive group, a balance group, required to maintain constant the mass between the different isotopes, and a reporter group, used for the relative quantification at the MS2 level. The advantage of iTRAQ is the possibility of multiplexing, using up to 8 distinct isobaric tags. However, the labeling efficiency can be variable and dependent on the sample complexity, thus generating possible quantification variability. Moreover, an additional source of error can derive from the later step of labeling within the sample preparation workflow prior to MS. The TMT approach is similar to iTRAQ and consists in the isobaric-labeling of the N-terminus and lysine residues of peptides through a tag composed of four regions: a mass reporter region, a cleavable linker, a region that works as “normalizator of total mass” and a protein-reactive group (Figure 2). Pairs of TMT-labeled peptides have identical reagent structure with the same overall mass, but contain a different combination of carbon (C) and nitrogen (N) isotopes, which make them distinguishable and quantifiable upon MS/MS fragmentation. The major advantage of TMT labeling is the high multiplexing possibility, with up to 11 different samples that can be combined and profiled in parallel with no effects on the quantity and quality of the

detected peptides (Thompson et al., 2003; Stepanova et al., 2018).

In the metabolic labeling strategies, the isotope is provided as a metabolic precursor to dividing cells, so that it is incorporated in the proteome during cell replication and protein neo-synthesis. This strategy is advantageous when applied to *in vivo* studies and its reliability relies on the fact that the sample mixing upon differential labeling can be carried out at the very early stages of the sample preparation workflow, thus avoiding possible biases deriving from the variability in the experimental procedure. Thus, metabolic labeling approaches are particularly well-suited for accurate protein quantitation when complex, multi-step sample preparation protocols are needed. The most popular strategy is *Stable Isotope Labeling by Amino Acids in Cell Culture* (SILAC) (Ong et al., 2004; Ong and Mann, 2006; Cao et al., 2013; Lau et al., 2014), which consists in the growth of replicating cells in culturing media complemented with the isotope-encoded version of specific essential amino acids, allowing the incorporation of stable isotopes into proteins during *de novo* protein synthesis. In particular, cells are grown in an *ad hoc* medium that contains either the heavy (^2H , ^{13}C , and ^{15}N) or light (H, ^{12}C , and ^{14}N) versions of lysine and arginine for a number of passages that enable their full incorporation into the proteomes (Figure 2). Heavy (H) and light (L) cells are then harvested, mixed in equal amounts and proteins are separated, digested and subjected to MS-analysis. In this setup, each protein-derived peptide exists in MS1 as a peak-pair, with the H and L counterparts being distinguishable by the delta mass. Hence, the intensity of the two peaks is an indicator of the corresponding amount in the two SILAC states. SILAC labeling has been applied to a wide variety of studies, including protein expression profiling, global PTM analysis, protein-protein and nucleic acids-protein interaction analyses. Moreover, a modified version of the standard SILAC, named pulsed-SILAC (pSILAC) was successfully adopted to

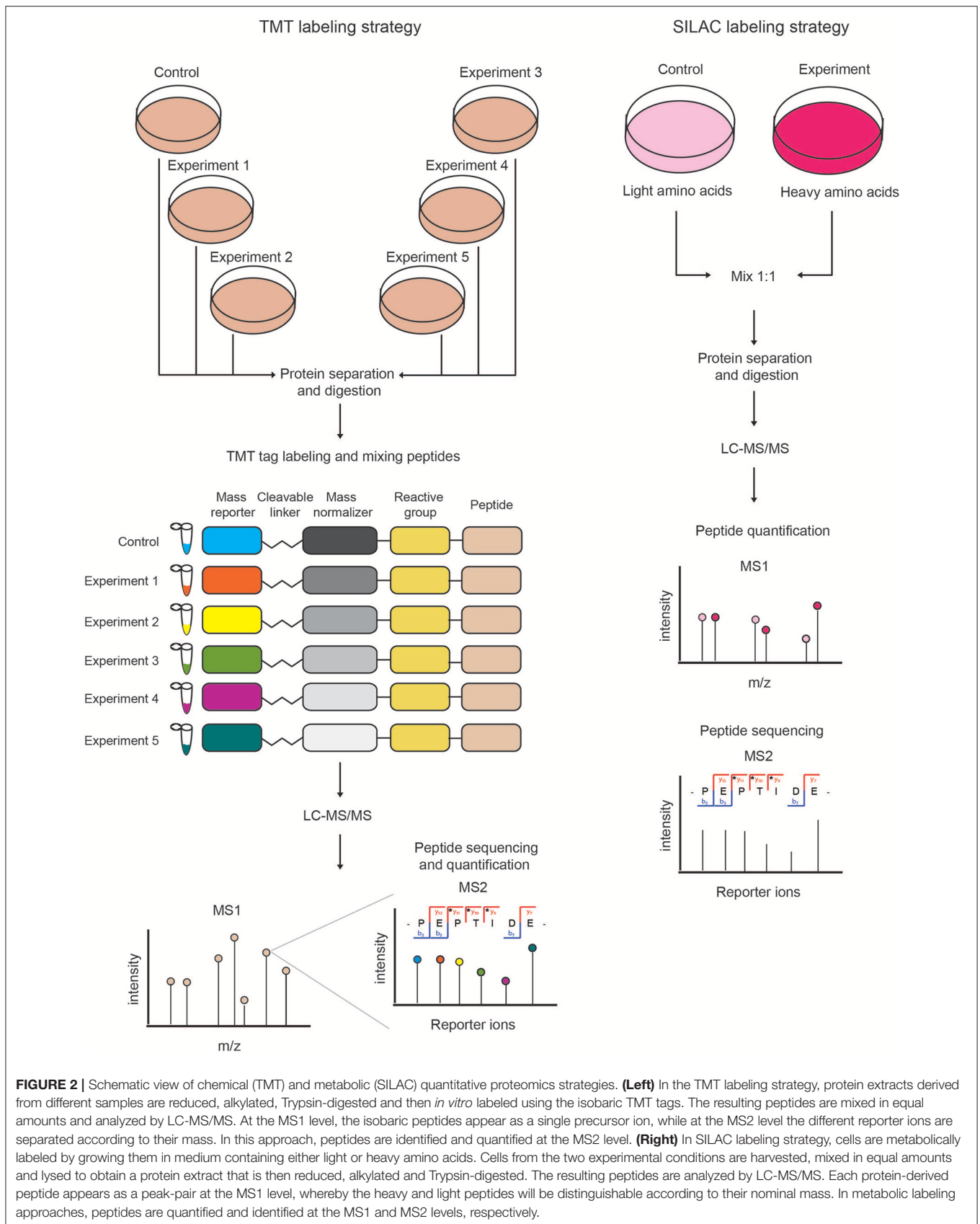


FIGURE 2 | Schematic view of chemical (TMT) and metabolic (SILAC) quantitative proteomics strategies. **(Left)** In the TMT labeling strategy, protein extracts derived from different samples are reduced, alkylated, Trypsin-digested and then *in vitro* labeled using the isobaric TMT tags. The resulting peptides are mixed in equal amounts and analyzed by LC-MS/MS. At the MS1 level, the isobaric peptides appear as a single precursor ion, while at the MS2 level the different reporter ions are separated according to their mass. In this approach, peptides are identified and quantified at the MS2 level. **(Right)** In SILAC labeling strategy, cells are metabolically labeled by growing them in medium containing either light or heavy amino acids. Cells from the two experimental conditions are harvested, mixed in equal amounts and lysed to obtain a protein extract that is then reduced, alkylated and Trypsin-digested. The resulting peptides are analyzed by LC-MS/MS. Each protein-derived peptide appears as a peak-pair at the MS1 level, whereby the heavy and light peptides will be distinguishable according to their nominal mass. In metabolic labeling approaches, peptides are quantified and identified at the MS1 and MS2 levels, respectively.

quantitatively assess protein translation dynamics, as described below.

MS-Based Proteomics for the Systematic Analysis of ncRNA-Protein Interactions

In the last two decades, various strategies have been developed to comprehensively analyze ribonucleoprotein (RNP) complexes. They can be grouped based on whether the RNAs or the proteins are used as baits for RNP complex enrichment and dissection: in the *RNA-centric* approaches, the RNA of interest is used as a bait to enrich and identify the respective protein interactors by MS-proteomics; in *protein-centric* approaches, specific RNA-binding proteins (RBPs) are used as baits to characterize the respective RNA interactors by RNA-sequencing analysis. Given the focus of this review on proteomic methods in ncRNA research, we will focus only on RNA-centric strategies.

Historically, *in vitro* RNA pull-down coupled to MS-based proteomics has been the most successful approach to identify RBPs regulating various RNAs. Improvements of this method led to the development of *in vivo* RNA immuno-purification assays, which enabled the purification of native RNPs from living cells. In the last few years, more global approaches to study RBPs have shifted the focus from the analysis of single RNAs to the comprehensive investigation of global cellular RNA-interactomes (Castello et al., 2013).

RNA Affinity-Purification Assays

RNA affinity-purification assays coupled to MS-based proteomics are used for the purification of ncRNA interactomes, both *in vitro* and *in vivo* [reviewed in Yang et al. (2015), Jazurek et al. (2016), Faoro and Ataide (2014)]. In this assay, a co-transcriptionally labeled, or tagged, RNA is used as a bait that can be either directly incubated with the protein extract, or bound to a solid support, prior to incubation. The proteins interacting to the labeled/tagged RNA are first immuno-precipitated and then identified by MS-proteomics. Variants of this basic RNA affinity-purification strategy have been developed, with technical improvements that have led to: (a) the implementation of different tagging molecules that favor the proper folding of the bait-RNA and minimize structural interference during the assembly of RNP complexes; (b) the introduction of two-step purification systems that enable more specific elution; (c) quantitative proteomics strategies that potentiate the discrimination of specific RNA-interactors from unspecific binders. Among them, SILAC is particularly well-suited to analyze low abundant and/or transient interactions, which is often the case when investigating protein-RNA associations.

Biotin-Labeled RNA Pull-Downs

Biotin (vitamin H) is the most common tag used in RNA pull-down experiments. It can either be incorporated during *in vitro* transcription, when the RNA is synthesized in the presence of biotinylated nucleotides, or it can be added at the 3'- or 5'-ends of pre-synthesized RNA by enzymatic reaction. Upon incubation with the protein extract, proteins bound to biotinylated-RNAs are purified through streptavidin beads. The interaction between

biotin and streptavidin is very strong, highly specific and resistant to high salt concentrations, high temperatures and extreme pH, which is particularly advantageous because it allows performing very stringent washes (Jazurek et al., 2016). However, this interaction is so potent that elution using soluble biotin in excess is precluded and the elution is thus performed with strong denaturing buffers, or with RNase A, which may release sticky proteins from the beads. Also, the incorporation of biotinylated nucleotides can interfere with the proper folding of the bait-RNA, thus impairing the proper binding of some interactors to specific secondary structures of the RNA; on the other hand, 3'- or 5'-end tagging can be inefficient (Jazurek et al., 2016). In spite of these limitations, biotin has led to the characterization of the interactomes of various lncRNAs, such as HOTAIR, *Firre* and lincRNA-p21 (Rinn et al., 2007; Huarte et al., 2010; Hacisuleyman et al., 2014).

Aptamers

Aptamers are short oligonucleotides or peptides that bind with high affinity to specific target molecules. They are very attractive RNA-tags, thanks to their suitability for purification under native conditions, for both *in vivo* and *in vitro* experiments. Some aptamers are naturally-occurring RNA stem-loop sequences (e.g., MS2 and PP7), while others are derived from library screenings (S1, D8, tobramycin and streptomycin). The MS2 and PP7 aptamers are based on the bacteriophage system: they bind with high specificity to the *Escherichia coli* bacteriophage coat protein MS2 (MS2cp) and *Pseudomonas aeruginosa* bacteriophage coat protein PP7, respectively (Johansson et al., 1997; Lim et al., 2001). In a RNA pull-down setting, the bait RNA is tagged at the 3'- or 5'- end with repeats of either the MS2-binding or the PP7-binding RNA stem-loops, which enable the affinity-purification of the RNPs containing the tagged RNA by using immobilized MS2 or PP7, respectively. MS2 can also be fused to the maltose-binding protein (MBP), which allows the selective elution of the RNP complexes by the addition of molar excess of soluble maltose (Zhou and Reed, 2003). This purification strategy, in combination with SILAC-based quantitative protein profiling, was applied to study the proteins associated to the lncRNA HOTAIR (Meredith et al., 2016). MS2-tagged HOTAIR and control-RNA were incubated with nuclear extracts from differentially SILAC-labeled HeLa and MDA-MB-231 cells; binding proteins were purified via MS2-MBP conjugated to amylose resin and *bona-fide* HOTAIR interactors were distinguished from the background proteins on the basis of their bait-over-control SILAC ratio. With the same approach, the interactomes of the lncRNAs MEG3 and trRNA were also characterized, upon the implementation of an additional cross-linking step, by either UV or formaldehyde, to stabilize the RNP complexes (Gumireddy et al., 2013; Liu et al., 2015).

Two-Step Purification Approaches

Aptamers are also commonly used in the two-step purification strategies, developed to increase the purity of the isolated RNP complexes reducing the contamination from nonspecific proteins. The first two-step purification method, named RAT (*RNA Affinity in Tandem*), was developed by Hogg and colleagues

for the characterization of proteins associated to the 7SK RNA (Hogg and Collins, 2007). In the RAT approach, the 7SK RNA is tagged with both PP7 and tobramycin aptamers and expressed in cells together with the recombinant PP7 carrying a TEV protease cleavage site. PP7 is used for the first purification step of native RNPs; upon elution with TEV, the tobramycin resin is used for the second purification step. RAT paved the way to various two-step purification variants, the majority taking advantage of the MS2 aptamer. For instance, Slobodin and Gerst developed the RaPID (*RBP Purification and Identification*) method that employs a MS2-based fluorescent reporter (MS2-CP-GFP) fused with the streptavidin-binding protein (SBP) tag, thus allowing both the visualization of the mRNAs bearing the MS2 aptamer and the purification of the protein interactors of the MS2-tagged RNAs by streptavidin beads (Slobodin and Gerst, 2010). Tsai and colleagues developed the MS2-BioTRAP [*MS2 in vivo Biotin Tagged RNA Affinity Purification*, (Tsai et al., 2011)], where the cells are co-transfected with both the MS2-tagged RNA of interest and the MS2 protein fused with the HB-tag, which, in turn, consists of the following elements: a hexahistidin tag, a TEV cleavage site and a bacterially-derived signal peptide for *in vivo* biotinylation in mammalian cells (Tagwerker et al., 2006). Upon expression, the MS2-tagged RNA with its interacting proteins is specifically bound to the MS2-HB biotinylated protein and the whole MS2-RNP complex is first stabilized by UV cross-linking, and then purified on streptavidin beads. Moreover, Tsai and colleagues used SILAC labeling to distinguish genuine interactors from unspecific background. The Gorospe lab improved further this method developing the MS2-TRAP (*MS2-Tagged RNA Affinity Purification*), whereby the MS2-tagged RNAs and the chimeric MS2-glutathione-S-transferase (MS2-GST) protein are co-expressed in mammalian cells, thus permitting the purification of native RNP complexes formed on the MS2-tagged RNA with a glutathione-sepharose resin (Yoon et al., 2012).

Hybridization-Based Purification Strategies

Hybridization-based purification strategies have been originally used for the systematic mapping of lncRNAs along the genome; subsequently, these techniques have been applied also to the investigation of lncRNA interactomes. For example, Chu and colleagues modified the ChIRP-Seq technique (Chu et al., 2011) into the ChIRP-MS approach (*Comprehensive Identification of RNA-binding Proteins by Mass Spectrometry*), to enable the large-scale identification of lncRNA-bound proteins *in vivo*. The workflow of both methods is the same: cells are first cross-linked *in vivo*, chromatin is extracted and sheared by sonication, and then biotinylated oligonucleotides complementary to the complete lncRNA sequence are added and let hybridize with the lncRNA. The hybrids, which comprise the target lncRNA, the associated cross-linked proteins and the chromatin, are purified using streptavidin beads. While in ChIRP-Seq the aim is to capture and sequence the genomic DNA binding to the lncRNA of interest, in ChIRP-MS the biotin-elution is followed by de-cross-linking,

protein extraction and MS-analysis. This method was successfully applied to comprehensively identifying the Xist- interactome (Chu et al., 2015). A very similar method is CHART-MS (*Capture Hybridization Analysis of RNA Targets by Mass Spectrometry*), a derivative of CHART-Seq (Simon et al., 2011) that differs from ChIRP mainly by the smaller size of the biotinylated probes used for purification, which are exclusively complementary to the lncRNA domain binding to DNA. CHART-MS was successfully used to identify the proteins interactors of the lncRNAs MALAT1 and NEAT1 (West et al., 2014).

The Xist interactors were also studied by the RAP-MS (*RNA Antisense Purification by Mass Spectrometry*) method (McHugh et al., 2015) that is a proteomic adaptation of the RAP method (Engreitz et al., 2013). Although also RAP and RAP-MS make use of biotinylated antisense probes, they are more efficient in RBPs identification than previous strategies, thanks to a number of improvements, which include: (a) the use of longer (>60 nucleotides) biotinylated antisense probes that enable the formation of very stable RNA-DNA hybrids, thus permitting more stringent washing steps during the lncRNA-protein complex purification; (b) the use of UV instead of formaldehyde for cross-linking, which allows fixing exclusively the direct binders to the RNAs; (c) SILAC-based protein profiling was employed in combination with RAP, to facilitate the discrimination of specific binders from background proteins. All these aspects have made RAP the elective strategy for the combined identification of the DNA-binding sites and protein-interactors of lncRNAs, such as in the case of the Firre lncRNA (Hacisuleyman et al., 2014).

Another attractive hybridization-based method that enables to directly monitor local RNPs *in vivo*, is PAIR (*PNA (Peptide Nucleic Acid)-Assisted Identification of RBPs*), which was successfully applied to study proteins associated to *ank*, a pan-neuronal dendritic mRNA (Zielinski et al., 2006). The PAIR method is based on the use of specific RNA-binding probes (PNAs) that hybridize with high specificity and selectivity to the endogenous target RNA and are cell permeable, thanks to the presence of a cell-penetrating peptide (Margus et al., 2012). PNAs contain also a photo-activated compound, which covalently cross-links to the associated RBPs when cells are exposed to UV light and thus stabilizes all direct interactors, allowing stringent washing steps. RNP complexes are then purified through biotinylated sense (antisense to PNA) oligonucleotides, coupled to streptavidin beads; afterwards co-associated proteins are identified by MS-based proteomics.

Pitfalls in ncRNA-Interactor Identification by MS-Based Proteomics

The choice of the ideal RNA affinity-purification strategy for specific research goals should take into account the various experimental parameters that may influence the isolation of RNP complexes prior to proteomics analysis, which include: (1) the expression level and folding of the RNA used as bait; (2) the cellular localization of the RNA; (3) the type of cell-lysis protocol employed; (4) the RNA-to-RBP stoichiometry; and

(5) the stability of the RNP complex under investigation. These experimental issues are reviewed in more detail in Oeffinger (2012).

In vitro RNA affinity-purification approaches are generally the first experimental choice because they are relatively fast and manageable and require overall small amount of starting material. However, *in vitro* purifications usually suffer from a large excess of bait-RNA that affects the physiological RNA-to-RBP stoichiometry. In addition, *in vitro* synthesized RNA might not fold properly and the protein extract used for the purification is not restricted to the cellular compartment where the endogenous RNA resides. All together, these aspects might favor the formation of artificial and unspecific protein-RNA interactions (McHugh et al., 2014).

In vivo RNA affinity-purification strategies are generally more challenging from a technical point of view, but they allow preserving the native protein-RNA interactions, catching the physiological RNP complexes in the cell. Such experiments have been mainly carried out using highly abundant or over-expressed RNAs, as baits. In the latter condition, however, the over-expressed RNA may display either aberrant cellular localization or altered RNA-RBP stoichiometry (Riley et al., 2012; Jazurek et al., 2016).

In addition, both *in vitro* or *in vivo* RNA affinity-purification strategies are often biased toward the most abundant RBPs, such as hnRNPs, RNA helicases, ribosomal and spliceosomal proteins, which might promiscuously associate with any RNA sequence (Butter et al., 2009; McHugh et al., 2014). To this regard, crosslinking strategies which stabilize protein-RNA interactions and thus allow using stringent washing conditions are useful both to limit the cross-contamination from abundant sticky proteins and to increase the identification of transient and weak, although specific, interactions. In general, crosslinking strategies are particularly suitable for the identification of the protein interactors of low abundant RNAs (McHugh et al., 2014; Ferrè et al., 2016).

RNA affinity-purifications in non-denaturing conditions followed by MS-based proteomic analysis often leads to the identification of hundreds of proteins, due to general stickiness of RNP complexes under investigation and to the high sensitivity of modern mass spectrometers. In this context, quantitative MS approaches, such as SILAC, iTRAQ, TMT are crucial to discriminate “true” interactors from the protein background (Meyer and Selbach, 2015; Aebersold and Mann, 2016). Nevertheless, also quantitative MS-approaches are affected by the over-representation of highly abundant proteins in the list of the identified specific interactors (Duncan et al., 2010; Hu et al., 2016; Ankney et al., 2018).

MS-Based Proteomics for the Characterization of Multi-Protein Complexes Involved in ncRNA Processing and Function

Various RNA-centric methods have been adopted for the characterization of RBPs involved in miRNA biogenesis. For instance, a method exploiting the endoribonuclease Csy4, a part of the CRISPR system in *Pseudomonas aeruginosa*, was

applied for the purification of RBPs associated to pre-let-7a, pre-miR-200a and pre-miR-342 (Lee et al., 2013). This purification method relies on various Csy4-specific features: (i) the selective recognition and cleavage of a 16-nt hairpin sequence; (ii) the strong affinity for the substrate; (iii) the presence of a histidine in the catalytic domain, which -when mutated to alanine H29A- does not affect the substrate binding affinity and specificity; (iv) the possibility of rescuing the cleavage activity of Csy4 H29A by imidazole. The peculiar conditional activity of Csy4 H29A, together with the high selectivity and affinity of the enzyme for its substrate, permits a very selective affinity-purification of RNP complexes. Typically, pre-miRNAs are tagged with 16-nt hairpin sequences at their 5'-end. Upon incubation with the cellular extract, the complex comprising the Csy4 hairpin tagged-pre-miRNA and the co-associated proteins is captured by biotinylated-Csy4 H29A, immobilized on avidin resin; upon stringent washing steps, the RNPs are eluted by imidazole. The selective transcript isolation with imidazole ensures very low background contamination which allows omitting a SDS-page purification step prior to MS, thus favoring the detection of low-abundant RBPs.

A RNA-centric approach was also employed by the group of G. Meister to annotate the most comprehensive dataset of RBPs involved in miRNA biogenesis: they used 72 *in vitro* synthesized pre-mRNA bearing the same 5' extension, complementary to biotinylated 2'-O-methyloligonucleotides, to uncover 180 proteins that show preferential or selective binding to single, or small sets of precursor-miRNAs (Treiber et al., 2017). The results collected in this study corroborated the model whereby miRNA biogenesis is a dynamic process mediated by a temporally-defined association of different RBPs.

In addition to RNA-centric approaches, classical strategies for multi-protein complexes purification, typically based on the overexpression of the protein of interest, followed by immuno-purification and MS-identification of the co-associated proteins, have been employed to characterize biochemically the nucleoprotein machineries involved in miRNA biogenesis and function. In 2004, three groups independently identified the Microprocessor complex, composed of DROSHA and DGCR8 and responsible for the pri-to-pre- miRNA cleavage step (Denli et al., 2004; Han et al., 2004; Liu et al., 2004). Immunoprecipitation of a FLAG-tagged version of both DROSHA and DGCR8 led to the identification of 21 additional auxiliary proteins that associate to the Microprocessor, forming a multi-protein complex named *Large Drosha Complex* (LDC). These auxiliary proteins are required to modulate the ribonuclease activity of the core dimer (Liu et al., 2004; Shiohama et al., 2007). The RISC (*RNA-Induced Silencing Complex*) was characterized with a similar strategy, in the same period (Chendrimada et al., 2005; Haase et al., 2005; Höck et al., 2007).

MS-Based Quantitative Proteomics to Study the Function of ncRNAs Proteomics Analyses to Assess the Translation of ncRNA Into Polypeptides

Advances in sequencing technologies have led to the discovery that pervasive transcription in eukaryotes produces an excess

of lncRNAs (Djebali et al., 2012). Initially these molecules were believed not entailing a protein-coding potential, but subsequent ribosome profiling analyses have demonstrated that lncRNAs can interact with the translational machinery (Bazzini et al., 2014; Ingolia et al., 2014; Calviello et al., 2016). Nevertheless, the question whether the observed ribosome-footprints on lncRNAs and the presence of small open reading frames (smORFs) may truly reflect an active translation into short polypeptides is still debated (de Andres-Pablo et al., 2017; Verheggen et al., 2017b). Several groups demonstrated- first in fruit-fly and zebrafish and then in mammalian cells- that some very short polypeptides (micropeptides) are produced from putative lncRNA transcripts and are functionally relevant (Pauli et al., 2014; Albuquerque et al., 2015; Anderson et al., 2015; Nelson et al., 2016). While in all these studies micropeptides were discovered starting from RNA-seq analyses followed by *in silico* prediction, the Pandolfi group took advantage of MS-based proteomics to identify for the first time a smORF encoded by the lncRNA LINC00961, thus demonstrating its translatability (Matsumoto et al., 2017). Nonetheless, as of today, <1% of the micropeptides encoded by lncRNAs has been experimentally validated by MS-proteomics (Volders et al., 2015), raising concerns about the technical possibility to annotate micropeptides by MS. Recently, Verheggen and colleagues have systematically addressed the possible biases in micropeptide detection by MS, assessing how various factors, such as size, amino acid composition, abundance and half-life- could affect the identification of peptides encoded by lncRNAs (Verheggen et al., 2017b). They observed that MS is not specifically biased against the detection of lncRNA-encoded micropeptides, so that their lack of detection with this technique may indeed be suggestive of their overall absence. Recently, U. Ohler and colleagues developed the RiboTaper statistical approach to identify translated regions from ribosome profiling data (Calviello et al., 2016). Interestingly, the experimental proteome which they acquired by MS-proteomics showed excellent match with the putative proteome predicted by RiboTaper. Indeed, only 504 ORFs within noncoding genes were identified as translated, with the majority belonging to pseudogenes; this is in line with the idea that most lncRNAs are, in fact, non-coding.

Quantitative Proteomics to Characterize miRNA Function and Targets

In the last decade, miRNAs have emerged as pervasive post-transcriptional regulators, with each miRNA being able to dampen the expression of hundreds of targets, simultaneously. They repress target genes by either destabilizing mRNAs and/or inhibiting their translation. MiRNAs bind their targets through the interaction of a seed region with the matching binding site located in the 3' untranslated region (3' UTR) of the transcript. One mRNA may have different variants bearing 3' UTRs regions of different length. Any variation in the 3' UTR length of a transcript can affect its regulation and stability as a consequence of the changes in the repertoire of regulatory elements present within the region, including regulatory RBPs and binding sites for miRNAs (Mihailovic et al., 2012; Tian and Manley, 2013). Although many computational approaches for miRNA targets

prediction have been developed, *in silico* predictions overall lack accuracy, as they do not take into account the 3' UTRs heterogeneity, as well as imperfect mRNA/miRNAs match. Hence, experimental validation of miRNA target prediction is mandatory when investigating the genuine biological activity of individual miRNAs in a specific context. Vinther and his colleagues pioneered the use of MS-based proteomics for the unbiased experimental identification of miRNA targets using SILAC-based quantitative profiling of protein level changes, upon miR-1 overexpression (Vinther et al., 2006). Although only 12 high-confident miR-1 targets were identified, this study represented the proof-of-concept that a single miRNA can reduce the translation of several proteins in parallel, and showed that SILAC-based quantitation is particularly suitable for assessing the very mild protein changes induced by miRNA modulation. The steady-state proteomic analysis may, however, underestimate the effect of differential turnover among proteins, as well as the combinatorial effect of multiple regulatory events impinging on the overall protein levels. For example, on the one hand, very stable proteins may result unaltered when the proteomics analysis is performed too shortly after miRNA modulation; on the other hand, if MS analysis is carried out too late, the proteome changes may be the result of both direct (miRNA activity) and indirect (e.g., the modulation of some transcription factors) events.

This limitation can be overcome by using the pulsed-SILAC (pSILAC) strategy, which was developed by M. Selbach in 2008 and represents a technological milestone for experimental miRNA target identification by MS (Selbach et al., 2008). In pSILAC, both control and treated cells are pulse-labeled with two distinct versions of the heavy amino acids, so that pre-existing proteins are visualized in the light channel, and the newly synthesized proteins from different functional states (e.g., untreated and perturbed) exist in the "heavy" and "medium-heavy" channels, respectively. The comparison of the heavy and medium-heavy channels for the same proteins reveals the impact of a specific perturbation (e.g., miRNA modulation) only on protein neo-synthesis/translation (Figure 3). Specifically, Selbach and colleagues globally assessed the effect of protein translation upon the modulation of both endogenous and exogenous miRNAs in HeLa cells and showed that a single miRNA can affect the expression of hundreds of proteins. Pulsed-SILAC allows a better identification of direct miRNA targets, therefore it is well-suited to distinguish targets of distinct miRNAs belonging to the same family (Ebner and Selbach, 2014), which is almost impossible with *in silico* predictions.

While studies using transient modulation of miRNAs are almost exclusively focused on target identification, the studies involving the stable overexpression or down-regulation of miRNAs can offer additional insights into the physiological role of the miRNA under investigation. For example, Baek and colleagues interrogated the effect of *mir-223* gene knockout in mouse neutrophils. They isolated bone marrow hematopoietic progenitors from wild-type and *mir-223* deficient mice and differentiated them *in vitro* in SILAC conditions, to analyze *mir-223* targets by quantitative proteomics (Baek et al., 2008). To discern direct from indirect targets, they intersected the list of up-regulated proteins with miRNA target *in silico* prediction.

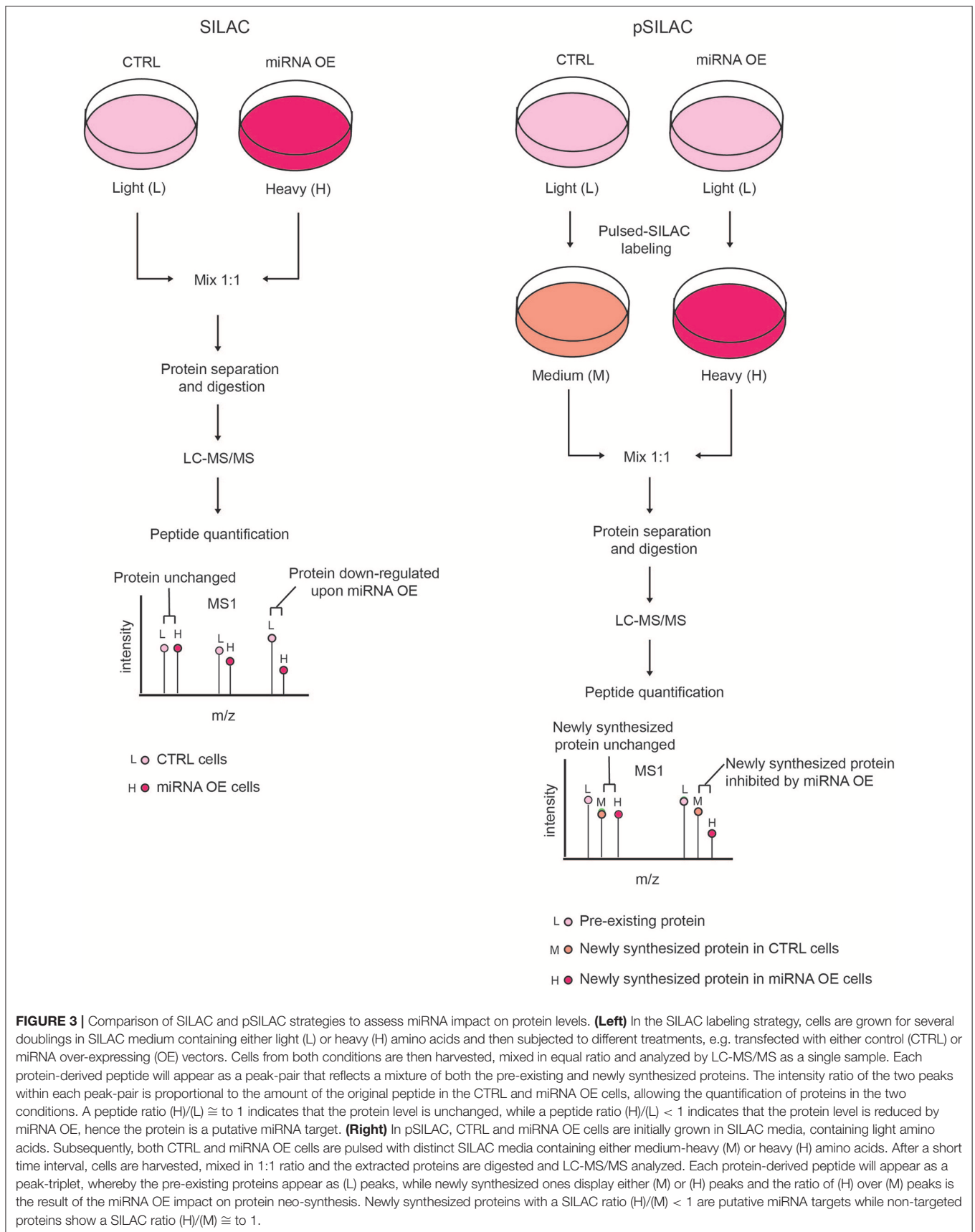


FIGURE 3 | Comparison of SILAC and pSILAC strategies to assess miRNA impact on protein levels. **(Left)** In the SILAC labeling strategy, cells are grown for several doublings in SILAC medium containing either light (L) or heavy (H) amino acids and then subjected to different treatments, e.g. transfected with either control (CTRL) or miRNA over-expressing (OE) vectors. Cells from both conditions are then harvested, mixed in equal ratio and analyzed by LC-MS/MS as a single sample. Each protein-derived peptide will appear as a peak-pair that reflects a mixture of both the pre-existing and newly synthesized proteins. The intensity ratio of the two peaks within each peak-pair is proportional to the amount of the original peptide in the CTRL and miRNA OE cells, allowing the quantification of proteins in the two conditions. A peptide ratio (H)/(L) \cong 1 indicates that the protein level is unchanged, while a peptide ratio (H)/(L) < 1 indicates that the protein level is reduced by miRNA OE, hence the protein is a putative miRNA target. **(Right)** In pSILAC, CTRL and miRNA OE cells are initially grown in SILAC media, containing light amino acids. Subsequently, both CTRL and miRNA OE cells are pulsed with distinct SILAC media containing either medium-heavy (M) or heavy (H) amino acids. After a short time interval, cells are harvested, mixed in 1:1 ratio and the extracted proteins are digested and LC-MS/MS analyzed. Each protein-derived peptide will appear as a peak-triplet, whereby the pre-existing proteins appear as (L) peaks, while newly synthesized ones display either (M) or (H) peaks and the ratio of (H) over (M) peaks is the result of the miRNA OE impact on protein neo-synthesis. Newly synthesized proteins with a SILAC ratio (H)/(M) < 1 are putative miRNA targets while non-targeted proteins show a SILAC ratio (H)/(M) \cong 1.

Although the experimental validation of the predicted targets through quantitative proteomics accurately reflects changes in their protein abundances, these changes might still be the sum of both the direct activity of miRNAs and other adaptive responses of the cell. We investigated this aspect in 2015, by carrying out an integrated analysis of SILAC-based proteomics, transcriptomics, *in silico* prediction and in-depth 3' UTR-analysis, in full-blown B cell lymphoma, where we modulated the expression of the miR-17-92 cluster. This integrated analysis of multiple -omics data revealed the importance of 3' UTR shortening in defining the effective miR-17-92 activity in full-blown B cell lymphoma, with the miRNA cluster shifting from an oncogenic to a tumor-suppressor role due to differences in mRNA landscape at different stages of tumor progression (Mihailovich et al., 2015).

In addition to SILAC and its variations, also chemical labeling methods -such as ICAT, iTRAQ and TMT- were employed in various miRNAs studies, when either the model systems under investigation were not amenable to metabolic labeling, or more extensive multiplexing was required. For instance, Li and colleagues isolated splenic B cells from three miR-146a-overexpressing transgenic mice and three controls and labeled them with 6-plex isobaric TMT tags, which enabled the simultaneous analysis of all samples in a single MS run. They identified and quantified over 5,000 proteins, from which ~200 were differentially expressed between B cells from miR-146a transgenic mice and controls (Li et al., 2017).

Taken together, MS-based quantitative proteomics has contributed significantly to the large-scale identification of miRNA targets, thanks to the implementation of *ad hoc* strategies and experimental designs.

CONCLUSIONS AND FUTURE PERSPECTIVES

Different classes of ncRNAs, including miRNAs and lncRNAs, have been described as key regulators of gene expression; with the increasing knowledge gained through HTS, novel and less characterized classes of ncRNAs will be probably included in the same regulatory group, as was the case for the recently reported circRNAs (Huang et al., 2017). A feature in common to all these molecules is their association with proteins to form RNP complexes, whose composition, organization and dynamics is spatial- and temporal- specific, and dependent on both intra- and extra- cellular stimuli. Both protein-centric and RNA-centric high-throughput methods are required to characterize the dynamicity of ncRNP complexes. The development of the *in vivo* RNA-Interactome Capture (RIC) method, coupled with MS-proteomics analysis (Castello et al., 2012) enabled the systematic and comprehensive identification of RBPs from various cell types and organism models (Baltz et al., 2012; Kwon et al., 2013; Mitchell et al., 2013; Marondedze et al., 2016; Nandan et al., 2017), offering important insights into RNA biology. The RIC approach, in fact, almost doubled the number of annotated RBPs and uncovered dozens of the so-called "enigma" RBPs, for which the corresponding RNA-binding module and RNA-partners are unknown (Hentze et al., 2018). The continuous annotation of

RBP compendia has paved the way to novel approaches to study RBPs; for example, Zappulo and colleagues identified a sub-population of 29 RBPs specifically localized in neurites, by comparing a local proteome acquired from neurites with a published repository of 1,542 RBPs (Gerstberger et al., 2014; Zappulo et al., 2017). Interestingly, a parallel ncRNA-analysis carried out in the same neuronal compartment unraveled 12 lncRNAs of unknown function and 41 circRNAs, thus suggesting that the newly identified neurite-specific RBPs might be involved in the regulation of these ncRNAs.

Very recently three similar strategies, named respectively OOPS, XRNAX, and PTex, have been developed as very promising global approaches to expand the current knowledge on RBPs and RNA-binding domains (Queiroz et al., 2018; Trendel et al., 2018; Urdaneta et al., 2018). They all adopt an organic phase separation step for the extraction and enrichment of UV-crosslinked RNA-protein complexes from whole cell extracts. The isolated complexes are then analyzed by either RNA-sequencing or MS-based proteomics for the identification of the RNA- and protein- constituents, respectively. By avoiding any RNA/protein affinity-purification step, these methodologies allow more systematic and unbiased characterization of RNA-protein interactions, including less characterized ncRNA species, in both eukaryotic and prokaryotic systems.

Although more extensively employed in eukaryotes, most of the approaches based on affinity purification coupled to MS-based proteomics discussed in this review can in principle be adopted to identify ncRBPs in prokaryotes, a less developed field which has been recently reviewed in Holmqvist and Vogel (2018). For example, already in 2009 Said et al. used the MS2-aptamer to identify *in vivo* the proteins bound to various small non-coding RNAs (sRNAs) from *Salmonella Typhimurium* (Said et al., 2009). Similarly, Rieder et al. employed *in vitro* MS2-tagged RNA pull-downs to characterize the RBPs of sRNAs from *Helicobacter Piloni* (Rieder et al., 2012). Differently, Osborne and colleagues developed the SSAC-MS/MS (Sequence-Specific Affinity Chromatography and tandem Mass Spectrometry) method, in which biotinylated cDNA probes, complementary to the sRNA target, were used to purify UV-crosslinked sRNA-RBP complexes formed *in vivo*. SSAC-MS/MS was employed to define the interactomes of the iron-responsive PrrF and PrrH sRNAs in *Pseudomonas aeruginosa* (Osborne et al., 2014). All these approaches are based on the characterization of the RBPs interacting with a specific tagged sRNA used as bait. Very recently, the successful application of the PTex strategy in *Salmonella Typhimurium* demonstrated the possibility of applying an unbiased and global identification of the numerous RNP complexes in living prokaryotic systems (Urdaneta et al., 2018).

With the expanding knowledge on multifunctional RBPs, the RNA field has focused its interest on understanding the link between post-transcriptional regulation and signaling pathways that respond to environmental and/or developmental stimuli. Along this line, the modulation of the protein components of RNPs by post-translational modifications (PTMs) has gained much attention. It has been suggested that various PTMs, such as phosphorylation, methylation, glycosylation, acetylation,

NEDDylation and ubiquitination, modulate both protein-protein and protein-RNA interactions within RNPs, thus influencing both the processing, stability, turnover and translation of mRNAs (Will and Luhrmann, 2011; Chen and Moore, 2014; García-Mauriño et al., 2017; Gonçalves et al., 2018) and the biogenesis and function of ncRNAs (Heo and Kim, 2009; Herbert et al., 2013; Hu et al., 2014; Jee and Lai, 2014; Golden et al., 2017). Interestingly, RBPs are the most frequently arginine-methylated proteins in the mammalian cells (Liu and Dreyfuss, 1995; Blackwell and Ceman, 2012) and arginine-methylation was shown to be required for the correct assembly and function of various RNP complexes, such as the spliceosome. The spliceosome is the largest RNP complex characterized in cells so far, comprising 5 snRNAs, around 50 small-nuclear RNPs (snRNPs) and more than a hundred of non-snRNP proteins that are required for proper pre-mRNA splicing (Wahl et al., 2009; Newman and Nagai, 2010). The composition of the spliceosome is highly dynamic and regulated by several PTMs, among which arginine methylation. In particular, the core components SmD1, SmD3, and SmB are methylated by protein arginine methyltransferase 5 (PRMT5) to mediate the correct assembly of some snRNP particles within the Spliceosome, in the cytoplasm compartment [Reviewed in Yu (2011)]. While a lot has been described on the role of spliceosome-methylation, the extent and function of this modification in other RBPs are largely unknown. In 2013,

through a systematic analysis of protein arginine-methylation by MS we reported for the first time that the LDC is hypermethylated at arginine residues, suggesting a possible regulatory role of this modification in miRNA biogenesis (Bremang et al., 2013).

MS-based modification-proteomics will play a pivotal role in the future to dissect the extent and regulatory impact of arginine methylation, as well as other PTMs, in defining the biogenesis, stability, location and function of RBPs.

AUTHOR CONTRIBUTIONS

RG and MM wrote the draft. TB proposed the topic and revised the text.

FUNDING

Research in TB group is supported by grants from the Italian Ministry of Health (GR-2011-02347880), the Epigen Flagship project (CNR), the Italian Association for Cancer Research (AIRC n. 15741).

ACKNOWLEDGMENTS

We thank A. Cuomo and R. Noberini for critical reading of the manuscript.

REFERENCES

- Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355. doi: 10.1038/nature19949
- Albuquerque, J. P., Tobias-Santos, V., Rodrigues, A. C., Murry, F. B., and da Fonseca R. N. (2015). small ORFs: a new class of essential genes for development. *Genet. Mol. Biol.* 38, 278–283. doi: 10.1590/S1415-475738320150009
- Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., Olson, E. N., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160, 595–606. doi: 10.1016/j.cell.2015.01.009
- Ankney, J. A., Muneer, A., and Chen, X. (2018). Relative and absolute quantitation in mass spectrometry-based proteomics. *Annu. Rev. Anal. Chem.* 11, 49–77. doi: 10.1146/annurev-anchem-061516-045357
- Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71. doi: 10.1038/nature07242
- Baltz, A. G., Munschauer, M., Schwanhauser, B., Vasile, A., Murakawa, Y., Schueler, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690. doi: 10.1016/j.molcel.2012.05.021
- Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 404, 939–965. doi: 10.1007/s00216-012-6203-4
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389, 1017–1031. doi: 10.1007/s00216-007-1486-6
- Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Giraldez, A. J., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 33, 981–993. doi: 10.1002/embj.201488411
- Beermann, J., Piccoli, M. T., Viereck, J., and Thum, T. (2016). Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiol. Rev.* 96, 1297–1325. doi: 10.1152/physrev.00041.2015
- Blackwell, E., and Ceman, S. (2012). Arginine methylation of RNA-binding proteins regulates cell function and differentiation. *Mol. Reprod. Dev.* 79, 163–175. doi: 10.1002/mrd.22024
- Bremang, M., Cuomo, A., Agresta, A. M., Stugiewicz, M., Spadotto, V., and Bonaldi, T. (2013). Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Mol. Biosyst.* 9, 2231–2247. doi: 10.1039/c3mb00009e
- Brodbeck, J. S. (2016). Ion activation methods for peptides and proteins. *Anal. Chem.* 88, 30–51. doi: 10.1021/acs.analchem.5b04563
- Butter, F., Scheibe, M., Morl, M., and Mann, M. (2009). Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10626–10631. doi: 10.1073/pnas.0812099106
- Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., et al. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* 13, 165–170. doi: 10.1038/nmeth.3688
- Cao, X. J., Zee, B. M., and Garcia, B. A. (2013). Heavy methyl-SILAC labeling coupled with liquid chromatography and high-resolution mass spectrometry to study the dynamics of site-specific histone methylation. *Methods Mol. Biol.* 977, 299–313. doi: 10.1007/978-1-62703-284-1_24
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Hentze, M. W., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406. doi: 10.1016/j.cell.2012.04.031
- Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L., et al. (2013). System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* 8, 491–500. doi: 10.1038/nprot.2013.020
- Chen, W., and Moore, M. J. (2014). The spliceosome: disorder and dynamics defined. *Curr. Opin. Struct. Biol.* 24, 141–149. doi: 10.1016/j.sbi.2014.01.009
- Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Shiekhattar, R., et al. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436, 740–744. doi: 10.1038/nature03868
- Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

- Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Chang, H. Y., et al. (2015). Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404–416. doi: 10.1016/j.cell.2015.03.025
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805. doi: 10.1021/pr101065j
- de Andres-Pablo, A., Morillon, A., and Wery, M. (2017). LncRNAs, lost in translation or licence to regulate? *Curr. Genet.* 63, 29–33. doi: 10.1007/s00294-016-0615-1
- Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* 432, 231–235. doi: 10.1038/nature03049
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Gingeras, T. R., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233
- Duncan, M. W., Aebersold, R., and Caprioli, R. M. (2010). The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* 28, 659–664. doi: 10.1038/nbt0710-659
- Ebner, O. A., and Selbach, M. (2014). Quantitative proteomic analysis of gene regulation by miR-34a and miR-34c. *PLoS ONE* 9:e92166. doi: 10.1371/journal.pone.0092166
- Elias, J. E., and Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214. doi: 10.1038/nmeth1019
- Elias, J. E., and Gygi, S. P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* 604, 55–71. doi: 10.1007/978-1-60761-444-9_5
- Eng, J. K., McCormack, A. L., and Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5, 976–989. doi: 10.1016/1044-0305(94)80016-2
- Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341:1237973. doi: 10.1126/science.1237973
- Faoro, C., and Ataide, S. F. (2014). Ribonomic approaches to study the RNA-binding proteome. *FEBS Lett.* 588, 3649–3664. doi: 10.1016/j.febslet.2014.07.039
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64–71. doi: 10.1126/science.2675315
- Ferré, F., Colantoni, A., and Helmer-Citterich M. (2016). Revealing protein-lncRNA interaction. *Brief. Bioinformatics* 17, 106–116. doi: 10.1093/bib/bbv031
- García-Mauriño, S. M., Rivero-Rodríguez, F., Velázquez-Cruz, A., Hernandez-Vellica, M., Diaz-Quintana, A., Diaz-Moreno M. A., et al. (2017). RNA binding protein regulation and cross-talk in the control of AU-rich mRNA Fate. *Front. Mol. Biosci.* 4:71. doi: 10.3389/fmolb.2017.00071
- Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813
- Golden, R. J., Chen, B., Li, T., Braun, J., Manjunath, H., Chen, X., et al. (2017). An argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* 542, 197–202. doi: 10.1038/nature21025
- Gonçalves, V., Pereira, J. F. S., and Jordan, P. (2018). Signaling pathways driving aberrant splicing in cancer cells. *Genes* 9:E9. doi: 10.3390/genes9010009
- Gumireddy, K., Li, A., Yan, J., Setoyama, T., Johannes, G. J., Huang, Q., et al. (2013). Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. *EMBO J.* 32, 2672–2684. doi: 10.1038/emboj.2013.188
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999. doi: 10.1038/13690
- Haase, A. D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gatignol, A., et al. (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* 6, 961–967. doi: 10.1038/sj.embor.7400509
- Hacisuleyman, E., Goff, L. A., Trapnell, C., Williams, A., Henao-Mejia, J., Rinn, J. L., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* 21, 198–206. doi: 10.1038/nsmb.2764
- Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., and Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* 18, 3016–3027. doi: 10.1101/gad.1262504
- Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341. doi: 10.1038/nrm.2017.130
- Heo, I., and Kim, V. N. (2009). Regulating the regulators: posttranslational modifications of RNA silencing factors. *Cell* 139, 28–31. doi: 10.1016/j.cell.2009.09.013
- Herbert, K. M., Pimienta, G., DeGregorio, S. J., Alexandrov, A., and Steitz, J. A. (2013). Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile. *Cell Rep.* 5, 1070–1081. doi: 10.1016/j.celrep.2013.10.017
- Higgs, R. E., Butler, J. P., Han, B., and Knierman, M. D. (2013). Quantitative proteomics via high resolution MS quantification: capabilities and limitations. *Int. J. Proteomics* 2013:674282. doi: 10.1155/2013/674282
- Höck, J., Weinmann, L., Ender, C., Rudel, S., Kremmer, E., Raabe, M., et al. (2007). Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. *EMBO Rep.* 8, 1052–1060. doi: 10.1038/sj.embor.7401088
- Hogg, J. R., and Collins, K. (2007). RNA-based affinity purification reveals 7SK RNP with distinct composition and regulation. *RNA* 13, 868–880. doi: 10.1261/rna.565207
- Holmqvist, E., and Vogel, J. (2018). RNA-binding proteins in bacteria. *Nat. Rev. Microbiol.* 16, 601–615. doi: 10.1038/s41579-018-0049-5
- Hu, A., Noble, W. S., Wolf-Yadlin A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res.* 5:F1000 Faculty Rev-419. doi: 10.12688/f1000research.7042.1
- Hu, F., Lai, E. C., and Okamura, K. (2014). A signaling-induced switch in dicer localization and function. *Dev. Cell* 31, 523–524. doi: 10.1016/j.devcel.2014.11.032
- Huang, S., Yang, B., Chen, B. J., Bliim, N., Ueberham, U., Janitz, M., et al. (2017). The emerging role of circular RNAs in transcriptome regulation. *Genomics* 109, 401–407. doi: 10.1016/j.ygeno.2017.06.005
- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Rinn, J. L., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409–419. doi: 10.1016/j.cell.2010.06.040
- Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Weissman, J. S., et al. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379. doi: 10.1016/j.celrep.2014.07.045
- Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K., and Krzyzosiak, W. J. (2016). Identifying proteins that bind to specific RNAs—focus on simple repeat expansion diseases. *Nucleic Acids Res.* 44, 9050–9070. doi: 10.1093/nar/gkw803
- Jee, D., and Lai, E. C. (2014). Alteration of miRNA activity via context-specific modifications of Argonaute proteins. *Trends Cell Biol.* 24, 546–553. doi: 10.1016/j.tcb.2014.04.008
- Johansson, H. E., Liljas, L., and Uhlenbeck, O. C. (1997). RNA recognition by the MS2 phage coat protein. *Semin. Virol.* 8, 176–185. doi: 10.1006/smvy.1997.0120
- Kwon, S. C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K., et al. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1122–1130. doi: 10.1038/nsmb.2638
- Lai, X., Wang, L., and Witzmann, F. A. (2013). Issues and applications in label-free quantitative mass spectrometry. *Int. J. Proteomics* 2013:756039. doi: 10.1155/2013/756039
- Lau, H. T., Lewis, K. A., and Ong, S. E. (2014). Quantifying *in vivo*, site-specific changes in protein methylation with SILAC. *Methods Mol. Biol.* 1188, 161–175. doi: 10.1007/978-1-4939-1142-4_12
- Lee, H. Y., Haurwitz, R. E., Appfel, A., Zhou, K., Smart, B., Doudna, J. A. et al. (2013). RNA-protein analysis using a conditional CRISPR nuclease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5416–5421. doi: 10.1073/pnas.1302807110
- Li, F., Huang, Y., Huang, Y. Y., Kuang, Y. S., Wei, Y. J., Wan, Y., et al. (2017). MicroRNA-146a promotes IgE class switch in B cells

- via upregulating 14-3-3sigma expression. *Mol. Immunol.* 92, 180–189. doi: 10.1016/j.molimm.2017.10.023
- Lim, F., Downey, T. P., and Peabody, D. S. (2001). Translational repression and specific RNA binding by the coat protein of the *Pseudomonas* phage PP7. *J. Biol. Chem.* 276, 22507–22513. doi: 10.1074/jbc.M102411200
- Lindemann, C., Thomaneck, N., Hundt, F., Lerari, T., Meyer, H. E., Marcus, K. et al. (2017). Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biol. Chem.* 398, 687–699. doi: 10.1515/hsz-2017-0104
- Liu, Q., and Dreyfuss, G. (1995). *In vivo* and *in vitro* arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* 15, 2800–2808. doi: 10.1128/MCB.15.5.2800
- Liu, S., Zhu, J., Jiang, T., Zhong, Y., Tie, Y., Wu, Y., et al. (2015). Identification of lncRNA MEG3 binding protein using MS2-tagged RNA affinity purification and mass spectrometry. *Appl. Biochem. Biotechnol.* 176, 1834–1845. doi: 10.1007/s12010-015-1680-5
- Liu, X., Taylor, J., Wang, S., Yianni, J., Gregory, R., Stein, J., et al. (2004). Reversal of hypertonic co-contraction after bilateral pallidal stimulation in generalised dystonia: a clinical and electromyogram case study. *Mov. Disord.* 19, 336–340. doi: 10.1002/mds.10655
- Margus, H., Padari, K., and Pooga, M. (2012). Cell-penetrating peptides as versatile vehicles for oligonucleotide delivery. *Mol. Ther.* 20, 525–533. doi: 10.1038/mt.2011.284
- Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S., and Gehring, C. (2016). The RNA-binding protein repertoire of *Arabidopsis thaliana*. *Sci. Rep.* 6:29766. doi: 10.1038/srep29766
- Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., et al. (2017). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* 541, 228–232. doi: 10.1038/nature21034
- McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., Guttman, M., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232–236. doi: 10.1038/nature14443
- McHugh, C. A., Russell, P., and Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol.* 15:203. doi: 10.1186/gb4152
- Megger, D. A., Bracht, T., Meyer, H. E., and Sitek, B. (2013). Label-free quantification in clinical proteomics. *Biochim. Biophys. Acta* 1834, 1581–1590. doi: 10.1016/j.bbapap.2013.04.001
- Meredith, E. K., Balas, M. M., Sindy, K., Haislop, K., and Johnson, A. M. (2016). An RNA matchmaker protein regulates the activity of the long noncoding RNA HOTAIR. *RNA* 22, 995–1010. doi: 10.1261/rna.055830.115
- Meyer, K., and Selbach, M. (2015). Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. *Front. Genet.* 6:237. doi: 10.3389/fgene.2015.00237
- Mihailovic, M., Wurth, L., Zambelli, F., Abaza, I., Militti, C., Mancuso, F., et al. (2012). Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR. *RNA* 18, 53–64. doi: 10.1261/rna.029603.111
- Mihailovich, M., Bremang, M., Spadotto, V., Musiani, D., Vitale, E., Varano, G., et al. (2015). miR-17-92 fine-tunes MYC expression and function to ensure optimal B cell lymphoma growth. *Nat. Commun.* 6:8725. doi: 10.1038/ncomms9725
- Mitchell, S. F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* 20, 127–133. doi: 10.1038/nsmb.2468
- Nandan, D., Thomas, S. A., Nguyen, A., Moon, K. M., Foster, L. J., and Reiner, N. E. (2017). Comprehensive identification of mRNA-binding proteins of *Leishmania donovani* by interactome capture. *PLoS ONE* 12:e0170068. doi: 10.1371/journal.pone.0170068
- Nelson, B. R., Makarewicz, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Olson, E. N., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351, 271–275. doi: 10.1126/science.aad4076
- Newman, A. J., and Nagai, K. (2010). Structural studies of the spliceosome: blind men and an elephant. *Curr. Opin. Struct. Biol.* 20, 82–89. doi: 10.1016/j.sbi.2009.12.003
- Oeffinger, M. (2012). Two steps forward—one step back: advances in affinity purification mass spectrometry of macromolecular complexes. *Proteomics* 12, 1591–1608. doi: 10.1002/pmic.201100509
- Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 4, 709–712. doi: 10.1038/nmeth1060
- Ong, S. E., and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* 1, 2650–2660. doi: 10.1038/nprot.2006.427
- Ong, S. E., Mittler, G., and Mann, M. (2004). Identifying and quantifying *in vivo* methylation sites by heavy methyl SILAC. *Nat. Methods* 1, 119–126. doi: 10.1038/nmeth715
- Osborne, J., Djapagne, L., Tran, B. Q., Goo, Y. A., and Oglesby-Sherrouse A. G. (2014). A method for *in vivo* identification of bacterial small RNA-binding proteins. *Microbiol. Biotechnol.* 3, 950–960. doi: 10.1002/mbo3.220
- Pasut, A., Matsumoto, A., Clohessy, J. G., and Pandolfi, P. P. (2016). The pleiotropic role of non-coding genes in development and cancer. *Curr. Opin. Cell Biol.* 43, 104–113. doi: 10.1016/j.ceb.2016.10.005
- Pauli, A., Norris, M. L., Valen, E., Chew, G. L., Gagnon, J. A., Schier, A. F., et al. (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* 343:1248636. doi: 10.1126/science.1248636
- Paulo, J. A. (2013). Practical and efficient searching in proteomics: a cross engine comparison. *Webmedcentral* 4:WMCPLS0052. doi: 10.9754/journal.wplus.2013.0052
- Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–67. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
- Queiroz, R. M. L., Smith, T., Villanueva, E., Monti, M., Pizzinga, M., Marti-Solano, M., et al. (2018). Unbiased dynamic characterization of RNA-protein interactions by OOPS. *BioRxiv*. doi: 10.1101/333336
- Rieder, R., Reinhardt, R., Sharma, C., and Vogel, J. (2012). Experimental tools to identify RNA-protein interactions in *Helicobacter pylori*. *RNA Biol.* 9, 520–531. doi: 10.4161/rna.20331
- Riley, K. J., Yario, T. A., and Steitz, J. A. (2012). Association of argonaute proteins and microRNAs can occur after cell lysis. *RNA* 18, 1581–1585. doi: 10.1261/rna.034934.112
- Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Chang, H. Y., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. doi: 10.1016/j.cell.2007.05.022
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Pappin, D. J., et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169. doi: 10.1074/mcp.M400129-MCP200
- Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H., and Vogel, J. (2009). *In vivo* expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res.* 37:e133. doi: 10.1093/nar/gkp719
- Selbach, M., Schwanhauser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63. doi: 10.1038/nature07228
- Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2007). Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Exp. Cell Res.* 313, 4196–4207. doi: 10.1016/j.yexcr.2007.07.020
- Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Kingston, R. E., et al. (2011). The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20497–20502. doi: 10.1073/pnas.1113536108
- Slobodkin, B., and Gerst, J. E. (2010). A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *RNA* 16, 2277–2290. doi: 10.1261/rna.2091710
- Stepanova, E., Gygi, S. P., and Paulo, J. A. (2018). Filter-based protein digestion (FPD): a detergent-free and scaffold-based strategy for TMT workflows. *J. Proteome Res.* 17, 1227–1234. doi: 10.1021/acs.jproteome.7b00840
- Su, Y., Wu, H., Pavlosky, A., Zou, L. L., Deng, X., Jevnikar, A. M. (2016). Regulatory non-coding RNA: new instruments in the orchestration of cell death. *Cell Death Dis.* 7:e2333. doi: 10.1038/cddis.2016.210
- Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9528–9533. doi: 10.1073/pnas.0402700101

- Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., et al. (2006). A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with *in vivo* cross-linking. *Mol. Cell. Proteomics* 5, 737–748. doi: 10.1074/mcp.M500368-MCP200
- Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904. doi: 10.1021/ac0262560
- Tian, B., and Manley, J. L. (2013). Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem. Sci.* 38, 312–320. doi: 10.1016/j.tibs.2013.03.005
- Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiss, J. L., Meister, G., et al. (2017). A compendium of RNA-binding proteins that regulate MicroRNA biogenesis. *Mol. Cell* 66, 270–284.e13. doi: 10.1016/j.molcel.2017.03.014
- Trendel, J., Schwarzl, T., Prakash, A., Bateman, A., Hentze, M. W., Krijgsvelde, J. (2018). The human RNA-binding proteome and its dynamics during arsenite-induced translational arrest. *BioRxiv*. doi: 10.1101/329995
- Tsai, B. P., Wang, X., Huang, L., and Waterman, M. L. (2011). Quantitative profiling of *in vivo*-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol. Cell. Proteomics* 10:M110.007385. doi: 10.1074/mcp.M110.007385
- Urdaneta, E. C., Vieira-Vieira, C. H., Hick, T., Wessels, H. H., Figini, D., Moschall, R., et al. (2018). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *BioRxiv*. doi: 10.1101/333385
- Vandermarliere, E., Mueller, M., and Martens, L. (2013). Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom. Rev.* 32, 453–465. doi: 10.1002/mas.21376
- Verheggen, K., Raeder, H., Berven, F. S., Martens, L., Barsnes, H., and Vaudel, M. (2017a). Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows. *Mass Spectrom. Rev.* doi: 10.1002/mas.21543. [Epub ahead of print].
- Verheggen, K., Volders, P. J., Mestdagh, P., Menschaert, G., Van Damme, P., Vandesompele, J., et al. (2017b). Noncoding after all: biases in proteomics data do not explain observed absence of lncRNA translation products. *J. Proteome Res.* 16, 2508–2515. doi: 10.1021/acs.jproteome.7b00085
- Vinther, J., Hedegaard, M. M., Gardner, P. P., Andersen, J. S., and Arctander, P. (2006). Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res.* 34:e107. doi: 10.1093/nar/gkl590
- Volders, P. J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., et al. (2015). An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.* 43, 4363–4364. doi: 10.1093/nar/gkv295
- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi: 10.1016/j.cell.2009.02.009
- West, J. A., Davis, C. P., Sunwoo, H., Simon, M. D., Sadreyev, R. I., Kingston, R. E., et al. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802. doi: 10.1016/j.molcel.2014.07.012
- Will, C. L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3:a003707. doi: 10.1101/cshperspect.a003707
- Yang, Y., Wen, L., and Zhu, H. (2015). Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein. *Cell Biosci.* 5:59. doi: 10.1186/s13578-015-0050-x
- Yates, J. R. III, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* 67, 1426–1436. doi: 10.1021/ac00104a020
- Yoon, J. H., Srikantan, S., and Gorospe, M. (2012). MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* 58, 81–87. doi: 10.1016/j.jymeth.2012.07.004
- Yu, M. C. (2011). The role of protein arginine methylation in mRNP dynamics. *Mol. Biol. Int.* 2011:163827. doi: 10.4061/2011/163827
- Zappulo, A., van den Bruck, D., Ciolli Mattioli, C., Franke, V., Imami, K., McShane, E., et al. (2017). RNA localization is a key determinant of neurite-enriched proteome. *Nat. Commun.* 8:583. doi: 10.1038/s41467-017-00690-6
- Zhou, Z., and Reed, R. (2003). Purification of functional RNA-protein complexes using MS2-MBP. *Curr. Protoc. Mol. Biol. Chapter 27:Unit 27 3*. doi: 10.1002/0471142727.mb2703s63
- Zielinski, J., Kilk, K., Peritz, T., Kannanayakal, T., Miyashiro, K. Y., Eberwine, J., et al. (2006). *In vivo* identification of ribonucleoprotein-RNA interactions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1557–1562. doi: 10.1073/pnas.0510611103

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Giambruno, Mihailovich and Bonaldi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.