Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein

Yongfang Yang, Liwei Wen and Hongliang Zhu*

Abstract
Tens of thousands of long non-coding RNAs (lncRNAs) have been discovered in eukarya, but their functions are largely unknown. Fortunately, lncRNA–protein interactions may offer details of how lncRNAs play important roles in various biological processes, thus identifying proteins associated with lncRNA is critical. Here we review progress of molecular archetypes that lncRNAs execute as guides, scaffolds, or decoys for protein, focusing on advantages, shortcomings and applications of various conventional and emerging technologies to probe lncRNAs and protein interactions, including protein-centric biochemistry approaches such as nRIP and CLIP, and RNA-centric biochemistry approaches such as ChIRP, CHART and RAP. Overall, this review provides strategies for probing interactions between lncRNAs and protein.

Keywords: Long non-coding RNA, Function, Interaction, Protein, Biochemistry approaches

Background
Recently, an explosion of microarray tiling and high-throughput deep sequencing analysis has led to the discovery of thousands of previously presumed non-coding transcripts [1, 2]. Global transcriptional analyses of the human genome have revealed that non-coding RNA far exceed the protein-coding mRNAs which account for only about two percent of the human genome [3]. Non-coding RNA include many small regulatory RNAs and tens of thousands of polyadenylated and nonpolyadenylated lncRNAs which have been shown to be essential for many rapidly growing research areas [4]. Although only a few lncRNA have been documented to have important biological functions, increasing evidences suggest that the regulation of lncRNAs on target genes is complicated [5].

lncRNAs, through interactions with protein, DNA and RNA, regulate gene expression at multiple levels, including chromatin remodeling and nuclear transcription, pre-mRNA splicing and cytoplasmic mRNA translation [6]. Moreover, virtually all functional RNA molecules interact with protein complexes and protein is confirmed to be the first and principal partner of lncRNA [7]. Thus, understanding lncRNA functions can be accomplished by identification of lncRNA-bound proteomes. Substantial effort is being devoted to depicting RNA–protein interactions for gaining insight into molecular mechanisms, but lncRNA–protein interplay is poorly understood [8–10]. In this review, we highlight molecular modes and functions of lncRNA–protein interactions and summarize conventional and emerging techniques to probe these interactions, with hopes of illuminating hidden lncRNA regulatory mechanisms.

Characteristics and function of lncRNA
LncRNA are a group of non-coding RNAs defined as being larger than 200 nucleotides in length, which distinguish them from small RNAs such as microRNAs, small nucleolar RNAs (snoRNAs) and small interfering RNAs (siRNAs) [11]. According to the relative position of the coding gene, lncRNA exist in four groups: intergenic, introngenic, overlap and antisense [12]. Compared with protein coding RNA, lncRNAs are typically shorter.
with fewer exons, less abundance, less coding potential and more restrictions to particular tissues or cells [13]. Moreover, lncRNAs sequences are less conserved than mRNA among related species. Recently, secondary structures of lncRNAs have been shown to be conserved, having ‘repeat A’ region in lncRNA Xist (X inactive-specific transcript) and a ‘roX-boxes’ sequence motif comprised of two lncRNAs roX1 and roX2 [14, 15]. Research that lncRNAs have various important regulatory effect on target gene expression contributing to epigenetic modification, transcription and post-transcriptional processing via specific interactions with proteins and other cellular factors [16–18].

LncRNAs mediate epigenetic changes by recruiting chromatin remodeling complexes to specific genomic loci [16]. For example, the lncRNA HOX antisense intergenic RNA (HOTAIR) which is initiated from the HOXC cluster interacts with ploycomb repressive complex 2 (PRC2) to silence transcription across 40 kb of the HOXC locus in trans by inducing a repressive chromatin state [19]. lncRNAs, Xist, RepA and Kcnqot1 all recruit the polycomb complex to the target genome and they trimethylate lysine 27 residues (me3K27) of histone H3 to induce heterochromatin formation and repress gene expression [20, 21]. In addition, lncRNA also regulate target gene at transcriptional [18]. Proximal promoters can be transcribed into long ncRNAs that recruit and integrate RNA binding proteins function into the transcriptional process [22]. For example, an lncRNA induced by DNA damage and transcribed from the cyclin D1 gene promoter, recruits and integrates RNA binding protein TLS to silence cyclin D1 gene expression [22]. LncRNAs could act as co-factors to modulate transcription factor activity. LncRNA Exf-2 is transcribed from an enhancer and recruits transcription factor DIX2 to this same enhancer to modulate transcription factor DlX2 to this same enhancer to Evf-2 is transcribed from an enhancer and recruits to modulate transcription factor activity. LncRNA gene expression [22]. LncRNAs could act as co-factors and may serve as an adaptor for Xist [36].

Molecular archetypes of the lncRNA–protein interaction

Recently, how lncRNAs control gene expression and molecular function archetypes of lncRNAs have been concerned. Wang’s group discussed four emerging archetypes of molecular functions that lncRNAs execute as signals, decoys, guides, and scaffolds via proteins, DNA and RNA interaction [27]. Here, we distill the myriad functions of lncRNA into three archetypes of molecular mechanisms to illustrate how lncRNAs directly interacting with proteins and serve as ‘guide’ to recruit protein complexes to target genes, serve as ‘scaffold’ to assemble proteins into RNPs, and serve as ‘decoy’ to sequester regulatory proteins away from target gene [28]. We then offer examples of each archetype’s lncRNA–protein interactions.

LncRNAs act as protein guides

First, lncRNA acts as guide to recruit proteins to chromatin sites through RNA-DNA base pairing, to regulate downstream gene expression (Fig. 1a). For example, lncRNA fetal-lethal non-coding developmental regulatory RNA (Fendrr), which is specifically transcribed in nascent lateral plate mesoderm of the developing mouse embryo, guides PRC2 to target genes to increase PRC2 occupancy and trimethylate of H3K27me3, subsequently leading to attenuation of target gene expression [29]. Cold assisted intronic non-coding RNA (COLDAIR), a cold-induced lncRNA with a capped 5’ end lacking a 3’ poly A tail, is transcribed from the intron of the FLC locus gene in Arabidopsis thaliana and recruits PRC2 to establish and maintain stable repressive chromatin of FLC through H3K27 trimethylation during vernalization [30]. Other than recruiting PRC2 to repress target gene expression, some lncRNAs, such as lncRNA-p21 can guide hnRNP-K protein to the promoter of the p21 gene and act as a co-activator for p53-dependent p21 transcription [31]. Most known human proteins have nucleic acid binding domains [32], so potential lncRNA–protein interactions may act as adaptors to link lncRNA to target loci (Fig. 1a). The telomere complex is a classical model for proteins serving as adaptors between RNA and DNA [33, 34]. The telomere repeat factor TRF2 forms a stable complex with telomere-repeat-encoding RNA (TERRA) and telomere DNA repeats [35]. These interactions could be applied to lncRNA to illustrate archetypes for lncRNA and protein interactions. A well-studied protein that acts as an adaptor between regulatory lncRNA and chromatin is through YY1 tethering lncRNA of Xist to the inactive X nuclear center through repeat C. YY1 protein can bind both RNA and DNA through different sequence motifs and may serve as an adaptor for Xist [36].
LncRNAs scaffolds bring proteins together

LncRNA can be scaffolds to create discrete protein complexes: LncRNA–RNPs (Fig. 1b). HOTAIR could both bind to PRC2 and LSD1 to repress gene transcription and the catalytic methyl-transferase subunit EZH2 of PRC2 is confirmed to be recruited via a structural domain at the 5′-end of HOTAIR to impart repressive histone modifications. Meanwhile the 3′-end of HOTAIR associates with LSD1, inducing H3K4 demethylated modification [37]. Another nascent antisense LncRNA, ANRIL, which is transcribed by RNA polymerase II at the TSS of the p16INK4a gene, recruits PRC2 and PRC1 to mediate protein-coding gene repression in cis [38]. LncRNA roX is transcribed from the Drosophila X genome, which is thought to be critical scaffold for assembly of a functional MSL dosage compensation complex to activate transcription through acetylation of H4K16 [15, 39]. Moreover, LncRNAs act as protein scaffolds to control gene expression by modulation of nuclear architecture. The sub-nuclear structure-specific LncRNAs taurine upregulated

---

**Fig. 1** Schematic illustration of possible molecular archetypes of LncRNA–protein. **a** LncRNA guides protein to target genomic loci or the protein acts as an adaptor for LncRNA to link IncRNA to the target gene. **b** LncRNA brings two or three proteins together to form discrete complexes of LncRNA–RNPs. **c** LncRNA act as decoys to draw proteins away from target genes.
gene 1 (TUG1) and nuclear-enriched autosomal transcript 2 (NEAT2) bind to methylated and unmethylated polycomb 2 protein (Pc2) respectively to mediate assembly of multiple corepressor or coactivator protein complexes [40], which switch non-histone protein methylated mark recognition to relocation of transcription units in the nuclear three-dimensional space, achieving coordinated gene expression regulation.

**LncRNAs act as decoys to titrate away proteins**

LncRNAs act as decoys to remove proteins away from target loci (Fig. 1c). The lncRNA p21 associated ncRNA DNA damage activated (PANDA) is a well-studied lncRNA that acts as a decoy for transcription factors. PANDA is located 5 kb upstream of the CDKN1A with a 5'-cap and a 3'-polyadenylated but non-spliced tail. PANDA binds to and sequestrates NF-YA transcription factor from target gene promoters to repress gene expression [41]. Also, lncRNA could also be decoy of other proteins. LncRNA growth arrest-specific 5 (Gas5) has been identified to interact with glucocorticoid receptor (GR) to prevent binding to DNA response elements, thereby blocking glucocorticoid signal pathway [42]. The lncRNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) binds to and sequesters several serine/arginine splicing factors to modulate their nuclear distribution and phosphorylation states to ensure splicing factor regulation of alternative splicing of cellular pre-mRNAs at a precise time, place, and concentration [24].

**Technologies to probe the lncRNA–protein interactions**

As the significance of lncRNA–protein interactions is better understood, their biochemistry attributes are being discovered and novel bioinformatics approaches are being developed to identify and predict proteins that interact with target lncRNA. Previously, most methods to study RNA–protein interactions are protein-centric, including native RNA immunoprecipitation (nRIP), cross-linking and immunoprecipitation (CLIP) [43, 44]. Recently, the discovery of numerous non-coding RNA has led to concern about RNA-centric approaches, such as the RNA pull-down assay, chromatin isolation by RNA purification (ChiRP), capture hybridization analysis of RNA targets (CHART), and RNA antisense purification (RAP) [45–48]. So other methods are being sought [49, 50], and here we depict detailed strategies for identifying lncRNA–protein interactions.

**Protein-centric approach to probing lncRNA–protein interactions**

**nRIP and nRIP-seq**

nRIP is used to detect the association of individual proteins with specific RNA species in vivo. An antibody-based technology, this can be used to immunoprecipitate an RNA binding protein complex from a heterogeneous cellular homogenate in vivo. RNAs stably associated with the protein complex will immunoprecipitated together and associated RNAs including lncRNA and small RNA or even coding RNA can be measured with real-time PCR or be identified by RNA sequencing [51]. One important consideration prior to commencing RIP-based approaches designed to identify protein–RNA interactions is whether cross-linking is necessary. Native RIP is the method without crosslinking before cell lysis, which is a more appropriate for proteins that bind RNA directly. To reduce RNA recovery rates and identify the proteins that bind RNA indirectly, RIP assay is performed to crosslink cells with formaldehyde before cell lysis step [52]. nRIP is economical, requiring no specialized equipment and can be carried out under native conditions and thus allows identification of kinetically stable interactions [43]. However, the nRIP approach has limitations. First, differentiating direct from indirect interactions is difficult and the read length of associated RNA is too large for identifying the actual binding site. Finally, nRIP assays are commonly known to have variability (reproducibility of 60 % or more) [53]. Thus, multiple biological replicates for nRIP are required. Thus, RIP-based approaches have been used to identify many lncRNA–protein interactions (see Table 1). Zhao’s group discovered that Xist, Tsix, and RepA interacted with PRC2 by native RIP [14]. Recently, Fendrr was identified to bind to PRC2 complex and WDR5 protein by nRIP [29]. Previously, RNA isolated by native RIP was analyzed with microarray technologies (RIP-ChIP) [54], but this is naturally limited by array design and coverage, so nRIP has recently been coupled with high-throughput sequencing (nRIP-seq) to capture the genome-wide RNA pools bound to target proteins [55]. Recent studies suggest that a genome-wide pool of thousands of PRC2-interacting RNAs including some characterized lncRNA, and many unannotated RNAs were identified in embryonic stem cells by nRNA-seq [56].

**CLIP and CLIP-seq**

CLIP is a powerful protein-centric tool used to isolate cross-linked RNA and protein complexes in tissues or cultured cells and subsequently purify the RNA targets. CLIP overcomes the drawbacks of nRIP by cross-linking RNA–protein complexes with ultraviolet light. Based on strong cross-links, followed by RNase treatment of the cell lysate to shorten RNA fragments, immunoprecipitation is performed to purify the covalently cross-linked lncRNA–protein. Importantly, covalent cross-linking permits stringent washing of immunoprecipitates, and this reduces background noise [44]. CLIP has been
used to reveal that many intronic lncRNA directly bind to the PRC2 complex (see Table 1) [64]. For instance, Takashi’s group reported that lncRNA Air interacted with the H3K9 histone methyl-transferase G9a by CLIP [57], but their data suggest that the post-assay RNA quantity is small and the assay is tedious steps. Consequently, CLIP combined with high-throughput sequencing (HITS-CLIP or CLIP-seq) is used to identify many RBPs, such as Nova, Ago2, and TDP-43 [65–67]. The CLIP assay occasionally produces false–positive interactions and it determining the exact binding sites is not always straightforward [68]. Therefore, modified protocols have been developed to define cross-linking events, such as photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP) and individual-nucleotide resolution CLIP (iCLIP) [69, 70]. With PAR-CLIP, cells are cultured in the presence of 4-SU (4-thiouridine) or 6-SG (6-thioguanosine), which are incorporated into RNA and induce strong cross-linking between RNAs and RBPs. Thus, this method can eliminate nonspecific targets and identify exact binding sites at a single nucleotide resolution [71]. A disadvantage of PAR-CLIP is difficulties in the use of 4-SU and 6-SG in living animals due to the toxicity. Thus, the PAR-CLIP method has been successfully applied to RBPs, such as HuR, FMRP, and Ataxin-2 [72–74]. To identify RNA binding motifs and novel functions, iCLIP introduces an adapter at the 5′ end through the primer used for reverse transcription by cDNA circularization and subsequent linearization. Thus, both truncated and read-through cDNAs are captured. Importantly, iCLIP also provides information about the cross-link site that permits precise mapping of RNA–protein contacts at a nucleotide resolution [70]. Rossbach’s group performed iCLIP combined with deep-sequencing (iCLIP-Seq) to reveal global regulatory roles of hnRNPII L protein [75].

RNA-centric approach to dissecting lncRNA–protein interactions

RNA pull-down assay

RNA pull-down assay is a preliminary RNA-centric in vitro method that enabling identification and characterization of various proteins which interact with a given lncRNA of interest. First, lncRNA probes were synthesized and labeled with high affinity tags, such as biotin, then cell lysate was prepared from a in vitro sample. Next, the lncRNA probe was incubated with lysate or recombinant protein to form a specific lncRNA–protein complex. Subsequently, the protein complex was pulled down with streptavidin agarose or magnetic beads. Finally, the retrieved protein was identified with Western blot or mass spectrometry (MS) [45]. Rinn’s group used RNA pull-down to discover that HOTAIR was directly associated with the PCR2 complex, which repressed transcription of the HOXD loci in trans [19].

ChIRP and ChIRP-MS

Advances in RNA-centric biochemical purification offer new opportunities for systematically mapping lncRNA interactions with proteins and chromatin. The ChIRP method is based on using the biotinylated oligonucleotides complementary to the lncRNA of interest as a “handle” to pull down lncRNA-associated proteins and chromatin DNA. In brief, cultured cells are cross-linked, chromatin is extracted and sonicated, and then biotinylated tilling oligos that tile whole lncRNA were added and hybridized. Finally, the hybrids including target lncRNA, proteins and chromatin DNA were eluted with
magnetic streptavidin beads and subjected to q-PCR or deep sequencing for DNA analysis, or to Western blotting or MS for protein analysis (Fig. 2) [46]. In addition, the recently developed ChIRP-seq method allows global and high-throughput discovery of genomic DNA associated with IncRNA. Chu’s group used ChIRP-seq to reveal the precision genomic occupancy of roX2, TERC, and HOTAIR: three rather different IncRNAs in two species successfully (see Table 1) [47]. To dissect the IncRNA functional domains in situ, domain-specific chromatin isolation by RNA purification (dChIRP) has been developed based on the ChIRP method and with dChIRP, biotinylated oligos are used as specific pools to recover specific domains of target RNA that improve the RNA genomic localization signal-to-noise ratio by >20-fold over traditional ChIRP [61]. dChIRP allowed IncRNA to be characterized at the domain level and revealed the IncRNA architecture and function with high precision and sensitivity [76]. Recently, Quinn’s group reported a ‘three-fingered hand’ topology of roX1 and that the three D domains of roX1 bind directly to the MSL protein complex to individually rescue male lethality by dChIRP [77].

MS based proteomics is a common tool for studying cellular interactions. Baltz and colleagues and Castello’s group used MS to identify hundreds of novel RBPs in human cells [78, 79]. Recently, to enable quantitation and accurate discovery of novel RNA–protein interactions from complexes assembled in vivo, Klass and colleagues used quantitative MS combined with RNase treatment of affinity-purified RNA–protein complexes to identify proteins that bind to RNA concurrently with an RBP of interest [80]. Kramer’s group developed an experimental and computational workflow method combining photo-induced cross-linking, high-resolution MS and automated analysis of the resulting spectra for identification of RNA interactions with proteins [81]. This MS-based workflow based on MS can be applied to map any RNA–protein complex of interest. Recently, ChIRP-MS, an optimized ChIRP method for systematically discovering IncRNA-bound proteome in vivo proteomes by MS was developed to identify 81 endogenous proteins that associated with Xist in two waves to coordinate X chromosome spreading and silencing. Interestingly, HrnpK protein participates in Xist-mediated gene silencing and chromatin modifications, but not Xist biogenesis or localization. Thus, the results suggested ChIRP-MS assay achieved high output and specificity regarding IncRNA–protein interactions in vivo [62].

**CHART and CHART-MS**

Another hybridization-based purification strategy is CHART which is used to confirmed the genome-wide localization of IncRNA in chromatin and isolate the protein associated with the IncRNA of interest. CHART is more similar than different with ChIRP (Fig. 2), but one significant difference is the design criteria of the oligonucleotide probe. With ChIRP, short antisense DNA oligonucleotides tile across the entire target IncRNA without a priori knowledge of target RNA function domain cover all potential hybridization spots [47]. In contrast, probes of CHART are empirically determined after RNase H assay which determines the candidate hybridization region [82]. CHART is a useful method for biochemically defining DNA and proteins associated with IncRNAs. CHART-seq, which combines CHART and RNA-seq, was applied to discover hundreds of trans-genomic binding sites for NEAT1 and MALAT1 [60]. Moreover, West’s group initially adapted the CHART assay to identify the full complement of proteins associated with RNAs in vivo with MS. CHART-MS was performed for two human IncRNAs, NEAT1 and MALAT1, to identify many nuclear speckle and para-speckle components and several new proteins not previously associated with them (see Table 1) [60].

**RAP and RAP-MS**

Similar to ChIRP and CHART, the RAP method is also used to capture a target IncRNA of interest through hybridization with antisense biotinylated oligos (Fig. 2) [47]. With RAP, various cross-linking conditions can be performed to identify different molecules that interact with the target RNA via different mechanisms. For direct RNA–RNA interactions, psoralens are used for crosslinking, but for protein–RNA interactions, formaldehyde or ultraviolet (UV) light is applied to crosslink. Compared to ChIRP and CHART, the most distinctive feature of RAP is its use of long capture probes (>60 nucleotides), which form very stable RNA-DNA hybrids [7]. Such a probe design strategy robustly captures any RNA and enables the use of stringent hybridization and washing conditions that dramatically reduce nonspecific interactions of off-target nucleic acids or proteins. Long DNA probes are considerably more costly but the background signals may be reduced due to the fewer probes used compared with short probes [48]. Hacisuleyman’s group applied RAP to discover the genomics sites and proteins that associated with lincRNA FIRE. And confirmed that it interacts with hnPNPU protein in an RRD-dependent manner and localizes across several trans-chromosomal binding sites (see Table 1) [58]. To developed a high-throughput method to identify proteins associated with a specific IncRNA in vivo, McHugh and colleagues combined RAP with MS to obtain high yields of RNA complex and identified ten proteins associated with IncRNA Xist, including SHARP, RBM15, MYEF2, CELF1, HNRNPC,
Fig. 2 Schematic representation of ChIRP, CHART and RAP to identify associated proteins and chromatin DNA. RNA–protein–DNA complexes were cross-linked in vivo and solubilized by sonication. Corresponding biotinylated oligonucleotides of three methods were designed and synthesized to be hybridized to target lncRNAs under stringent conditions. Associated proteins and chromatin DNA were efficiently pulled down with streptavidin magnetic beads. Co-purified RNA, protein and DNA were isolated with RNase elution and subjected analyzed downstream analysis. Isolated proteins were analyzed by mass spectrometry or Western blotting and chromatin DNA was used for q-PCR or deep sequencing.
LBR, SAF-A, RALY, HNRNPM, and PTBP1 (see Table 1). Also, they reported that the Xist interacts directly with SHARP to silence transcription through HDAC3 and that the recruitment of PCR2 by Xist depended on SHARP and HDAC3. These data is contrasted with previous work indicating that Xist directly interacted with PCR2 across the X chromosome [63]. Thus, the RAP-MS can be useful for investigating lncRNA regulation mechanism.

Bioinformatics approach to predicting lncRNA–protein interactions
Biochemical approaches to identify the lncRNA–protein complexes are constantly expanding along with computational technologies. Compared with biochemical assays, the bioinformatics is more convenient and rapid for large-scale predictions of protein–lncRNA associations. Tartaglia’s group developed the algorithm, ‘fast predictions of RNA and protein interactions and domains at the Center for Genomic Regulation, Barcelona, Catalonia’ (catRAPID), which evaluates interaction propensities of polypeptide and nucleotide chains using their physicochemical properties [49]. catRAPID was used to predict RNA–protein interactions in neurodegenerative disorders, in which RNA-binding proteins apparently have a major role [83]. Recently, this method was used to predict protein interactions in the Xist regulatory network [84], and data show that catRAPID is powerful for predicting RNA–protein interactions from sequences. However, prediction of lncRNAs function is generally hampered by poor sequence homology and lack of interaction data. Consequently, Lu and colleagues developed a new computational method, IncPro, to predict lncRNA–protein interactions. Compared to CatRAPID, IncPro is computational-friendly and does not lead to nonsensical cross terms. Applying IncPro to all human proteins, this laboratory reported that long non-coding RNAs tend to interact with nuclear and RNA-binding proteins [50]. However, this technique is limited for finding the direct lncRNA–protein interaction s due to the volume of proteins. Recently, to gain insights into global relationships between lncRNAs and their binding proteins, Shang and colleagues constructed an lncRNA–protein network (LPN) including 177 lncRNAs, 92 proteins and confirmed 683 relationships between them, based on experimentally determined functional interactions [85]. Therefore, bioinformatics approaches to predicting lncRNA–protein interactions may guide future experimental approaches and facilitate a deeper understanding of the role of lncRNAs.

Conclusions and perspectives
Given the multitude of non-coding transcripts discovered by second-generation deep sequencing, lncRNAs arouse interest to biological and bio-medical researchers. Recently, evidence has accumulated to support the idea that lncRNAs are critical to numerous biological processes, whereas the mechanisms by which lncRNA are poorly understood. Protein, an important partner for RNA in vivo, has been associated with molecular archetypes of lncRNA and we observed scaffolds, guides and decoys in these associations. However, some lncRNA interact with protein through more than one kind molecular mechanism. For example, HOTAIR is a scaffold for PCR2 and LSD1 as well as a guide to recruit PCR2 to target loci. Therefore molecular mechanisms behind lncRNA–protein interactions are complicated and rarely described. Study of lncRNAs interaction partners and the use of technologies to isolate and identify molecules associated with lncRNA are assisting researchers with the study of proteins and genomic DNA that directly and indirectly interplay with target lncRNAs. However, these interactions have not been studied across diverse species. As technologies improve, we may 1 day better understand evolution and functional mechanisms of lncRNAs.

Abbreviations
lncRNA: long non-coding RNA; RBP: RNA binding protein; RIP: RNA immunoprecipitation; PCR2: ploidycomb repressive complex 2; HOTTIP: HOX transcript at the distal tip; HOTAIR: HOX antisense intergenic RNA; BC1: brain cytoplasmic RNA 1; Fendr: fetal-lethal non-coding developmental regulatory RNA; COLDAR: cold assisted intronic non-coding RNA; TUG1: taurine upregulated gene 1; NEAT2: nuclear-enriched autosomal transcript 2; Pcp2: ploidycomb 2 protein; PANDA: p21 associated ncRNA DNA damage activated; Gas5: growth arrest-specific 5; MALAT1: metastasis-associated lung adenocarcinoma transcript 1; CLIP: cross-linking and immunoprecipitation; ChIRP: chromatin isolation by RNA purification; CHART: capture hybridization analysis of RNA targets; RAP: RNA antisense purification; PAR-CLIP: photoactivatable-ribonucleoside-enhanced CLIP; iCLIP: individual-nucleotide resolution CLIP; HITS-CLIP: high throughputssequencing of CLIP; MS: mass spectrometry; LPN: lncRNA–protein network.

Authors’ contributions
HZ and YY planned the manuscript outline and wrote the draft. YY generated figures and tables. LW and HZ revised and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We thank Chen-Guang Zhou and Jing Shi for stimulating discussions and critical review of the manuscript.

Competing interests
The authors declared that they have no competing interests.

Received: 18 June 2015 Accepted: 8 October 2015
Published online: 22 October 2015

References
1. Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet. 2005;21:93–102.
2. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by
RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
3. Djebali S, Davis CA, Merkel A, Debain A, Lassmann T, Mortazavi A, Tanzer A, Lagero J, Lin W, Schlesinger F. Landscape of transcription in human cells. Nature. 2012;489:101–8.
4. Novkova IV, Hennelly SP, Sanbonmatsu K. Zing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? Bioarchitecture. 2012;189–99.
5. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Gene Dev. 2009;23:1494–508.
6. Zhao J, Yang Y, Li R, Fu D, Wen L, Luo Y, Zhu H. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. J Exp Bot. 2015;66:4848–53.
7. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG. The GENCODE v7 catalog of human long non-coding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775–89.
8. Wang S, Tran E. Unexpected functions of IncRNAs in gene regulation. Commun Integr Biol. 2013; 6.
9. Yoon J, Abdelmohsen K, Gorospe M. Functional interactions among microRNAs and long noncoding RNAs. Seminars in cell and developmental biology. Elsevier. 2014; v. 9–14.
10. Chen C, Spitali RC, Chang HY. Technologies to probe functions and mechanisms of long noncoding RNAs. Nat Struct Mol Biol. 2015;22:29–35.
11. Rinn JL, Uriel O. IncRNAs in RNA–protein interactions. Genome Biol. 2014;15:401.
12. Buenrostro JD, Ayara CL, Chirico LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ. Quantitative analysis of RNA–protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nat Biotechnol. 2014;32:562–8.
13. Tome JM, Ozer A, Pagano JW, Gheba D, Schroth GP, Lis JT. Comprehensive analysis of RNA–protein interactions by high-throughput sequencing–RNA affinity profiling. Nat Methods. 2011;8:1683–8.
14. Wang S, Tran E. Unexpected functions of IncRNAs in gene regulation. Commun Integr Biol. 2013; 6.
15. Lee JT. Epigenetic regulation by long noncoding RNAs. Science. 2013;338:1435–9.
16. Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Gene Dev. 2009;23:1494–508.
17. Yoon J, Abdelmohsen K, Gorospe M. Functional interactions among microRNAs and long noncoding RNAs. Seminars in cell and developmental biology. Elsevier. 2014; v. 9–14.
18. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-2 noncoding RNA regulates alternative splicing by modulating SR splicing factor phospho-status. Genes Dev. 2006;20:1470–84.
19. Bilaud T, Brun C, Ancelin K, Koering CE, Laroche T, Gilson E. Telomeric localization of TRF2, a novel human telobox protein. Nat Genet. 1997;17:236–8.
20. Jeon Y, Lee JT. Y11 tethers Xist RNA to the inactive X nucleation center. Cell. 2011;146:119–33.
21. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 2011;43:689–93.
22. Yang L, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y. Role of histone H2A ubiquitination in Polycomb silencing. Nature. 2004;431:873–8.
23. Wang Y, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-2 noncoding RNA regulates alternative splicing by modulating SR splicing factor phospho-status. Genes Dev. 2006;20:1470–84.
24. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA Gas5 is a receptor. Sci Signal. 2010;3:a8.
25. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 2011;43:689–93.
26. Lu Q, Ren S, Lu M, Zhang Y, Zhu D, Zhang X, Li T. Computational prediction of associations between long non-coding RNAs and proteins. BMC Genom. 2013;14:651.
51. Selth LA, Close P, Svejstrup JQ. Studying RNA–protein interactions in vitro by RNA immunoprecipitation. Epigenetics protocols. Springer. 2011. p. 253–64.

52. Niranjankumari S, Lasda E, Brazas R, Garcia-Blanco MA. Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions in vivo. Methods. 2002;26:182–90.

53. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Morales DR, Thomas K, Presser A, Bernstein BE, van Oudenaarden A. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci. 2009;106:11667–72.

54. Keene JD, Komisarow JM, Friedensood MB. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat Protoc Electron Ed. 2006;1:302.

55. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Stepete AL, Wani S, Bethel G. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods. 2008;5:613–9.

56. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell. 2010;40:939–53.

57. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science. 2008;322:1717–20.

58. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR. Topological organization of multichromosomal regions by the long intergenic non-coding RNA Firre. Nat Struct Mol Biol. 2014;21:198–206.

59. Wang KC, YangYW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472:120–4.

60. 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 to coordinate homeotic gene expression. Nat Cell Biol. 2014;15:1488–97.

61. Russo A, Blume JE, Wang X. HITS-CLIP yields genome-wide insights into brain protein–RNA interaction sites in living cells. Methods. 2005;37:376–86.

62. McHugh CA, Chen C, Chow A, Surka CF, Tran C. The Xist long non-coding RNA interacts with H3K4me1 and affects Xist RNA binding of Malat1 bind active chromatin sites. Mol Cell. 2014;55:791–802.

63. Quin JJ, Chang HY. In situ dissection of RNA functional subunits by domain-specific chromatin isolation by RNA purification (dChIRP). Nuclear bodies and noncoding RNAs. Springer. 2015. p. 199–213.

64. Chou C, Zhang QC, Da Rocha ST, Flynn RA, Bharendra M, Calabrese JM, Magnuson T, Heard E, Chang HY. Systematic discovery of Xist RNA binding proteins. Cell. 2015;161:404–16.

65. McHugh CA, Chen C, Chow A, Surka CF, Tran C. The Xist IncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature. 2015;554:791–802.

66. Li Q, Uemura Y, Kawahara Y. Cross-linking and immunoprecipitation of nuclear RNA-binding proteins. nuclear bodies and noncoding RNAs. Springer. 2015. p. 253–67.

67. Yang et al. Cell Biosci (2015) 5:59

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit