Long noncoding RNA and protein abundance in IncRNPs

MAN WU,1 LIANG-ZHONG YANG,1 and LING-LING CHEN1,2,3
1State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
2School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
3School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

Editor’s Note:
Ling-Ling Chen was the recipient of the 2021 Mid-Career Research Award presented by the RNA Society.

ABSTRACT
Although long noncoding RNAs (lncRNAs) are generally expressed at low levels, emerging evidence has revealed that many play important roles in gene regulation by a variety of mechanisms as they engage with proteins. Given that the abundance of proteins often greatly exceeds that of their interacting lncRNAs, quantification of the relative abundance, or even the exact stoichiometry in some cases, within lncRNA–protein complexes is helpful for understanding of the mechanism(s) of action of lncRNAs. We discuss methods used to examine lncRNA and protein expression at the single cell, subcellular, and suborganellar levels, the average and local lncRNA concentration in cells, as well as how lncRNAs can modulate the functions of their interacting proteins even at a low stoichiometric concentration.

Keywords: lncRNA; IncRNP; concentration; abundance; stoichiometry; single cell; subcellular; suborganellar

INTRODUCTION
The vast majority of the human transcriptome consists of non–protein coding RNAs (ncRNAs), among which those longer than 200 nt are defined as long noncoding RNAs (lncRNAs) (Derrien et al. 2012). The latest annotation suggests that human lncRNAs exceed 170,000 (Zhao et al. 2021), although in a cell-/tissue-specific expression manner (Kim et al. 2015; Seifuddin et al. 2020). Currently annotated lncRNAs comprise species that are derived from different genomic origins and are processed by distinct mechanisms (for reviews, see St Laurent et al. 2015; Wu et al. 2017; Kopp and Mendell 2018). Over the past decade, well-documented studies focused on IncRNA modes of action have shown that these large and single-stranded molecules are often functionally engaged with distinct sets of proteins (Statello et al. 2021). They can act as guides, recruiters, scaffolds and decoys to modulate the structure and function of chromosome architecture and nuclear condensates, to regulate transcriptional and post-transcriptional processes, or to directly interfere with mRNA translation, post-translational modifications and signaling pathways. Thus, lncRNAs can trigger significant cellular responses in many biological contexts.

Given their large size and flexible conformation, it remains a challenge to fully understand how lncRNAs exert their effects. In particular, the abundance of RNA binding proteins (RBPs) often greatly exceeds that of interacting lncRNAs, except in a few cases, such as the human telomerase reverse transcriptase (hTERT) and the telomerase RNA (hTR) that form an IncRNP at about a 1:1 ratio (Xi and Cech 2014; Ghanim et al. 2021). Quantitative comparison of lncRNA and protein levels shows that the copy number of the majority of proteins per cell varies from 10,000 to 80,000,000 (Hein et al. 2015), remarkably exceeding that of lncRNAs, with expression of 0.3–1000 molecules per cell (Table 1). Further, most proteins associate with many other factors, not just lncRNAs to which they can bind. Thus, distinct mechanisms have been adapted by lncRNAs to overcome their low abundance compared to their interacting proteins for achieving measurable effects. Understanding the subcellular localization of lncRNAs and the local concentrations of both proteins and RNAs is...
| lncRNA          | Length (nt) | Localization                      | Copies of lncRNA (per cell) | Interacting protein(s)     | Copies of interacting proteins per cell | Function and mode of action                                                                 | Key references                                                                 |
|-----------------|-------------|-----------------------------------|-----------------------------|---------------------------|----------------------------------------|--------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| hTR             | 451         | Nucleoplasm and telomeres         | 1150 (HeLa)                 | hTERT                     | 500 (per cell)                         | An essential telomerase component required for telomere homeostasis                         | (Xi and Cech 2014)                                                              |
| HOTTIP          | 3764        | Nucleus (chromatin)               | ~0.3 (foreskin fibroblast)  | WDR5/MLL                  | N/A                                    | Recruits WDR5/MLL for H3K4me3 across HOXA to enhance transcription with an AML-like phenotype in Hottip transgenic mice | (Wang et al. 2011; Cheng et al. 2015; Luo et al. 2019) |
| lincRNA-EPS     | 2531        | Nucleus (chromatin)               | 11 (BMDM)                   | hnRNP L                   | N/A                                    | Suppresses immune gene transcription by controlling nucleosome positioning                  | (Atianand et al. 2016; Mumbach et al. 2019)                                          |
| Xist            | 17,000      | Nucleus (inactivated X chromosome) | ~100 (MEF)                  | RBM15, PCGF5, SPEN, CELF1, and others | N/A                                    | Essential for X-chromosome inactivation in female mammals                                    | (Brown et al. 1992; Hartshorn et al. 2003; McHugh et al. 2015; Sunwoo et al. 2015; Markaki et al. 2020) |
| PNCTR           | >10,000     | Nucleus (perinucleolar compartment, PNC) | ~36 (HeLa)                  | PTBP1                     | ~286,000 (per cell)                    | Modulates cancer cell proliferation by sequestering the splicing factor PTBP1                | (Yap et al. 2018)                                                              |
| NEAT1_2         | 22,700      | Nucleus (paraspeckles)            | 53 (per paraspeckle in HeLa)| NONO, SFPO, FUS, RBM14   | ~2000 NONO per paraspeckle             | Nucleates paraspeckle formation                                                              | (Chujo et al. 2017; Yamazaki et al. 2018)                                           |
| LETN            | 4500        | Nucleolus (granular component, GC) | ~200 (HUH7 and HCC827)      | NPM1                      | 8,800,000 (per cell)                   | Maintains cancer cell proliferation by modulating nucleolar function                          | (Wang et al. 2021)                                                             |
| SLERT           | 694         | Nucleolus (dense fibrillar component, DFC) | 826 (PA1)                   | DDX21                     | 13,000 (per DFC), 78,000 (per nucleolus), and 1,040,000 (per cell) | Controls FC/DFC phase separation and promotes Pol I transcription                           | (Xing et al. 2017; Wu et al. 2021)                                                |
| circFAM120A     | 556         | Cytoplasm (monoribosomes)         | 22–25 (HT29, 293FT and HT29) | IGF2BP2                   | ~136,000 (per cytoplasm) and ~200,000 (per cell) | Promotes cell proliferation by outcompeting FAM120A mRNA binding to IGF2BP2                  | (Li et al. 2021)                                                               |
| FAST            | 547         | Cytoplasm                         | 140 (H9)                    | β-TrCP                    | 3269 (per cytoplasm) and 8173 (per cell) | Maintains hESC pluripotency by activating WNT pathway                                        | (Guo et al. 2020)                                                              |
| IncRNA-ACOD1    | Variants: 2330 and 2259 | Cytoplasm                         | ~100 (BMDM)                 | GOT2                      | N/A                                    | Facilitates viral replication by enhancing glutamic-oxaloacetic transaminase GOT2 activity | (Wang et al. 2017)                                                             |
| NORAD           | 5300        | Cytoplasm (NORAD-PUM bodies)      | 400 (HCT116)                | PUMILLO (PUM1/PUM2)       | 15,000 PUM1/2,000 PUM2 (per cell)      | Sequesters PUM to maintain genomic stability                                                 | (Lee et al. 2016; Elguindy and Mendell 2021)                                        |
|                |             | Cytoplasm                         | 70 (U2OS)                   | PUMILLO (PUM1/PUM2)       | 200 PUM1/550 PUM2 (per cell)            |                                                                                              | (Tichon et al. 2016)                                                            |
therefore more informative mechanistically than only knowing average cellular concentrations.

As one example, for cis-acting lncRNAs, an abundance of one to 10 molecules per cell is likely sufficient for functioning at a single locus or at several loci by directly base-pairing with the genomic DNA or by recruiting chromatin modifiers to impact transcriptional outputs. This would explain how a number of low-abundance lncRNAs have been shown to exert pronounced biological effects (Table 1). For trans-acting lncRNAs, a higher abundance (10–1000) has been reported for their potential roles as recruiters, scaffolds, or decoys of proteins. Of note, such a higher abundance is still far below that of their interacting proteins; thus, the exact working model of a particular lncRNA depends not only on how many RBP binding motifs or structural modules exist within the lncRNA but also on its subcellular distribution (Table 1). In addition to relative abundance, it is worthwhile noting that the exact stoichiometry of lncRNA and protein within a functional lncRNP can provide clues into detailed lncRNA working mechanisms, as illustrated by well-studied lncRNPs, such as NORAD interacting with PUMILIO (PUM) (Elgindy and Mendell 2021), and SLERT interacting with DDX21 (Wu et al. 2021). In this review, we summarize methods used to quantify lncRNA and interacting RBPs at the cellular, subcellular and suborganellar levels. We also discuss how cells deal with the generally low lncRNA–protein concentration ratios through a variety of regulatory modes of action.

STOICHIOMETRY OF CLASSICAL lncRNPs

Several classical cellular ribonucleoprotein complexes (RNPs) have now been well studied such as the catalytic ribonucleoprotein complex RNase MRP involved in precursor ribosomal RNA (Lan et al. 2020), the RNase P complex that selectively cleaves transfer RNA-like substrates (Guerrier-Takada et al. 1983; Bartkiewicz et al. 1989; Evans et al. 2006) and the telomerase complex that acts in lengthening telomeres during DNA replication (Kim et al. 1994). For each of these functional RNPs a defined stoichiometry of RNAs and proteins is required for their assembly (Gray and Gopalakrishnan 2020). A useful example is the human telomerase that is minimally composed of a catalytic hTERT and an RNA component, hTR. hTR is a 451 nt lncRNA acting as the template for telomere elongation by associating 1:1 with hTERT that is capable of de novo DNA synthesis (Schmidt and Cech 2015). Stoichiometric analyses by spike-in qRT-PCR and northern blotting (NB) experiments showed ∼1150 hTR per HeLa cell, and by a two-step immunoprecipitation (IP)–western blotting (WB) method showed ∼500 hTR per HeLa cell (Xi and Cech 2014). Interestingly, quantification of endogenous levels of telomerase revealed ∼240 active telomerase monomers or ∼120 dimers, a number comparable to that of telomeres in late S phase per HeLa cell (Cohen et al. 2007; Xi and Cech 2014). These analyses suggested that telomerase was not an abundant RNP and that unassembled nontelomerase hTR or hTERT existed in cells. Overexpression of either hTR or hTERT led to increased telomerase activity, supporting the notion of a quick assembly of the subpopulations of nontelomeric hTERT and hTR into functional telomerases in cells (Xi and Cech 2014). Methodologically, the nonequivalence of copy number between telomerase and its components indicates that only measuring the total cellular copy number of a lncRNA of interest and its corresponding protein may be inadequate to indicate the relevant abundance of functional lncRNPs.

METHODS TO DETECT THE ABUNDANCE OF lncRNAs AND THEIR ASSOCIATED RBPs

Different from hTR and hTERT in telomerase, most lncRNA-binding proteins also bind to many other RNAs in cells; thus, it is important to detect the relative abundance of lncRNAs and their associated RBPs within the particular lncRNPs for functional insights. Also, unlike telomerase, which can be quantified by telomerase activity, we lack biochemical assays to detect most lncRNPs. Methods to quantify lncRNA expression have therefore moved from classic biochemical quantification (Xi and Cech 2014; Lee et al. 2016; Xing et al. 2017) to combining subcellular fractionation and quantitative imaging (Guo et al. 2020; Li et al. 2021), and more recently, high resolution microscopy at the suborganelle level to dissect the relative abundance of lncRNA and protein in lncRNPs (Wu et al. 2016; Yao et al. 2019; Elgindy and Mendell 2021; Wu et al. 2021).

BIOCHEMICAL QUANTIFICATION OF lncRNAs AND RBPs IN CELLS

Biochemical quantification is the most well-established method to calculate the average lncRNA copy number per cell. To do so, one needs to construct an in vitro benchmark system, termed a “standard curve,” which can be generated by using NB or qRT-PCR (Xi and Cech 2014; Xing et al. 2017; Li et al. 2021; Wu et al. 2021) analyses with a series of dilutions of purified in vitro transcribed (IVT) lncRNAs of interest (Fig. 1A). Then, the total RNA extracted from a defined number of cells (i.e., 1 million cells) is collected and subjected to NB followed by band quantification or by qRT-PCR to obtain the respective Cć values. Based on the standard curve (Fig. 1A), the copy number of a particular lncRNA per cell can be determined by dividing the total number of cells used to extract the total RNA.

One must take two matters into consideration during this biochemical quantification protocol. First, for RNAs that are strongly encapsulated with proteins, chromatin,
nuclear matrix or in dense condensates in cells, it is compulsory to utilize an improved RNA extraction method by shearing or heating to disrupt such dense complexes, thereby leading to enhanced extractability of RNA by TRIzol (Chujo et al. 2017; Wu et al. 2021). Second, we highly recommend the use of IVT RNAs as the template for standard curves. Using IVT RNAs as the template for standard curves involves a cDNA synthesis step and thus takes the efficiency of the reverse transcription reaction into account. Furthermore, knowing a copy number of a particular RNA per cell, by combining the RNA-seq transcriptomic data, one can roughly get an idea of the copy number of additional RNAs by comparing their expression in samples with the same sequencing depth (Liu et al. 2019).

In principle, in these biochemical quantifications, specificities of NB probes and qRT-PCR primers are of importance to be considered to avoid potential bias in calculating concentrations (Li et al. 2021). For instance, using different qRT-PCR primer sets targeting hTR gave noteworthy differences in hTR levels, even as much as 10-fold variation. To solve this concern, gradient dilutions of in vitro purified hTR as spike-in were incorporated into the endogenous hTR to obtain qRT-PCR signals for abundance (Xi and Cech 2014).

Similarly, the copy number of a protein of interest per cell can be calculated, by first generating the standard curve as a reference concentration with a series dilution of purified proteins by WB and then calculating the protein copy number from a defined number of cells (Li et al. 2021). Of note, the accuracy of antibodies for WB should be considered to avoid potential misdetection. On the other hand, though, mass spectrometry (MS)-based quantitative proteomics provides a sensitive way to analyze proteins on a global scale across cell and tissue type (Aebersold and Mann 2003; Schubert et al. 2017). Generally, there are two quantitative MS strategies, isotopic labeling (Gygi et al. 1999; Ong et al. 2002; Ross et al. 2004) and label-free approaches (Chelius and Bondarenko 2002; Ahrme et al. 2013). Among isotopic labeling methods, SILAC (stable isotope labeling with amino acids in cell cultures) (Ong et al. 2002; Wang et al. 2018a) is the most popular strategy in which specific amino acids with a nonradioactive, isotopically labeled form are
incorporated into all proteins in given samples. For label-free absolute quantification, all proteins are digested into peptides and subjected to liquid chromatography to measure the summarized ion counts; then these counts are converted into signal values compared to the total amount of proteins injected into the mass spectrometer, or correlated to a set of spike-in reference peptides with known concentrations (Chelius and Bondarenko 2002; Ludwig et al. 2012). Although large scale quantification is allowed in these measurements, a huge number of cells is required to gain confidence in quantitative MS, because the readout is limited as a proxy for protein levels. Consequently, single-cell proteomics, termed scMS, has major limitations on throughput and proteomic depth (Budnik et al. 2018; Zhu et al. 2018). Nevertheless, in principle, the copy number of a protein of interest can be extracted for particular types of cells that have been analyzed by quantitative MS (Castello et al. 2012, 2016). This can then be used to determine the levels of RBPs containing intrinsically disordered regions (IDR) in RNA interactomes that have emerging roles in phase separation (Das et al. 2014; Zhang et al. 2015; Maharana et al. 2018; Zhou et al. 2018; Guillen-Boixet et al. 2020).

**BIOCHEMICAL QUANTIFICATION OF lncRNAs AND RBPs AT THE SUBCELLULAR LEVEL**

lncRNAs often possess subcellular preferences for function instead of being evenly distributed throughout the cell (Cabili et al. 2015; Wu et al. 2016; Sun et al. 2018). More than the copy number of a lncRNA per cell, its specialized localization indeed more profoundly impacts the function of its interacting proteins. One simple way used to determine subcellular localization involves separating cytoplasmic and nuclear RNAs followed by NB and qRT-PCR quantification (Fig. 1B). In brief, cells are treated with nonionic detergent, such as Igepal, to solubilize membrane lipids and release cytoplasmic contents. After centrifugation at low speed (i.e., 1000 g), the soluble and cytoplasmic RNAs in the supernatant are extracted by TRIzol or other RNA extraction reagents (Chen et al. 2008; Lo Piccolo et al. 2015; DeCaprio and Kohl 2020). The pelleted nuclear fractions are further incubated with prolonged nonionic detergent or gentle ionic detergent, such as deoxycholate, to destroy the nuclear membrane and release nuclear soluble and insoluble fractions. The latter insoluble fractions mainly consist of chromatin-associated RNAs (Ramsby and Makowski 2011) and nuclear condensate-enriched RNAs (Chujo et al. 2017). In theory, the nuclear insoluble fraction is resistant to high salt, nonionic detergent and nuclease treatment; thus, special attention should be paid to isolate lncRNAs from the insoluble fraction using appropriate buffers (i.e., by including high salt, 2 M NaCl) (Takata et al. 2009).

In the cytoplasm, some lncRNAs are specifically localized to certain organelles for function. For example, the lncRNA SAMMON is mitochondrial localized and plays a role in promoting melanoma growth and survival via binding mitochondrial p32 protein (Leucci et al. 2016). Some lncRNAs are enriched in ribosomal fractions and affect mRNA translation (Essers et al. 2015; Hansji et al. 2016). For instance, ZFAS1 binds to the small 40S ribosomal subunit and regulates ribosome production and assembly (Hansji et al. 2016). A recent study showed that circFAM120A bound the translational inhibitor IGF2BF2 to prevent this protein from engaging with FAM120A mRNA on monoribosomes, resulting in enhanced FAM120A mRNA translation (Li et al. 2021). It should be noted that FAM120A mRNA is expressed at 250–600 copies per cell, while circFAM120A is expressed at only 22 copies per cell. Interestingly, to be competitively competent for IGF2BF2 binding, the majority of circFAM120A, but just a small fraction (<10%) of FAM120A mRNA, is specifically present on monoribosomes, resulting in a 3:1 to 5:1 of FAM120A mRNA and circFAM120A stoichiometry on the monoribosomes, as shown by polyribosome profiling assays that segregate single ribosomes and polyribosomes via sucrose gradients (Li et al. 2021). Such specific localization of circFAM120A provides a sufficient level to trap IGF2BF2 on monoribosomes, thereby facilitating FAM120A mRNA translation and ultimately cell proliferation (Li et al. 2021).

Inside the nucleus, isolating soluble and insoluble nuclear fractions can roughly indicate a subnuclear localization of lncRNAs, but it remains a challenge to delineate their precise subnuclear localization patterns using biochemical assays. Different types of membrane-less nuclear bodies are difficult to isolate biochemically. The largest nuclear structures, nucleoli (1–5 μm [Lafontaine et al. 2021]), are more amenable to be isolated from a nuclear fraction while other smaller nuclear bodies are far more difficult to purify using sucrose cushions. Isolation of nucleoli could be achieved by treating sonicated nuclear fractions with various concentrations of sucrose and magnesium chloride (Andersen et al. 2005), which allowed investigators to uncover the role of Alu-containing RNAs in the maintenance of nucleolar structure and function (Caudron-Herger et al. 2015). Paraspeckles (0.2–0.5 μm [Sasaki and Hirose 2009; Mao et al. 2011b]) are difficult to isolate biochemically and were discovered by immunofluorescence (IF) imaging validation of proteomic results of the sucrose density—isolated nucleolar components including PSP1 and NONO (Fox et al. 2002). IF with anti-PSP1 and anti-NONO revealed that they did not localize to nucleoli but rather to previously undefined regions that are localized adjacent to nuclear speckles (Fox et al. 2002, 2005).

Beyond the above-mentioned traditional biochemical analyses, recently developed proximity labeling-based methods including BioID (biotin identification) (Choi-Rhee et al. 2004; Roux et al. 2012), APEX (engineered
ascorbate peroxidase) (Martell et al. 2012; Hung et al. 2016), HRP (horseradish peroxidase) (Hopkins et al. 2000; Li et al. 2014), and PUP-IT (pupylation-based interaction tagging) (Liu et al. 2018), coupled with quantitative MS offer a high-throughput means for systematic analysis of spatially restricted proteomes (Bosch et al. 2021). In addition, RAP-MS (RNA antisense purification coupled with mass spectrometry) revealed Xist interacting proteins (McHugh et al. 2015), IP-MS identified thousands of proteins across almost all protein classes in HeLa cells (Hein et al. 2015) and RNA CHART (capture hybridization analysis of RNA targets) -seq using antisense oligonucleotides for NEAT1_2 revealed novel RNA components of paraspeckles (West et al. 2016) as well as the crosstalk between paraspeckles and mitochondria (Wang et al. 2018b). Moreover, DNA CHART-seq and CHART-MS of NEAT1 and MALAT1 have also been used to show the shared DNA and protein components between paraspeckles and nuclear speckles (West et al. 2014).

**IMAGING AND SUPER RESOLUTION IMAGING QUANTIFICATION OF IncRNPs AT THE SUBCELLULAR AND SUBORGANELLE LEVEL**

Compared to biochemical separation that is performed using bulk cells, imaging and super resolution imaging provide an in situ measurement of cellular components. Based on the biochemical quantification of IncRNA copy number per cell (Fig. 1A,B), the exact molecule number in designated localized regions can be examined by calculating fluorescent intensities of the regions of interest (ROI) and whole-cell regions through single molecule RNA FISH (smFISH) experiments of the target IncRNAs (Fig. 1C,D; Wu et al. 2016; Yao et al. 2019; Wu et al. 2021). Of note, ROI is represented by the known marker protein in the targeted subdomain (Fig. 1C,D). It is worthwhile mentioning that quantitative imaging demands high-resolution and high-quality images, devoid of unwanted contamination by background signal intensity.

The subregion quantification of a particular protein is similar to that of IncRNA by circling the interest region through protein IF imaging (Wu et al. 2016; Guo et al. 2020; Wu et al. 2021). Since the expression level of RNA and protein per cell is based on biochemical quantification from the standard curve generated from bulk cells (Fig. 1A, B), the averaged quantification per cell does not account for cell-to-cell variability, and accordingly, the calculated copy number at a particular region can vary.

A more precise way for quantification of proteins in living cells is to fuse proteins of interest to a fluorescent protein such as monomeric enhanced green fluorescent protein (mEGFP) using CRISPR/Cas9 technology. This can be used to indicate both protein expression and localization. By generating a fluorescence intensity standard curve of purified mEGFP in vitro at different concentrations, the exact copy of protein in the whole cell and the specific subcellular region can be retrieved (Fig. 1C; Yao et al. 2019). However, one potential disadvantage of this method is that the fused mEGFP domain (∼27 kDa) may change the localization, stability or function of the protein of interest.

Recently developed super-resolution microscopy has enabled the observation and quantification of subcellular structures at the nanometer level (West et al. 2016; Wu et al. 2016, 2021; Xing et al. 2017; Colognori et al. 2019; Elguindy and Mendell 2021). By consolidating super-resolution imaging and specific cellular organelles of interest, the meaningful local concentration of RNA and protein in one IncRNP could be accurately calculated. For instance, structural illumination microscopy (SIM) allied with CHART RNA-seq using antisense oligonucleotides that target different NEAT1_2 regions revealed protein and RNA components arranged as core-shell structures in paraspeckles (West et al. 2016; Wang et al. 2018b). SIM, together with iCLIP RNA-seq, revealed the distinct binding preferences of IncRNA and RBPs in terms of sno-IncRNAs, SPA IncRNAs and splicing factors RBFOX2, hnRNP M and TDP43 in a particular nuclear region that is absent from patients with Prader–Willi syndrome (PWS) (Wu et al. 2016).

**THE SUBCELLULAR LOCALIZATION AND ABUNDANCE OF IncRNAs PROVIDE CLUES ABOUT MODES OF ACTION**

Knowing the subcellular localization gives clues about whether an IncRNA acts in cis or in trans; combining its intrinsic sequence, conformational features and the IncRNA local concentration can further illustrate how an IncRNA exerts transcriptional or post-transcriptional effects by interacting with proteins that are expressed at different levels in cells (Fig. 2).

**LOW ABUNDANCE CHROMATIN-ASSOCIATED IncRNAs CAN BE INVOLVED IN THE IN CIS AND IN TRANS REGULATION OF TRANSCRIPTIONAL OUTPUTS**

Nuclear soluble and insoluble fractionation followed by qRT-PCR or RNA-seq analyses could first give clues about chromatin-associated IncRNAs (Fig. 1B) and their copy number per cell can be retrieved by biochemical quantification (Fig. 1A). Chromatin-associated IncRNAs are found to be expressed at low copy number, at a range of 0.3–100 copies per cell (Table 1). Their modes of action can be classified into two groups: base-pairing with DNA through complementarity to form RNA:DNA hybrids (R-loops) or RNA:DNA triplexes and recruiting or sequestering chromatin-associated RBPs at one locus or several loci (Fig. 2A). It is worthwhile noting that these modes are not mutually exclusive, because base-pairing with DNA very often recruits specific proteins for functions.
In the first scenario, a few copies per cell of lncRNAs that are associated with chromatin are likely adequate for their proposed modes of action. Formation of complex structures at specific genomic regions may in turn recruit additional chromatin modifiers and transcriptional factors to achieve a measurable gene expression change (Li et al. 2016; Arab et al. 2019; Feretzaki et al. 2020). This can be exemplified by the lncRNA TARID (TCF21 antisense RNA inducing demethylation) that forms an R-loop at the TCF21 promoter in cis to trigger the sequential process of GADD45A recruitment, TET1 association, DNA demethylation and TCF21 transcription activation (Fig. 2A; Arab et al. 2014, 2019).}

**FIGURE 2.** The subcellular localization and relative abundance of IncRNA determine its mode of action. (**A**) Examples of working modes of chromatin-associated lncRNAs. (**a**) Base-pairing with the complementary genomic DNA directs lncRNA localization and mode of action. (Top) lncRNA TARID (TCF21 antisense RNA inducing demethylation) forms an R-loop in the TCF21 (transcription factor 21, a tumor suppressor gene) CpG island promoter to recruit a proposed R-loop reader, GADD45A (growth arrest and DNA-damage-inducible, alpha), and cooperates with TET1 (10–11 translocation 1) and TDG (thymine-DNA glycosylase) to drive demethylation of the TCF21 promoter, resulting in TCF21 transcription activation (Arab et al. 2014, 2019). (Bottom) The IncRNA Khps1 (antisense orientation to SPHK1) forms RNA–DNA–DNA triplexes upstream of the transcription start site of proto-oncogene SPHK1-B (Sphingosine kinase 1, isoform B) with a purine-rich sequence. Khps1 prompts an open chromatin structure by recruiting the histone acetyltransferase p300/CBP to the SPHK1 promoter to facilitate SPHK1-B transcription (Postepska-Igielska et al. 2015). (**b**) Chromatin-associated lncRNA Xist (X inactive specific transcript) serves as a recruiter paradigm along the female inactive X chromosome (Xi), which contains ~1000 genes (Tukiainen et al. 2017). Xist recruits different chromatin-modifying factors to form super-molecular complex (SMC) to silence adjacent genes on Xi (Brockdorff 2018). (**B**) Examples of working models of trans-acting lncRNAs. (**a**) One lncRNA possesses multiple binding motifs (top) (Yap et al. 2018) or forms multiple structural modules (bottom) (Guo et al. 2020) to decoy RBP function. (b) Locally enriched IncRNA interacts with multiple RBPs to form nuclear condensates. NEAT1_2 (nuclear paraspeckle assembly transcript 1, the long isoform) is the scaffold RNA of paraspeckles. It folds in half to keep the core region in the center and 5′/3′ termini in the shell within paraspeckles. The radially arranged NEAT1_2 provides an interacting network for paraspeckle proteins (West et al. 2016), among which SFPQ (splicing factor proline and glutamine rich) and NONO (non-POU domain containing octamer binding protein) play critical roles in initiating paraspeckle assembly (Yamazaki et al. 2018). (c) Multiple lncRNAs are enriched as a group to interact with RBPs. Five sno-IncRNAs (small nuclear RNA-ended IncRNA) and two SPAs (5′ small nucleolar RNA capped and 3′ polyadenylated) are transcribed from the Ch15q11–13, and are absent in patients with PWS (Prader–Willi syndrome). These IncRNAs interact with multiple RBPs to form microscopically visible foci that sequester RBPs to affect pre-mRNA alternative splicing (Wu et al. 2016). (d) Many circRNAs possess 16- to 26-bp duplexes and function as a group to inhibit PKR activation (Liu et al. 2019).
Arab et al. 2014, 2019). In terms of the formation of triplex structures, Khps1 (antisense transcript of SPHK1) is tethered to the SPHK1-B promoter in cis via DNA–RNA triplexes and recruits p300/CBP to acetylate histones, promoting an active chromatin conformation to enhance SPHK1-B transcription (Fig. 2A; Postepska-Igielska et al. 2015).

In the second scenario, lncRNAs are located in the proximity of chromatin to scaffold or recruit protein binding (Fig. 2A). P53 induced intergenic noncoding RNA p21 (lincRNA-p21), is expressed at 20–80 copies per cell (Yang et al. 2014), interacts with hnRNP K and prevents re-programming by maintaining the heterochromatic state of pluripotency gene promoters (Bao et al. 2015). Chromatin modifier complex deposition mediated by lncRNAs can also extend over a long range. The most prominent example is the lncRNA Xist, which is 17,000 nt in mouse and is transcribed exclusively from the inactive X chromosome with ~100 copies per cell (Brown et al. 1992; Hartshorn et al. 2003; McHugh et al. 2015; Sunwoo et al. 2015; Markaki et al. 2020). Xist spreads directly from its transcription locus to distal sites across the X chromosome (Lee and Bartolomei 2013; Markaki et al. 2020) in differentiated female cells to silence an X chromosome by recruiting SPEN and a range of other proteins sequentially to form a transcriptionally silent nuclear compartment (Fig. 2A; Chujo et al. 2017; Dossin et al. 2020; Markaki et al. 2020).

**LOCALIZED ENRICHED lncRNAs CAN CONTAIN MULTIPLE BINDING MOTIFS TO ACHIEVE HIGH EFFICIENCY AS DECOYS OR SCAFFOLDS FOR PROTEINS**

LncRNAs acting in trans as decoys or scaffolds in principle require higher abundance. However, even lncRNAs expressed at copy numbers ranging from dozens to hundreds of copies per cell are still outnumbered by their interacting proteins (Table 1). To overcome this issue, lncRNAs often harbor many binding motifs (Yap et al. 2018) or form multiple binding modules (Guo et al. 2020) to interact with proteins. Further, their localized accumulation offers a sufficiently high concentration to achieve measurable effects on interacting proteins, especially in that cellular neighborhood.

**PNCTR** is a >10,000 nt transcript produced by RNA polymerase I (Pol I). It is expressed only at ~36 copies per cell, is exclusively localized to the perinucleolar region (PNC) and is capable of recruiting a substantial amount of PTBP1 (Yap et al. 2018). PTBP1 is an apoptosis activator and prefers to bind YUCUYY and YYUCUY motifs (Fig. 2B), which are highly abundant in PNCTR sequence (2178 and 565, respectively). These motifs serve as PTBP1 concentration “sensors” and enable PNCTR to sequester 7%–27% cellular PTBP1 in theory, which indeed was confirmed as 11%–31% by quantifying IF-smFISH experiments. PNCTR thus provides ample binding sites for PTBP1 to antagonize PTBP1 splicing regulation in a pro-survival function for cells (Fig. 2B; Yap et al. 2018). In another example, FAST is a 547 nt, cytoplasmic transcript that is expressed at 140 copies per human embryonic stem cell (hESC). It forms five independent stem–loops that allow binding of ~700 molecules of β-TrCP (an E3 ubiquitin ligase that targets phosphorylated β-catenin in the WNT destruction complex) (Fig. 2B), accounting for >20% of the total ~3269 copies of β-TrCP in the cytoplasm (Guo et al. 2020). This has enabled a suppressed β-catenin degradation and activated WNT signaling that contributes to the maintenance of hESC pluripotency (Guo et al. 2020).

Except for being targeted by the same proteins, some lncRNAs can interact with a variety of proteins, exemplified by one of the most well-studied lncRNAs, NEAT1_2 (Fig. 2B). NEAT1_2 is a 22,700 nt transcript, expressed at ~1000 copies per HeLa cell (Chujo et al. 2017) and serves as the key scaffold lncRNA of paraspeckles (Sasaki et al. 2009; West et al. 2016), a type of compact and membrane-less nuclear body exhibiting liquid-like properties (Yamazaki et al. 2018). In HeLa cells, each paraspeckle contains ~50 NEAT1_2 molecules (Chujo et al. 2017), providing many binding sites for paraspeckle proteins (PSPs) via distinct subdomains (Hennig et al. 2015; Yamazaki et al. 2018). In the core-shell spheroidal structure of a paraspeckle, the 5′ and 3′ termini of NEAT1_2 display a shell localized pattern and the middle domain of NEAT1_2 contains several functional subdomains. The middle region of NEAT1_2 binds to NONO, SFPQ, and FUS, which is sufficient for paraspeckle assembly (West et al. 2016; Yamazaki et al. 2018). Among the over 40 identified paraspeckle proteins, NONO and SFPQ have strong affinity with NEAT1_2 subdomain 12–13 k, lower but substantial affinity with the adjacent 9.8–12 k and 15.4–16.6 k, to initiate paraspeckle assembly (Yamazaki et al. 2018). In addition, FUS provides an interaction network through its PLD (prion-like domain), leading to paraspeckle assembly via phase separation (Fox et al. 2018). This was also illustrated by RNA pull-down assays combining SILAC-MS analysis where quadruplex-forming G-rich sequences in NEAT1_2 were important for NONO, SFPQ, PSPC1 binding (Simko et al. 2020).

**DIFFERENT lncRNAs CAN ACT AS A GROUP TO ENSURE MEASURABLE EFFECTS ON PROTEIN INTERACTIONS**

In addition to the single type of lncRNA that acts as a decoy or scaffold, multiple different lncRNAs may localize proximately and act as a group by containing the same binding motifs for proteins; alternatively, different lncRNAs can form similar structural modules that allow the same protein to bind these RNAs to achieve a measurable effect.
Five sno-lncRNAs and two SPA lncRNAs are paternally transcribed from Ch15q11–13, a region that is not transcribed in patients with PWS due to parental chromosome deletion, uni-chromosome disomy or epigenetic defects (Yin et al. 2012; Zhang et al. 2014; Wu et al. 2016). While PWS region sno-lncRNAs range from 1100 nt to 2900 nt in size, SPA lncRNAs are larger transcripts, 34,000 nt for SPA1 and 16,000 nt for SPA2. These snoRNA-biogenesis-facilitated lncRNAs are processed from the same polycistronic transcript driven by a strong paternal Pol II promoter (Wu et al. 2016), ultimately yielding relatively abundant SPAs and sno-lncRNAs (approximately one-third of the level of lncRNA MALAT1 in H9 cells). Once produced, these lncRNAs are all localized to their site of production and form microscopically visible lncRNP accumulations, referred to here as PWS bodies (Fig. 2B; Wu et al. 2016). Super-resolution microscopy together with iCLIP RNA-seq revealed that these lncRNAs are colocalized with splicing factors RBFOX2, hnRNP M, and TDP43 and provide specific binding motifs with certain preferences (Wu et al. 2016). For example, sno-lncRNAs prefer to bind RBFOX2 (UGCAU and GCAUG motifs), whereas SPA1 prefers to bind TDP43 (tandem UG-repeat sequences) as revealed by iCLIP analysis (Wu et al. 2016). Quantitative imaging of each RBP and these PWS-region lncRNAs has shown that greater than 1% of each RBP could be sequestered by these lncRNAs, suggesting that the PWS body takes up 0.02%–0.1% of the nuclear volume but sequesters greater than 1% of each associated RBP. As a consequence, in a PWS hESC cellular model lacking all these seven lncRNAs, altered splicing events that are associated with the mislocalization of these RBPs on pre-mRNAs were identified (Wu et al. 2016). Among the misregulated pre-mRNA splicing events were confirmed in hiPSCs derived from a PWS patient (Wu et al. 2016).

Another example can be illustrated by circRNAs, which are known to be expressed at a low level in general (Chen 2020). Using an optimized circSHAPE-MaP approach to measure the folding conformation of circRNAs (Liu et al. 2019), it was found that 26 out of 34 examined circRNAs contain one to four intramolecular 16- to 26-bp RNA duplexes (intra-dsRNAs), which are seen less frequently in their linear mRNA cognates. This has allowed the short intra-dsRNA-containing circRNAs to work as a group to bind PKR and prevent it from dimerization for activation (Fig. 2B; Liu et al. 2019). PKR is a double-stranded RNA-dependent protein kinase associated with direct antiviral activity and immune responses, and in this process, PKR activation requires dsRNAs longer than 33 bp, while short dsRNAs of 16 to 33 bp in length can block its activation (Zheng and Bevilacqua 2004; Bou-Nader et al. 2019). Given this structure-dependent, but sequence-independent, manner for PKR activation, and that each cell has ∼10,000 copies of circRNAs many of which have one to four intra-dsRNA regions, circRNAs as a group can modulate PKR activation (Liu et al. 2019).

**IncRNA-INDUCED PROTEIN PHASE SEPARATION PROMOTES EFFICIENT lncRNP ASSEMBLY TO OUTCOMPETE THE SAME BINDING MOTIFS ON OTHER EXPRESSED RNAs**

Even for IncRNAs that are abundantly expressed at hundreds of copies per cell, they may still display a low stoichiometric ratio compared to total cellular expressed RNAs containing the same binding motifs for a particular protein. For example, NORAD plays a prominent role in genome stability by sequestering and antagonizing PUM1 and PUM2, which have inhibitory effects on mitotic, DNA repair, and DNA replication factors related to aberrant mitosis (Lee et al. 2016; Tichon et al. 2016). Depletion of Norad in mice causes premature aging due to augmented activity of PUM proteins (Kopp et al. 2019). NORAD is expressed at ∼400 copies per HCT116 cell, and each transcript contains 18 PREs (PUMILIO response elements) for PUM binding (Lee et al. 2016). Given that there are ∼15,000 PUM1 plus ∼2000 PUM2 proteins per cell, NORAD provides sufficient PUM1 and PUM2 binding sites to restrain functional PUM proteins (Lee et al. 2016). However, in terms of element number, these 7200 PREs (400 × 18) are unlikely to compete well against overall 130,000–325,000 PREs within the whole mRNA transcriptome (Fig. 3A; Elguindy and Mendell 2021).

PUM proteins contain large amino-terminally located IDRs. Super-resolution microscopy has shown that NORAD and PUM are colocalized in the cytosol to form microscopically visible foci named NP (NORAD and PUM) bodies. NP bodies have a liquid–liquid phase separation property, in which PUM protein phase separation is driven by NORAD in a PRE-number-dependent manner. NORAD thus efficiently sequesters PUM beyond simple competitive titration based on RNA valency and PRE abundance by creating a micro-environment to effectively recruit a large number of PUM to these sites. In such a model, once NORAD-PUM clustering via multivalent RNA-binding interactions is established, additional PUM will be recruited and concentrated by IDR–IDR interactions of PUM proteins and finally enlarged into a NP body.

Given the fact that many RBPs, for example, those enriched in paraspeckles, undergo phase separation (Hennig et al. 2015; Fox et al. 2018) and that the lncRNA NEAT1_2 mid-region itself contains a phase separation property (Yamazaki et al. 2018), it remains possible that paraspeckles, and likely other lncRNP complexes such as Xist RNP (Markaki et al. 2020), may nucleate their formation via an efficient multivalent mechanism, in addition to coupling with Pol II transcription (Mao et al. 2011a; Hnisz et al. 2017; Shrinivas et al. 2019; Rippe and Papantonis 2021).
IncRNAs CAN PROMOTE CONFORMATIONAL CHANGE OF PROTEIN CLUSTERS BY A MOLECULAR CHAPERONE-LIKE MECHANISM AT A SUBSTOICHIOMETRIC RATIO

Super resolution microscopy IncRNPs in situ studies have shown an extremely low stoichiometric ratio of the IncRNA SLERT to its binding protein DDX21, which together play important roles in human nucleolar organization and Pol I transcription (Fig. 3B; Xing et al. 2017; Wu et al. 2021). In the nucleolus, Pol I transcription of rDNA occurs at the border of fibrillar centers (FC) and dense fibrillar components (DFC) (Fig. 3B), making FC/DFC organization as a unit for Pol I transcription and nascent pre-rRNA processing (Yao et al. 2019). SLERT is a H/ACA-type snoRNA-ended IncRNA, transcribed and processed from the TBRG4 locus, and then translocated to the nucleolus for function (Xing et al. 2017). Depletion of SLERT led to reduced cell proliferation, consistent with its role in promoting Pol I transcription (Xing et al. 2017).

SLERT is exclusively localized at FC/DFC units, where it binds to DDX21 (Xing et al. 2017), a DEAD-box RNA helicase that can adopt an open or closed conformation upon substrate binding (Chen et al. 2020). DDX21 is largely enriched outside of each FC/DFC, where it forms a discontinuous shell-like structure coating each FC/DFC by switching its intermolecular and intramolecular interactions mediated by unstructured amino- and carboxy-termini. Increased intermolecular interaction led to DDX21 multimerization, resulting in decreased size of the shell and FC/DFC structure, reduced mobility of FC/DFC components, and suppressed Pol I transcription (Fig. 3B; Wu et al. 2021). Intriguingly, there are about 13 SLERT and 13,000 DDX21 molecules per FC/DFC (Wu et al. 2021), and yet these 13 SLERT are capable of promoting significantly more abundant DDX21 proteins in FC/DFC to adopt a closed conformation, ultimately leading to the loosened environment of an FC/DFC unit and proper Pol I transcription (Fig. 3B; Wu et al. 2021). SLERT prefers to bind DDX21 with open conformation. It was proposed that under this super-stoichiometric ratio, once SLERT induces a DDX21 protein to adopt the closed conformation, this DDX21 would be released, allowing the recruitment of another one with an open conformation to initiate conformational transition, resulting in the formation of hypomultimerized DDX21 molecules in FC/DFC, which is required for FC/DFC liquidity and Pol I processivity. Moreover, without SLERT, DDX21 clusters are strongly engaged with rDNAs to prevent Pol I binding to rDNAs for transcription (Fig. 3B).

How exactly SLERT acts as a molecular chaperone to coordinate DDX21 conformational change in clusters warrants future studies that will require the development of...
new methods for the analysis of phase-separated lncRNPs. Nevertheless, both studies (Fig. 3A,B) have raised the intriguing possibility that lncRNPs may work as clusters in certain condensates, rather than independently in solution as previously thought.

**PERSPECTIVES**

Functional studies on lncRNAs have been burgeoning in the past decade. Numerous efforts have been made to explore their biogenesis, expression, localization, turnover, conformation, and function. These efforts have led to our current understanding of how lncRNAs use different mechanisms to overcome their naturally low expression to effectively interact with protein partners. More efforts are certainly needed to understand particular lncRNP complexes by developing new approaches to decode their molecular basis for interactions and to visualize their dynamics at the single molecule and suborganelle levels.

**ACKNOWLEDGMENTS**

We apologize to colleagues whose work could not be discussed owing to space limitations. This work was supported by the National Natural Science Foundation of China (NSFC) (31725009, 31830108), the Chinese Academy of Sciences (CAS) (Y5BR-009), the Shanghai Municipal Commission for Science and Technology (20JC1410300), and the Howard Hughes Medical Institute (HHMI) International Program (55008728) to L.-L.C. L.-L.C. acknowledges support from the Xplorer Prize.

**REFERENCES**

Aebesold R, Mann M. 2003. Mass spectrometry-based proteomics. *Nature* **422**: 198–207. doi:10.1038/nature01511

Ahme E, Molzahn L, Glatte T, Schmidt A. 2013. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* **13**: 2567–2578. doi:10.1002/pmic.201300135

Andersen JS, Lam YW, Leung AKL, Ong SE, Lyon CE, Lamond AI, Mann M. 2005. Nucleolar proteome dynamics. *Nature* **433**: 77–83. doi:10.1038/nature03207

Arab K, Park YJ, Lindroth AM, Schafer A, Oakes C, Weichenhan D, Lukanova A, Lundin E, Risch A, Meister M, et al. 2014. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol Cell* **55**: 604–614. doi:10.1016/j.molcel.2014.06.031

Arab K, Karaulanov E, Musheev M, Trnka P, Schafer A, Grummt I, Niehrs C. 2019. GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet* **51**: 217. doi:10.1038/s41588-018-0306-6

Atianand MK, Hu W, Satpathy AT, Shen Y, Ricci EP, Alvarez-Dominguez JR, Bhatta A, Schattgen SA, McGowan JD, Blin J, et al. 2016. A long noncoding RNA lincRNA-EPS acts as a transcriptional brake to restrain inflammation. *Cell* **165**: 1672–1685. doi:10.1016/j.cell.2016.05.075

Bao X, Wu H, Zhu X, Guo X, Hutchins AP, Luo Z, Song H, Chen Y, Dai K, Yin M, et al. 2019. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res* **25**: 80–92. doi:10.1038/cr.2014.165

Bartkiewicz M, Gold H, Altman S. 1989. Identification and characterisation of an RNA molecule that copurifies with RNase P activity from HeLa cells. *Genes Dev* **3**: 488–499. doi:10.1101/gad.3.4.488

Bosch JA, Chen CL, Perimmon N. 2021. Proximity-dependent labeling methods for protomic profiling in living cells: an update. *Wires Dev Biol* **10**: e392. doi:10.1002/wdev.392

Bou-Nader C, Gordon JM, Henderson FE, Zhang JW. 2019. The search for a PKR code-differential regulation of protein kinase R activity by diverse RNA and protein regulators. *RNA* **25**: 539–556. doi:10.1261/ma.070169.118

Broddorf N. 2018. Local tandem repeat expansion in Xist RNA as a model for the functionalisation of ncRNA. *Noncoding RNA* **4**: 8. doi:10.3390/ncrna40400028

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. The human Xist gene: analysis of a 17 Kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**: 527–542. doi:10.1016/0092-8674 (92)05052-M

Budnik B, Levy E, Harmange G, Slavon N. 2018. SCOPe-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**: 161. doi:10.1186/s13059-018-1547-5

Cabili MN, Dunagin PD, McClanahan PD, Biadasher O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16**: 20. doi:10.1186/s13059-015-0586-4

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, 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, Fischer B, Frese CK, Horos R, Alleaume AM, Foehr S, Czk T, Krijgsvald J, Hentze MW. 2016. Comprehensive identification of RNA-binding domains in human cells. *Mol Cell* **63**: 696–710. doi:10.1016/j.molcel.2016.06.029

Caudron-Herger M, Pankert T, Seiler J, Nemeth A, Voit R, Grummt I, Rippe K. 2015. Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J* **34**: 2758–2774. doi:10.15252/embj.201591458

Chelius D, Bondarenko PV. 2002. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **1**: 317–323. doi:10.1021/pr025517J

Chen LL. 2020. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol* **21**: 475–490. doi:10.1038/s41580-020-0243-y

Chen LL, DeCerbo JN, Carmichael GG. 2008. Alu element-mediated gene silencing. *EMBO J* **27**: 1694–1705. doi:10.1038/emboj.2008.94

Chen ZJ, Li ZY, Hu XJ, Xie FY, Kuang SY, Zhan BW, Gao WQ, Chen XJ, Gao SQ, Li Y, et al. 2020. Structural basis of human helicase DDX21 in RNA binding, unwinding, and antiviral signal activation. *Adv Sci* **7**: 2000532. doi:10.1002/advs.202000532

Cheng YT, Jutoori I, Chadalapaka G, Corton JC, Safe S. 2015. The long non-coding RNA HOTTIP enhances pancreatic cancer cell proliferation, survival and migration. *Oncotarget* **6**: 10840–10852. doi:10.18632/oncotarget.3450

Choi-Rhee E, Schulman H, Cronan JE. 2004. Promiscuous protein bio-tyination by Escherichia coli biotin protein ligase. *Protein Sci* **13**: 3043–3050. doi:10.1111/j.0994-1053.2004.0911804

Chu C, Zhang QFC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. 2015.
Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, Mendell JT. 2016. Noncoding RNA NORAD regulates genomic stability by quenstering PUMILIO proteins. Cell 164: 69–80. doi:10.1016/j.cell.2015.12.017
Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, Radaelli E, Eeyckerman S, Leonelli C, Vanderheyden K, et al. 2016. Melanoma addiction to the long non-coding RNA SAMMSON. Nature 531: 518–522. doi:10.1038/nature17161
Li XW, Rees JS, Xue P, Zhang H, Hamaia SW, Sanderson B, Funk PE, Farnsdale RW, Lilley KS, Perretti S, et al. 2014. New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. J Biol Chem 289: 14434–14447. doi:10.1074/jbc.M113 .529578
Li Y, Syed J, Sugiyama H. 2016. RNA-DNA triplex formation by long noncoding RNAs. Chem Biol 23: 1325–1333. doi:10.1016/j.chembiol.2016.09.011
Li S, Li X, Xue W, Zhang L, Yang LZ, Cao SM, Lei YN, Liu CX, Guo SK, Shan L, et al. 2021. Screening for functional circular RNAs using the CRISPR-Cas13 system. Nat Methods 18: 51–59. doi:10.1038/s41592-020-01101-4
Liu Q, Zheng J, Sun WP, Huo YB, Zhang LY, Hao PL, Wang HP, Zhuang M. 2018. A proximity-tagging system to identify membrane-protein-protein interactions. Nat Methods 15: 715. doi:10.1038/s41592-018-0100-5
Liu CX, Li X, Nan F, Jiang S, Gao X, Guo SK, Xue W, Cui Y, Dong K, Ding H, et al. 2019. Structure and degradation of circular RNAs regulate PKR activation in innate immunity. Cell 177: 865–880. e21. doi:10.1016/j.cell.2019.03.046
Lo Piccolo L, Bonaccorso R, Onorati MC. 2015. Nuclear and cytoplasmic soluble proteins extraction from a small quantity of Drosophila’s whole larvae and tissues. Int J Mol Sci 16: 12360–12367. doi:10.3390/ijms160612360
Ludwig C, Claassen M, Schmidt A, Aebersold R. 2012. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. Mol Cell Proteomics 11: M111.013987. doi:10.1074/mcp.M111.013987
Luo HC, Zhu GQ, Xu JF, Lai Q, Yan BW, Guo Y, Fung TK, Zeissig BG, Cui Y, Zha J, et al. 2019. HOTTIP IncRNA promotes hematopoietic stem cell self-renewal leading to AML-like disease in mice. Cancer Cell 36: 645. doi:10.1016/j.ccc.2019.10.011
Maharan S, Wang J, Papadopoulos DK, Richter D, Pozniakovsy A, Poser I, Bickle M, Rizk S, Guillem-Boixet J, Franzmann TM, et al. 2018. RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science 360: 918–921. doi:10.1126/science.aat7366
Mao YTS, Sunwoo H, Zhang B, Spector DL. 2011a. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol 13: 95–101. doi:10.1038/ncb2140
Mao YTS, Zhang B, Spector DL. 2011b. Biogenesis and function of nuclear bodies. Trends Genet 27: 295–306. doi:10.1016/j.tig.2011.05.006
Markaki Y, Chong JG, Luong C, Tan SYX, Wang Y, Jacobson EC, Maestrini D, Dror I, Mistry BA, Schöneberg J, et al. 2020. Xist-seeded nascent sites form local concentration gradients of silencing proteins to inactivate the X-chromosome. bioRxiv doi:10.1101/2020.11.22.393546
Martell JD, Deerinck TJ, Sancak Y, Poulos TL, Mootha VK, Sosinsky GE, Ellman MH, Ting AY. 2012. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat Biotechnol 30: 1143–1148. doi:10.1038/nbt.2375
McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandy-Jones A, Blanco M, Burghard C, Moradian A, et al. 2015. The Xist IncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521: 232–236. doi:10.1038/nature14443
Sunwoo H, Wu JY, Lee JT. 2015. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. Proc Natl Acad Sci 112: E4216–E4225. doi:10.1073/pnas.1503690112

Takata H, Nishijima H, Obara S, Sakaguchi T, Bubulya PA, Mochizuki T, Shibahara K. 2009. Proteome analysis of human nuclear insoluble fractions. Genes Cells 14: 975–990. doi:10.1111/j.1365-2443.2009.01324.x

Tichon A, Gil N, Lubselsky Y, Havkin Solomon T, Lemze D, Itzkovitz S, Stern-Ginossar N, Ultsky I. 2016. A conserved abundant cytoplasmic long noncoding RNA regulates repression by Pumilio proteins in human cells. Nat Commun 7: 12209. doi:10.1038/ncomms12209

Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, et al. 2017. Landscape of X chromosome inactivation across human tissues. Nature 550: 244. doi:10.1038/nature24265

Wang KC, Yang YY, Liu B, Sanjay A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011. A long non-coding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472: 120–124. doi:10.1038/nature09819

Wang P, Xu JF, Wang YJ, Cao XT. 2017. An interferon-independent gene expression. Nature 555: 1015–1055. doi:10.1126/science.aao4090

Wang XX, He Y, Ye Y, Zhao XY, Deng S, He G, Zhu HX, Xu NZ, Liang SF. 2018a. SILAC-based quantitative MS approach for real-time recording protein-mediated cell-cell interactions. Sci Rep 8: 8441. doi:10.1038/s41598-018-26262-2

Wang Y, Hu SB, Wang MR, Yao RW, Wu D, Yang L, Chen LL. 2018b. Genome-wide screening of NEAT1 regulators reveals cross-regulation between paraspeckles and mitochondria. Nat Cell Biol 20: 1145. doi:10.1038/s41564-018-0204-2

Wang X, Hu X, Song W, Xu H, Xiao Z, Huang R, Bai Q, Zhang F, Chen Y, Liu Y, et al. 2021. Mutual dependency between IncRNA LETN and protein NPM1 in controlling the nucleolar structure and functions sustaining cell proliferation. Cell Res 31: 664–683. doi:10.1038/s41422-020-00456-6

West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, Tolstorukov MY, Kingston RE. 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

West JA, Mitto M, Kurosaka S, Takumi T, Tanegashima C, Chuo T, Yanaka K, Kingston RE, Hirose T, Bond C, et al. 2016. Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. J Cell Biol 214: 817–830. doi:10.1083/jcb.201601071

Wu H, Yin QF, Luo Z, Yao RW, Zheng CC, Zhang J, Xiang JF, Yang L, Chen LL. 2016. Unusual processing generates SPA IncRNAs that sequester multiple RNA binding proteins. Mol Cell 64: 534–546. doi:10.1016/j.molcel.2016.10.007

Wu H, Yang L, Chen LL. 2017. The diversity of long noncoding RNAs and their generation. Trends Genet 33: 540–552. doi:10.1016/j.tig.2017.05.004

Wu M, Xu G, Han C, Luan PF, Xing YH, Nan F, Yang LZ, Huang Y, Yang ZH, Shan L, et al. 2021. IncRNA SLERT controls phase separation of FC/DFCs to facilitate Pol I transcription. Science 373: 547–555. doi:10.1126/science.abf6582

Xi L, Cech TR. 2014. Inventory of telomerase components in human cells reveals multiple subpopulations of HTR and hTERT. Nucleic Acids Res 42: 8565–8577. doi:10.1093/nar/gku560

Xing YH, Yao RW, Zhang Y, Guo CJ, Jiang S, Xu G, Dong R, Yang L, Chen LL. 2017. SLERT regulates DDX21 rings associated with Pol I transcription. Cell 169: 664–678.e16. doi:10.1016/j.cell.2017.04.011

Yamazaki T, Souquere S, Chuo T, Kobelke S, Chong YS, Fox AH, Bond CS, Nakagawa S, Pierron G, Hirose T. 2018. Functional domains of NEAT1 architectural IncRNA induce paraspeckle assembly through phase separation. Mol Cell 70: 1038–1053.e7. doi:10.1016/j.molcel.2018.05.019

Yang F, Zhang H, Mei Y, Wu M. 2014. Reciprocal regulation of HIF-1α and IncRNA-p21 modulates the Warburg effect. Mol Cell 53: 88–100. doi:10.1016/j.molcel.2013.11.004

Yao RW, Wu G, Yang Y, Shan L, Luan PF, Wang Y, Wu M, Yang LZ, Xing YH, Yang L, et al. 2019. Nascent pre-rRNA sorting via phase separation drives the assembly of dense fibrillar components in the human nucleolus. Mol Cell 76: 767–783.e11. doi:10.1016/j.molcel.2019.08.014

Yap K, Mukhina S, Zhang G, Tan JSC, Ong HS, Makeyev EV. 2018. A short tandem repeat-enriched RNA assembles a nuclear compartment to control alternative splicing and promote cell survival. Mol Cell 72: 525–540.e13. doi:10.1016/j.molcel.2018.08.041

Yin QF, Yang L, Zhang Y, Xiang JF, Wu WW, Carmichael GG, Chen LL. 2012. Long noncoding RNAs with snoRNA ends. Mol Cell 48: 219–230. doi:10.1016/j.molcel.2012.07.033

Zhang XO, Yin QF, Wang HB, Zhang Y, Chen T, Zheng P, Lu XH, Chen LL, Yang L. 2014. Species-specific alternative splicing leads to unique expression of sno-IncRNAs. BMC Genomics 15: 287. doi:10.1186/1471-2164-15-287

Zhang HY, Elbaum-Garfinkle S, Langdon EM, Taylor N, Occhipinti P, Bridges AA, Brangwynne CP, Gladfelter AS. 2015. RNA controls polyQ protein phase transitions. Mol Cell 60: 220–230. doi:10.1016/j.molcel.2015.09.017

Zhao L, Wang J, Li Y, Song T, Wu Y, Fang S, Bu M, Li H, Sun L, Pei D, et al. 2021. NONCODEv6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. Nucleic Acids Res 49: D165–D171. doi:10.1093/nar/gkaa1046

Zheng XF, Bevilacqua PC. 2004. Activation of the protein kinase PKR by short double-stranded RNAs with single-stranded tails. RNA 10: 1934–1945. doi:10.1261/rna.7150804

Zhou HK, Nguemaha V, Mazarakos K, Qin SB. 2018. Why do ordered and structured proteins behave differently in phase separation? Trends Biochem Sci 43: 499–516. doi:10.1016/j.tibs.2018.03.007

Zhu Y, Piekowska D, Zhao R, Chen J, Shen YF, Moore RJ, Shukla AK, Petyluk VA, Campbell-Thompson M, Mathews CE, et al. 2018. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. Nat Commun 9: 882. doi:10.1038/s41467-018-03367-w