Unusual semi-extractability as a hallmark of nuclear body-associated architectural noncoding RNAs

Takeshi Chujo¹, Tomohiro Yamazaki¹, Tetsuya Kawaguchi¹ , Satoshi Kurosaka², Toru Takumi², Shinichi Nakagawa³ & Tetsuro Hirose^{1,*}

Abstract

NEAT1_2 long noncoding RNA (lncRNA) is the molecular scaffold of paraspeckle nuclear bodies. Here, we report an improved RNA extraction method: extensive needle shearing or heating of cell lysate in RNA extraction reagent improved *NEAT1_2* extraction by 20-fold (a property we term “semi-extractability”), whereas using a conventional method *NEAT1_2* was trapped in the protein phase. The improved extraction method enabled us to estimate that approximately 50 *NEAT1_2* molecules are present in a single paraspeckle. Another architectural lncRNA, *IGS16*, also exhibited similar semi-extractability. A comparison of RNA-seq data from needle-sheared and control samples revealed the existence of multiple semi-extractable RNAs, many of which were localized in sub-nuclear granule-like structures. The semi-extractability of *NEAT1_2* correlated with its association with paraspeckle proteins and required the prion-like domain of the RNA-binding protein FUS. This observation suggests that tenacious RNA–protein and protein–protein interactions, which drive nuclear body formation, are responsible for semi-extractability. Our findings provide a foundation for the discovery of the architectural RNAs that constitute nuclear bodies.

Keywords long noncoding RNA; nuclear body; prion-like domain; RNA extraction; RNA-binding protein

Subject Categories Methods & Resources; Protein Biosynthesis & Quality Control; RNA Biology

DOI 10.15252/embj.201695848 | Received 7 October 2016 | Revised 2 February 2017 | Accepted 9 March 2017 | Published online 12 April 2017

The EMBO Journal (2017) 36: 1447–1462

Introduction

The mammalian cell nucleus is highly organized. In addition to chromatin, the nucleus contains distinct membraneless nuclear bodies, each of which consists of specific sets of RNAs and

proteins that act in various nuclear processes. Several types of nuclear bodies contain specific long noncoding RNAs (lncRNAs), some of which play architectural roles in the formation of the bodies in which they reside (Chujo *et al*, 2016). One remarkable example is paraspeckles, nuclear bodies initially defined as foci located in close proximity to nuclear speckles; paraspeckles are enriched in characteristic RNA-binding proteins such as PSPC1 and NONO (Fox *et al*, 2002). *Nuclear paraspeckle assembly transcript 1* (*NEAT1*) is a 23 kb lncRNA that localizes exclusively in paraspeckles and serves as an essential structural component of these bodies (Clemson *et al*, 2009; Sasaki *et al*, 2009; Sunwoo *et al*, 2009). *NEAT1* and paraspeckles regulate expression of multiple genes by sequestering specific proteins and RNAs (Prasanth *et al*, 2005; Chen & Carmichael, 2009; Hirose *et al*, 2014b). Physiologically, *NEAT1* and paraspeckles are involved in development of the corpus luteum and mammary gland in mice and also protect cells from viral and DNA damage-related stresses (Imamura *et al*, 2014; Nakagawa *et al*, 2014; Standaert *et al*, 2014; Adriaens *et al*, 2016). Paraspeckle formation is initiated by transcription of *NEAT1* from its chromosomal locus, located on human chromosome 11 and mouse chromosome 19, and proceeds in conjunction with co-transcriptional association of *NEAT1* lncRNA with > 40 paraspeckle proteins (PSPs), most of which exhibit features of canonical RNA-binding proteins (Mao *et al*, 2011). Paraspeckles form a spheroidal structure in which the outer layer contains the 5' and 3' regions of *NEAT1* and specific PSPs, whereas the inner core structure contains the middle region of *NEAT1* and distinct sets of PSPs (West *et al*, 2016). Among the > 40 PSPs present in paraspeckles, seven are essential for formation of these bodies (Naganuma *et al*, 2012). These essential PSPs contain RNA-binding domains, as well as characteristic low-complexity domains (also called prion-like domains, PLDs) (King *et al*, 2012; Naganuma *et al*, 2012; Yamazaki & Hirose, 2015), which were recently reported to promote liquid–liquid phase separation *in vitro*, resulting in dynamic liquids or hydrogels (Kato *et al*, 2012; Burke *et al*, 2015; Molliex *et al*, 2015). Moreover, the PLDs in two PSPs, RBM14 and FUS/TLS, are essential for cellular paraspeckle formation (Hennig *et al*, 2015).

¹ Institute for Genetic Medicine, Hokkaido University, Sapporo, Hokkaido, Japan

² Brain Science Institute, RIKEN, Wako, Saitama, Japan

³ Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo, Hokkaido, Japan

*Corresponding author. Tel: +81 11 706 5071; Fax: +81 11 706 7540; E-mail: hirose@igm.hokudai.ac.jp

Like *NEAT1*, several other lncRNAs play similar architectural roles in the formation of specific nuclear bodies, for example, *intergenic spacer* (IGS) lncRNAs in the nucleolar detention center (Audas et al, 2012), *satellite III* lncRNA in the nuclear stress body (Biamonti & Vourc'h, 2010), *Drosophila heat shock RNA omega* in the omega speckle (Prasanth et al, 2000), and fission yeast *meiRNA* in the Mei2 dot (Watanabe & Yamamoto, 1994). We proposed classifying these nuclear body-constructing RNAs as a distinct subclass of lncRNA called "architectural RNA (arcRNA)" (Chujo et al, 2016). Furthermore, a screen to identify RNase-sensitive nuclear bodies revealed that additional nuclear bodies (e.g., the Sam68 nuclear body and DBC1 body) require arcRNAs and prion-like RNA-binding proteins for their structural integrity (Mannen et al, 2016). As with paraspeckles, PLD-bearing RNA-binding proteins and/or chromatin remodeling complexes are involved in the formation of these subcellular structures, raising the possibility that a common molecular pathway underlies arcRNA-dependent formation of nuclear bodies (Kawaguchi et al, 2015; Chujo et al, 2016; Mannen et al, 2016).

Formation of arcRNA-dependent nuclear bodies occurs under various stresses and developmental stages and is initiated by induced expression of the arcRNAs. Accordingly, it is likely that as-yet-unknown arcRNA-dependent nuclear bodies remain to be identified in various cellular contexts. In this study, we discovered that *NEAT1* is exceptionally difficult to extract using conventional RNA purification reagents. Inefficient *NEAT1* extraction, a property that we refer to here as semi-extractability, resulted from entrapment of a substantial proportion of *NEAT1* in the protein phase during conventional RNA extraction. This feature required the *NEAT1*-binding protein FUS, particularly its PLD, suggesting that semi-extractability may reflect extensive RNA–protein and protein–protein interactions in paraspeckles. This discovery prompted us to search for additional semi-extractable RNAs by next-generation RNA sequencing to reveal novel arcRNA candidates that associate with prion-like RNA-binding proteins. This analysis identified multiple semi-extractable RNAs, many of which localized in subnuclear granule-like structures. Our findings strongly suggest that semi-extractability is a hallmark of nuclear body-associated arcRNAs.

Results

An improved RNA extraction method demonstrates that *NEAT1_2* lncRNA is semi-extractable under conventional RNA purification conditions

Conventional methods for RNA purification from cultured cells often use an acid guanidinium thiocyanate–phenol–chloroform (AGPC) reagent such as TRIzol; cellular RNA molecules are thought to be comprehensively and reproducibly extracted under these conditions. In the course of our experiments, however, we discovered that human *NEAT1_2* lncRNA was inefficiently extracted by conventional protocols. Needle shearing or heating of the HeLa cell lysate in the AGPC reagent markedly increased the extractability of *NEAT1_2* (Fig 1A and B). Specifically, shearing of the sample 100 times by passage through a 20-gauge needle dramatically increased the yield of *NEAT1_2*, representing a > 20-fold increase relative to the conventional protocol. Without needle shearing, the vast majority of *NEAT1_2* likely partitioned into the protein phase, that is, the

insoluble layer between the water and organic phases. Indeed, we confirmed that the level of *NEAT1_2* in the protein phase decreased > 10-fold following needle shearing (Fig 1C). More stringent needle shearing (up to 200 times, or using a 25-gauge needle) did not appreciably increase the yield (Fig 1B). Needle shearing caused little if any RNA degradation, as determined by RNA integrity number (RIN; Fig 1B). Heating of the sample at 55°C for 10 min gave almost the same yield as needle shearing, with no obvious RNA degradation. However, heating at 65°C resulted in slight RNA degradation (see RIN in Fig 1B). A combination of heating (55°C) and needle shearing (20 gauge, 100×) did not further increase the yield (Fig 1B). DNase treatment of cell lysate prior to RNA extraction did not change the semi-extractability of *NEAT1_2* (Fig EV1A). On the other hand, proteinase K treatment of the cell lysate prior to RNA extraction increased the extractability of *NEAT1_2* (Fig EV1B), suggesting that protein(s) may be responsible for the semi-extractability of *NEAT1_2*, consistent with the observation that *NEAT1_2* was trapped in the protein phase during conventional RNA extraction (Fig 1C). Even after proteinase K treatment, however, *NEAT1_2* was not fully extractable and still required needle shearing, reminiscent of proteinase K-resistant prion protein (Saborio et al, 2001). Collectively, these results demonstrate that needle shearing or heating of cell lysate in the AGPC reagent is the most efficient way to extract cellular *NEAT1_2*. Whereas the heating method is convenient and may be applied to extract large numbers of RNA samples for RT-qPCR, etc., the needle shearing method is labor-intensive but suitable for obtaining high-integrity RNA samples (e.g., for northern blot).

Semi-extractability of *NEAT1* is observed ubiquitously, yet differentially, in various human and mouse cells

Taking advantage of the improved RNA extraction method, we next performed quantitative characterization of *NEAT1*, including *NEAT1* isoform ratio, comparisons with other RNA species, *NEAT1* expression levels in various cells, and the number of *NEAT1* molecules in a single paraspeckle. An RNase protection assay (RPA) revealed that the increase in extraction efficiency was more conspicuous (> 20-fold) for the longer isoform *NEAT1_2*, whereas the yield of the shorter isoform *NEAT1_1* increased only ~threefold (Fig 2A). We previously estimated the *NEAT1_1*:*NEAT1_2* in HeLa cells as 7:3 (Sasaki et al, 2009); however, our new method revealed that the actual molar ratio is approximately 1:9 (Fig 2A). We confirmed that semi-extractability was a special feature of *NEAT1_2*; other cellular RNAs, including representative mRNAs (*GAPDH* and *HSP70*), unspliced or incompletely spliced mRNA precursors (Pre-MYC and Pre-CLK1), relatively small nuclear ncRNAs (*7SK* and *TERC*), rRNA precursors (pre-18S and pre-28S), and multiple previously characterized lncRNAs, were not semi-extractable (Fig 2B), even though some of the lncRNAs (e.g., *MALAT1* and *TUG1*) are known to localize in specific nuclear bodies. *MALAT1* localizes in nuclear speckles, but does not play an architectural role in this nuclear body (Tripathi et al, 2010; Nakagawa et al, 2012). On the other hand, needle shearing increased by ~twofold the extraction of *IGS16* lncRNA, the arcRNA of the nucleolar detention center, which forms under heat shock conditions (Fig 2C). Although this twofold increase is relatively small compared to that of *NEAT1*, it was still larger than those of any other RNAs we tested (Fig 2B). This observation supports the

Figure 1. An improved RNA extraction method to purify semi-extractable *NEAT1_2* lncRNA.

A Outline of the improved RNA extraction method and its application to identify arcRNA candidates. Prior to addition of chloroform, cell lysate in AGPC reagent was either sheared with a needle or heated at 55°C.

B Upon needle shearing or heating, > 20-fold more *NEAT1_2* was extracted from HeLa cells. RNA levels were quantitated by RT-qPCR. RIN, Bioanalyzer RNA integrity number.

C Entrapment of *NEAT1_2* in protein phase during conventional RNA extraction procedure. RNA was extracted from the water or protein phase.

Data information: Bar graphs show means $\pm$ SD ($n = 3$).

idea that semi-extractability is associated with the architectural role of lncRNAs, rather than their localization in nuclear bodies *per se*.

Paraspeckles of various sizes and numbers are present in multiple mammalian cell lines. Among nine human cell lines that we investigated, three cell lines (HeLa, T24, and MRC5) had easily recognizable paraspeckles, three (MCF7, A549, and HEK293) had detectable paraspeckles, and three (Huh7, HCT116, and U2OS) lacked detectable paraspeckles (Fig 2D and enlarged in Appendix Fig S1A). We extracted total RNA from each cell line using the needle shearing procedure and then performed RPA to detect the two *NEAT1* isoforms (Fig 2E). The presence of paraspeckles in these cell lines roughly correlated with expression levels of *NEAT1_2*, and the molar ratio of *NEAT1_1* and *NEAT1_2* varied among the cell lines. Comparison of RNA levels in samples treated with or without needle shearing revealed that the semi-extractability of *NEAT1_2* also roughly correlated with *NEAT1_2* expression levels and the presence of paraspeckles: upon needle shearing, RNA yield increased 14- to 27-fold in HeLa, T24, and MRC5, but only twofold–fivefold in Huh7, HCT116, and U2OS (Fig 2F). We also confirmed that mouse *Neat1_2* was semi-extractable in three mouse cell lines and in primary culture cells from the luteum, an adult mouse tissue in which *Neat1* is highly expressed, indicating that this feature of

NEAT1_2 is shared by at least two mammalian species (Fig 2G). Notably, *NEAT1_2* was semi-extractable to a considerable degree in three cell lines lacking visible paraspeckles (two- to fivefold increase in yield upon shearing). This result suggests that the semi-extractability of *NEAT1_2* is primarily associated with *NEAT1* RNP substructures and becomes more conspicuous when multiple *NEAT1* RNPs gather to form massive paraspeckles.

Absolute quantitation of *NEAT1* molecules in paraspeckles

Our improved RNA extraction method inspired us to precisely quantitate *NEAT1* molecules in individual cells and paraspeckles, as these numbers would provide basic information regarding the molecular mechanism of RNA-constructed nuclear bodies. For this purpose, the typical shell and core substructures of paraspeckles were visualized by super-resolution microscopy in HeLa cells (Fig EV2A and B). Specifically, RNA-FISH was performed with three antisense RNA probes: two corresponding to the 5' and 3' termini of *NEAT1_2*, which labeled the shell of the paraspeckle (green signals in Fig EV2A), and one targeting the middle of *NEAT1_2*, which labeled the internal area of the paraspeckle (magenta signals in Fig EV2A) (Souquere *et al*, 2010; Mito *et al*, 2016; West *et al*,

Figure 2.

2016). All paraspeckles in a cell with an intact shell and core structure were collectively imaged by maximum intensity projection. The average number of paraspeckles per cell was 18.8 (Fig 2H). Next,

we prepared total RNA from 7×10^4 cells by the needle shearing method and performed RPA using a ^{32}P -labeled *in vitro* transcribed fragment of *NEAT1* antisense RNA. The RPA product and a known

Figure 2. Quantitative characterizations of the semi-extractable *NEAT1* arcRNA.

- A Improved *NEAT1* extraction detected with RPA. The molar ratio of two *NEAT1* isoforms is shown below the panels.
- B Assessment of semi-extractability in various mRNAs, pre-mRNAs, and lncRNAs. Total RNA was extracted by the conventional (shearing –) or improved method (shearing +) from HeLa cells to analyze various RNAs, with the exception of *XIST* in HEK293 cells. The introns in the *CLK1* mRNA precursor are spliced out, with the exception of introns 3 and 4; this intron-retaining form is retained in the nucleus until the cells are exposed to stress, at which time these two introns are removed (Ninomiya *et al*, 2011).
- C Semi-extractability of *IGS16* architectural RNA. HeLa cells were heat-shocked at 42°C to induce *IGS16* RNA expression.
- D RNA-FISH of *NEAT1_2* (green) in various human cell lines. Nuclei were stained with DAPI (blue).
- E Steady-state levels of *NEAT1_2* and *NEAT1_1* in various human cell lines. An example of RPA using total RNA extracted with the improved extraction method is shown in the upper panel. In the lower panel, *NEAT1* levels (expressed as relative molarities) are shown relative to *NEAT1_2* level in HeLa cells.
- F Semi-extractability of *NEAT1_2* in various human cell lines, quantitated by RT-qPCR.
- G Semi-extractability of *Neat1_2* in various mouse cell lines and luteal cells, quantitated by RT-qPCR.
- H Quantitation of *NEAT1* molecules per paraspeckle.

Data information: Bar graphs show means $\pm$ SD ($n = 3$). Scale bar, 10 μ m.

quantity of intact 32 P-labeled transcript were electrophoresed next to each other to determine the absolute level of *NEAT1* (Fig EV2C). Finally, the RNA levels were normalized using the yields of RNAs spiked in prior to TRIzol extraction. Absolute quantitation of *NEAT1* isoforms indicated that 122 and 994 molecules of *NEAT1_1* and *NEAT1_2*, respectively, were present in each HeLa cell (Fig 2H). Based on these observations, we estimate that each paraspeckle contains 6.5 *NEAT1_1* and 53 *NEAT1_2* molecules on average (Fig 2H). It should be noted that this calculation is predicated on the assumption that *NEAT1* molecules are exclusively localized in paraspeckles, as strongly suggested by our previous observation that *NEAT1* is uniquely enriched in the high-density nucleoplasmic fraction (Sasaki *et al*, 2009).

Genomewide identification of semi-extractable RNAs

In light of the entrapment of *NEAT1_2* in the protein fraction during conventional RNA extraction, suggesting the presence of unusually strong RNA–protein interactions (Fig 1C), and the observation of semi-extractability of known arcRNAs such as *NEAT1_2* and *IGS* (Figs 1B and 2C), we reasoned that identification of novel semi-extractable RNAs could reveal as-yet-unknown nuclear body-associated arcRNAs. Previous work suggested that the chromosome-associated CoT-1 RNA, whose major constituents are likely derived from *LINE1* elements, is also semi-extractable (Hall *et al*, 2014). We confirmed that *LINE1* RNA was semi-extractable (~twofold increase upon needle shearing; Table 1), and an antisense probe for this RNA detected multiple nucleoplasmic foci (Fig EV3A). In addition, *Gomafu* lncRNA, which constitutes a nuclear body in mouse brain cells (Sone *et al*, 2007), was also semi-extractable (2.5-fold increase upon needle shearing) (Fig EV3B). Although *Gomafu* and *LINE1* have not been formally shown to be arcRNAs, these results further supported the idea that semi-extractability is associated with arcRNAs that constitute nuclear bodies.

To search for additional semi-extractable RNAs, which represent candidates for novel arcRNAs, we performed next-generation RNA sequencing analysis of RNA samples extracted from HeLa cells using the conventional and needle shearing protocols, and then compared the results to identify transcripts whose read numbers were higher in the needle-sheared sample (Fig 1A). We selected the top 49 semi-extractable RNA candidates, whose reads were > 3.5 -fold ($\log_2$ fold > 1.8) higher in the needle-sheared samples (Table 1). RT-qPCR confirmed that 45 of these RNAs were semi-extractable RNAs (Table 1). To form membraneless subcellular bodies, a scaffold

molecule (whether RNA or protein) must be present at a high concentration (Gilks *et al*, 2004; Shevtsov & Dundr, 2011). Thus, to acquire arcRNA candidates, we selected the top ten most abundant RNAs based on RT-qPCR C_p values and RNA-seq reads (Table 1). *NEAT1* and *LINE1* were present among the top ten RNAs, confirming that known arcRNAs and foci-forming RNAs were indeed included in our selection. The other eight RNAs consisted of four groups (Fig 3A): first, known lncRNAs including *LINC00473*, *PVT1*, and *CCAT1*; second, an uncharacterized antisense transcript of *PZP* (*PZP antisense*, *PZP-AS*); third, mRNA precursors of *PDE3A* and *ADARB2*; and fourth, uncharacterized RNAs from primate-specific repetitive sequences, including *SVA_D* (Wang *et al*, 2005) and sequences adjacent to low-copy *NPIP* repeats (*Peri-NPIP*) (Marques-Bonet & Eichler, 2009). RT-qPCR confirmed that needle shearing yielded an approximately two- to threefold increase in extraction of these RNAs (Fig 3B, upper graphs). Interestingly, although spliced forms of *LINC00473*, *PVT1*, *CCAT1*, *PZP-AS*, *PDE3A*, and *ADARB2* are registered in public databases, our improved RNA extraction method combined with RNA-seq demonstrated that the majority of these RNAs are unlikely to be spliced in HeLa cells (Fig 3A). Moreover, semi-extractability was observed only for the unspliced transcripts, whereas the spliced transcripts were extractable by conventional protocols (Fig 3A and C). By examining RNA-seq reads, we noticed that *NEAT1_2* has a minor spliced form lacking 794 nt from positions 20,360 to 21,153 (Fig EV3C). This spliced *NEAT1_2* was much more extractable than its unspliced counterpart (Fig 3D). However, CRISPR-mediated genomic deletion of a region overlapping with the *NEAT1_2* intron (position 20,156–22,647) in HAP1 cells, a human haploid cell line, preserved semi-extractability (Fig 3E). These results demonstrate that a splicing event, rather than removal of the specific *NEAT1* sequence, was negatively correlated with semi-extractability.

Exploiting semi-extractability to discover candidate architectural RNAs

We monitored the subcellular localizations of the eight new abundant semi-extractable RNAs (other than *NEAT1* and *LINE1*) by RNA-FISH (Figs 4A and EV4A). All eight formed distinct granule-like foci in the nucleus. *LINC00473*, *PVT1*, *CCAT1*, *PZP-AS*, *PDE3A*, *ADARB2*, and *Peri-NPIP* foci did not overlap with any marker proteins of known nuclear bodies (Figs 4B and EV5), indicating that these RNAs are components of nucleoplasmic substructures that have not yet been annotated. A subset of *SVA_D* foci overlapped with

Table 1. An Exploratory search for semi-extractable RNAs in HeLa cells.

RNA	Genomic region	log ₂ fold [("improved" FPKM)/("control" FPKM)]	RT-qPCR fold "Improved"/"control"	RT-qPCR C _p value
<i>USP3</i> pre-mRNA	15:63796792–63894627	3.97 = log ₂ (19.98/1.28)	1.86	25.4
<i>MAP2K4</i> pre-mRNA	17:11924134–12047499	3.91 = log ₂ (30.09/2.00)	1.69	26.8
<i>NEAT1</i> lncRNA ^a	11:65189925–65217667	3.70 = log ₂ (134.20/10.33)	32.30	18.5
<i>LINC00473</i> lncRNA ^a	6:166193756–166401734	3.65 = log ₂ (14.51/1.16)	2.61	22.9
<i>UBAC2</i> pre-mRNA	13:99853027–100038688	3.27 = log ₂ (11.53/1.20)	1.69	26.0
<i>DANT2</i> lncRNA ^a	X:115030507–115085504	3.08 = log ₂ (8.83/1.05)	2.75	29.0
<i>MEIS2</i> pre-mRNA	15:37181408–37393862	3.00 = log ₂ (10.30/1.28)	1.62	26.5
<i>PVT1</i> lncRNA ^a	8:128806760–129113503	2.99 = log ₂ (13.46/1.72)	2.01	23.8
<i>Peri-IPW locus</i> lncRNA ^a	15:25068793–25684237	2.97 = log ₂ (16.04/2.05)	2.97	26.1
<i>SIL1</i> pre-mRNA	5:138282408–138667360	2.96 = log ₂ (13.39/1.72)	2.74	28.2
<i>ZMYM4</i> pre-mRNA	1:35734192–35887659	2.79 = log ₂ (9.70/1.40)	2.92	26.6
<i>AK4P13</i> pre-mRNA	15:85734668–86293512	2.78 = log ₂ (8.02/1.17)	1.94	26.0
<i>ZFPM2</i> antisense 1 lncRNA ^a	8:105602960–107072962	2.66 = log ₂ (6.63/1.05)	1.98	24.8
<i>MIR100HG</i> lncRNA ^a	11:121899034–122293579	2.60 = log ₂ (6.76/1.12)	1.80	26.0
<i>PRH1-PRR4</i> read-through	12:10977558–11329814	2.60 = log ₂ (8.04/1.33)	1.76	26.2
<i>LINC-PINT</i> lncRNA ^a	7:130538478–131184451	2.56 = log ₂ (6.96/1.18)	2.03	25.5
<i>RABGAP1L</i> pre-mRNA	1:174084437–174964527	2.50 = log ₂ (8.52/1.51)	3.10	27.0
<i>SLCIA3</i> pre-mRNA	5:36606454–36725297	2.48 = log ₂ (8.45/1.52)	1.69	25.1
<i>LINE1</i> repeat RNA ^b	6:153552454–153668623	2.45 = log ₂ (9.42/1.72)	2.02	17.1
<i>FAM208B</i> pre-mRNA	10:5726703–5884095	2.43 = log ₂ (5.74/1.07)	2.07	25.3
<i>COL4A5</i> pre-mRNA	X:107682984–107940775	2.39 = log ₂ (10.88/2.07)	2.69	26.0
<i>RNF13</i> pre-mRNA	3:149478891–149942977	2.39 = log ₂ (6.99/1.33)	1.91	26.8
<i>ZRANB3</i> pre-mRNA	2:135809834–136483793	2.39 = log ₂ (6.22/1.19)	1.62	27.2
<i>5S rRNA pseudogene</i> lncRNA ^a	13:106705495–106895268	2.36 = log ₂ (14.28/2.78)	2.33	26.5
<i>EYA4</i> pre-mRNA	6:133561686–133853262	2.33 = log ₂ (6.58/1.30)	1.93	26.0
<i>CMSS1</i> pre-mRNA	3:99536670–99897447	2.31 = log ₂ (9.09/1.83)	2.00	27.5
<i>PDE3A</i> pre-mRNA	12:20514665–20837315	2.29 = log ₂ (23.76/4.84)	1.64	23.8
<i>RBMS3</i> pre-mRNA	3:29305684–30051927	2.29 = log ₂ (6.27/1.28)	2.47	27.9
<i>ADARB2</i> pre-mRNA	10:1227212–1781562	2.28 = log ₂ (20.08/4.14)	3.24	24.0
BTBD3 lncRNA	20:11871370–11923698	2.24 = log ₂ (9.82/2.07)	1.25	24.9
SLC7A11 pre-mRNA	4:138948575–139163670	2.21 = log ₂ (13.25/2.86)	1.28	27.0
<i>PALLD</i> pre-mRNA	4:169277329–169931440	2.19 = log ₂ (6.47/1.42)	2.56	25.4
<i>NIPBL</i> pre-mRNA	5:36876860–37068982	2.19 = log ₂ (4.80/1.05)	1.87	26.2
<i>SVA_D</i> repeat RNA ^b	X:122867009–122868339	2.17 = log ₂ (4.50/1.00)	1.94	20.6
<i>PZP</i> antisense RNA ^a	12:9296503–9360966	2.14 = log ₂ (11.64/2.64)	1.61	22.8
<i>C9orf41</i> pre-mRNA	9:77567883–77643339	2.11 = log ₂ (7.93/1.83)	1.86	27.3
TRIM5 pre-mRNA	11:5684424–5959849	2.10 = log ₂ (9.49/2.21)	1.33	27.3
<i>NKTR</i> pre-mRNA	3:42589460–42709186	2.08 = log ₂ (5.53/1.30)	1.91	25.8
<i>ASPH</i> pre-mRNA	8:61969716–62628262	2.08 = log ₂ (5.74/1.36)	1.96	26.6
<i>FTX</i> lncRNA ^a	X:73422668–73513400	2.04 = log ₂ (5.74/1.40)	2.18	26.7
<i>FGF2</i> pre-mRNA	4:123747862–123844123	2.02 = log ₂ (6.11/1.50)	1.62	30.1
<i>NLK</i> pre-mRNA	17:26368764–26531972	2.02 = log ₂ (5.54/1.37)	1.88	26.1
<i>APP</i> pre-mRNA	21:27252800–27548142	1.97 = log ₂ (9.87/2.52)	1.79	26.9
<i>CCAT1</i> lncRNA ^a	8:128121803–128122019	1.89 = log ₂ (3.78/1.02)	1.93	20.9
AP003900.6 lncRNA	21:11169665–11186935	1.89 = log ₂ (7.34/1.98)	1.14	20.7

Table 1 (continued)

RNA	Genomic region	log ₂ fold [("improved" FPKM)/("control" FPKM)]	RT-qPCR fold "Improved"/"control"	RT-qPCR C _p value
<i>IMMP2L</i> pre-mRNA	7:110303109–111202573	1.87 = log ₂ (3.85/1.05)	2.73	26.3
<i>DLEU</i> lncRNA ^a	13:50570023–51423145	1.87 = log ₂ (7.75/2.12)	2.29	26.1
<i>TSEN2</i> pre-mRNA	3:12525930–12705725	1.86 = log ₂ (17.11/4.71)	2.12	26.1
<i>Peri-NPIP</i> repeat RNA ^b	16:21845889–21930477	1.83 = log ₂ (18.54/5.23)	2.23	22.0

Forty-nine semi-extractable RNA candidates acquired by RNA-seq are listed. Of these RNAs, RT-qPCR analyses confirmed that 45 were semi-extractable; the four other candidates are shown in bold. The numbers in the genomic region are the output of computational analyses of RNA-seq data; the precise ends of the transcripts need to be experimentally determined.

^alncRNAs.

^brepetitive sequence-derived RNAs.

paraspeckles labeled with *NEAT1* or *EWSR1* (orange arrows in Figs 4A and EV4B), whereas the remaining *SVA_D* formed foci distinct from other nuclear bodies (yellow arrowheads in Figs 4A and EV4B, and EV5). The semi-extractability of *SVA_D* and the integrity of its foci were unchanged upon paraspeckle disruption by *NEAT1* knockdown (yellow arrowheads in Fig EV4B and C), suggesting that *SVA_D* foci form distinct nuclear structures that partially overlap with paraspeckles.

To confirm that the granule-like foci of semi-extractable RNAs are not simply sites of transcription, we determined the subcellular localizations of several semi-extractable RNAs (*LINC00473*, *CCAT1*, or *ADARB2*) along with intron-containing precursor transcripts of neighboring genes (*PDE10A*, *MYC*, or *LARP4B*, respectively), which should be present at their transcription sites (Fig 4C). Granule-like foci of *LINC00473*, *CCAT1*, and *ADARB2* were located not only at the transcription sites (orange arrows in Fig 4C), but also outside of transcription sites (yellow arrowheads in Fig 4C); this pattern is reminiscent of the paraspeckle, which is assembled at *NEAT1_2* transcription site and buds off or splits away as the body enlarges (Mao et al, 2011).

Semi-extractability of *NEAT1_2* requires association with paraspeckle proteins

Next, we sought to gain insights into the mechanisms that connect semi-extractability with formation of subnuclear granule-like structures. Since the data above suggested that protein binding made *NEAT1* semi-extractable, we investigated whether *NEAT1*–PSP interactions underlie the semi-extractability of *NEAT1_2* as a model case. To dissociate PSPs from *NEAT1*, we employed 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), a reversible inhibitor of RNA polymerase II. DRB induces rapid dissociation of PSPs from *NEAT1*, resulting in relocation of PSPs to perinucleolar cap structures by an unknown mechanism (Fox et al, 2005; Sasaki et al, 2009; Naganuma et al, 2012). As seen in Fig 5A (separate color channel images in Appendix Fig S2A), 1 h after DRB addition to the medium, the majority of *NEAT1* signals (green) and PSP (*EWSR1* or *FUS*) signals (magenta) segregated. The *NEAT1* signals became diffuse in the nucleus, whereas PSP signals relocated to the perinucleolar caps; the redistribution was more obvious after 4 h (Fig 5A). When DRB was removed from the medium after 4 h, paraspeckles started to become visible 2 h later (Fig 5A) and mostly recovered to their original pattern after 4 h. UV crosslinking immunoprecipitation (CLIP) demonstrated that co-precipitation of *NEAT1_2* with *FUS* markedly decreased when cells were treated with DRB for 1 h (Fig 5B and

Appendix Fig S2B). Specific co-precipitation of *NEAT1_2* with *FUS* depended on UV irradiation, indicating that we were detecting a direct interaction of *FUS* with *NEAT1_2*, which rapidly dissociated upon DRB treatment. Finally, we monitored the extractability of *NEAT1* over the course of paraspeckle disintegration following DRB treatment and reformation of paraspeckles following DRB removal. In control cells, needle shearing increased the yield of extracted *NEAT1* by ~20-fold (0 h in Fig 5C and Appendix Fig S2C). Treatment with DRB rapidly increased the yield of extracted *NEAT1_2* in the absence of needle shearing, and the ratio of extraction (with needle shearing/without shearing) dropped to ~twofold from > 20-fold within 1 h after addition of DRB (lower graph in Fig 5C). *NEAT1_2* levels in shearing+ samples gradually decreased for 4 h because *NEAT1_2* was gradually degraded during transcription arrest (upper graph in Fig 5C). Once DRB was removed from the medium, *NEAT1* levels started to increase, and easily recognizable paraspeckles reformed within 4 h (Fig 5A). Over the course of this process, the semi-extractability of *NEAT1_2* also gradually recovered, reaching its previous levels by 3 h after DRB removal (lower graph in Fig 5C). Taken together, these findings indicate that the *NEAT1*–PSP interaction, which is closely linked to paraspeckle integrity, is responsible for the semi-extractability of *NEAT1_2*. When cells were treated with DRB, semi-extractability was lost not only in *NEAT1_2* but also in other novel abundant semi-extractable RNAs (Fig EV4D). This suggests that the semi-extractability of these novel RNAs might also be dependent on the proteins that relocate to nucleoli upon transcriptional inhibition (Shav-Tal et al, 2005).

Treatment of cells with the proteasome inhibitor MG132 upregulates *NEAT1* expression, leading to elongation and proliferation of paraspeckles (Hirose et al, 2014b). Indeed, MG132 treatment increased the level of *NEAT1_2* by almost threefold (Fig 5D) but did not markedly change its extractability. This suggests that paraspeckle assembly, rather than paraspeckle elongation and proliferation, is the critical step for conferring semi-extractability on *NEAT1_2*.

The prion-like domain of *FUS* is required for the semi-extractability of *NEAT1_2*

Next, we sought to identify the PSPs involved in semi-extractability of *NEAT1_2*. To this end, we investigated whether the extractability of *NEAT1_2* was affected by depletion of specific PSPs. For this experiment, we used the CRISPR-Cas9 system to generate knockouts (KOs) of specific PSPs in HAP1 cells. We generated KOs of *FUS*, *DAZAP1*, and *HNRNPH3*, as well as a double KO of *BRG1* and *BRM*

Figure 3.

Figure 3. Genomewide identification of semi-extractable RNAs.

- A Exploratory search for novel semi-extractable RNAs in HeLa cells. RNA was extracted from cells using the conventional (control) or improved method (sheared). Examples of RNA-seq data of semi-extractable RNAs are shown. *GAPDH* mRNA is shown as a negative control. The semi-extractable regions of *CCAT1*, *PZP-AS*, and *Peri-NPIP* are preceded or followed by extractable regions, which may be due to an overlap of two transcripts or alternatively processed transcripts.
- B Confirmation of semi-extractability by RT-qPCR. RNAs present in the aqueous and protein phases during the RNA extraction procedure were analyzed. *GAPDH* mRNA is shown as a control.
- C, D Negative correlation between splicing and semi-extractability. Unspliced and spliced transcripts of novel abundant semi-extractable RNAs (C) or *NEAT1_2* (D) in HeLa cells were quantitated with RT-qPCR.
- E Semi-extractability of *NEAT1_2* in HAP1 cells upon CRISPR-mediated genomic deletion of regions overlapping with its intronic sequences.
- Data information: Bar graphs show means $\pm$ SD ($n = 3$).

(to functionally deplete SWI/SNF complexes), to deplete each of the four proteins required for paraspeckle assembly from *NEAT1* sub-RNPs (Naganuma *et al*, 2012; Kawaguchi *et al*, 2015). We also generated KO cells of TAF15, TDP43, or EWSR1, which contain PLDs (Appendix Fig S3A). PSPs such as NONO, SFPQ, RBM14, and HNRNPK are all required for *NEAT1_2* accumulation (Naganuma *et al*, 2012) and therefore could not be knocked out in our analysis of *NEAT1_2* extractability. Consistent with the results of our previous RNAi knockdown experiments, HAP1 cell lines lacking one of the four proteins required for paraspeckle assembly contained reduced numbers of paraspeckles; by contrast, KO of any of the three PLD-containing proteins barely affected paraspeckle appearance (Fig 6A, upper panels; enlarged in Appendix Fig S3B).

Among the PSPs we investigated, KO of FUS caused a dramatic increase in the extractability of *NEAT1_2*, decreasing the ratio of *NEAT1_2* extracted with vs. without needle shearing (lower graph in Fig 6A). This finding indicates that FUS makes a major contribution to the semi-extractability of *NEAT1_2*. Notably, KO of FUS in HAP1 cells decreased the semi-extractability of *PZP-AS* RNA, but did not markedly decrease the semi-extractability of *LINC00473* or *Peri-NPIP* RNA (Fig EV4E), suggesting the existence of other proteins that confer semi-extractability on different RNAs. Moreover, mouse embryonic fibroblasts (MEFs) prepared from *Fus* KO mice exhibited remarkable *Neat1_2* extractability (Fig 6B and Appendix Fig S3C), indicating that the role of the FUS association is common to both human and mouse. To validate this idea, we performed a FUS rescue experiment in *FUS* KO HAP1 cells (Fig 6C and Appendix Fig S3D). The introduction of *Fus* cDNA into *FUS* KO cells, in which functional FUS protein was produced at lower levels (47%) than in WT HAP1 (Appendix Fig S3E), partially rescued the semi-extractability of *NEAT1_2*. However, a FUS mutant lacking the PLD failed to rescue, indicating that FUS, as a protein required for paraspeckle assembly, contributes to the semi-extractability of *NEAT1_2* through its PLD.

Discussion

Although recent human transcriptome analyses revealed the presence of tens of thousands of lncRNAs, their functions remain largely unknown. Thus, molecular characterization and functional classification of these lncRNAs are important challenges for the next decade. A major obstacle to molecular characterization of lncRNAs is the fact that the lncRNA sequences themselves provide few clues regarding their functions. Accordingly, one strategy for studying the functions of lncRNAs is to first identify a detectable or tangible property that corresponds to a class of lncRNAs with a common function, and then use that property to enrich and identify other

members of the class. One class of lncRNAs, termed arcRNAs, functions as the essential molecular scaffolds or platforms of nuclear bodies (Hirose *et al*, 2014a; Chujo *et al*, 2016). All known arcRNAs are temporarily upregulated under specific cellular stresses, at particular developmental stages, or in various disease conditions, and they serve as molecular “sponges” to sequester specific regulatory proteins to modulate gene expression (Chujo *et al*, 2016). The data presented in this manuscript demonstrate that identification of semi-extractable RNAs is a powerful and relatively convenient method for identifying RNAs that localize in subnuclear structures, providing good candidates for arcRNAs. To become fully qualified as an arcRNA, a candidate RNA needs to meet two criteria: (i) The RNA is enriched in a specific nuclear body, and (ii) removal of the RNA abolishes the nuclear body, causing nuclear body marker proteins to disperse or exhibit an altered subcellular localization. Thus, to validate eight new abundant semi-extractable RNAs as genuine arcRNAs, future studies will need to identify body-specific marker proteins and determine whether removal of the RNAs abolishes the focal localizations of these proteins. The marker proteins bound to the semi-extractable RNAs in the cell may be identified using technologies such as capture hybridization analysis of RNA targets combined with mass spectrometry (ChIRP-MS or CHART-MS; West *et al*, 2014; Chu *et al*, 2015).

In this study, we discovered eight abundant semi-extractable RNAs with subnuclear granule-like localization. Because the eight RNAs were not efficiently solubilized by a standard extraction method using phenol and guanidinium thiocyanate (Fig 3A and B), like *NEAT1*, they are likely to form strong and extensive RNA–protein and protein–protein interactions in the cell. Among these RNAs, *LINC00473*, *PVT1*, and *CCAT1* have been reported to play roles in specific tissue development and cancer progression, although their specific molecular functions remain poorly characterized (Tseng *et al*, 2014; Chen *et al*, 2016; Liang *et al*, 2016; McClelland *et al*, 2016). *NEAT1*, which is also involved in cancer progression and tissue development (Chakravarty *et al*, 2014; Nakagawa *et al*, 2014; Standaert *et al*, 2014; Adriaens *et al*, 2016), sequesters various RNA-binding proteins, transcription factors, and specific RNAs (Prasanth *et al*, 2005; Chen & Carmichael, 2009; Hirose *et al*, 2014b; Imamura *et al*, 2014; West *et al*, 2016) and/or associates with specific active chromosomal loci (West *et al*, 2014). Therefore, under the cellular conditions in which specific semi-extractable RNAs are highly expressed, it would be intriguing to purify the complexes containing specific semi-extractable RNAs by ChIRP- or CHART-based methods to characterize the associated proteins, RNA, and/or DNAs.

Peri-NPIP RNA is transcribed from genomic regions adjacent to the *NPIP* family of primate-specific low-copy repeats. In humans, the *NPIP* genes and their surrounding loci have been duplicated

Figure 4. Semi-extractable RNA-seq identifies novel architectural RNA candidates.

A RNA-FISH of the eight novel abundant semi-extractable RNAs. Orange arrows, *SVA_D* puncta that overlapped with *NEAT1*; yellow arrowheads, distinct *SVA_D* puncta.

B RNA-FISH of novel semi-extractable RNAs and immunofluorescence of known nuclear body marker proteins. Body names are indicated under the marker protein names. As an example, *PZP-AS* RNA is shown. Images of seven other novel abundant semi-extractable RNAs are shown in Fig EV5.

C Dual RNA-FISH of semi-extractable RNAs and their neighboring gene precursor RNAs. *LINC00473*, *CCAT1*, or *ADARB2* gene loci are neighbored by *PDE10A*, *MYC*, and *LARP4B* gene loci, respectively. Precursor transcripts of *PDE10A*, *MYC*, and *LARP4B* genes were visualized using the intronic probes to show the approximate locations of transcription sites of *LINC00473*, *CCAT1*, or *ADARB2*, respectively. Orange arrows: foci of semi-extractable RNAs overlapping with the foci of neighboring gene mRNA precursor (transcription sites); yellow arrowheads: foci of semi-extractable RNAs outside of neighboring gene mRNA precursor foci. Semi-extractable RNAs were visualized with conventional FISH, and adjacent gene precursor RNAs were visualized by FISH combined with tyramide signal amplification.

Data information: Scale bars, 10 μ m.

into 17 blocks, 16 of which are on chromosome 16 (Marques-Bonet & Eichler, 2009). Indeed, a BLAT search of *Peri-NPIP* RNA in Fig 3A retrieved nine highly similar sequences exclusively from chromosome 16. The *SVA_D* repeat is another primate-specific

repetitive sequence (Wang *et al*, 2005). It is tempting to speculate that subnuclear granule-associated RNAs transcribed from such primate-specific repetitive sequences contribute to primate-specific gene regulation.

Figure 5. Dissociation of paraspeckle proteins (PSPs) from *NEAT1_2* abolishes the semi-extractable feature of *NEAT1_2*.

A–C Dissociation of PSPs from *NEAT1* upon DRB treatment, and reformation of paraspeckles following release from DRB. (A) *NEAT1_2* FISH combined with immunofluorescence of PSP (EWSR1 or FUS). Scale bar, 10 μ m. Separate images of *NEAT1_2*, PSP, and DAPI are shown in Appendix Fig S2A. (B) UV-CLIP of FUS after 1 h of DRB treatment. DMSO was used as a vehicle control. (C) *NEAT1_2* extraction level (upper panel) and semi-extractability (lower panel). Full results including vehicle (DMSO) controls are shown in Appendix Fig S2C.

D The effect of *NEAT1* upregulation and paraspeckle elongation on *NEAT1* semi-extractability. HeLa cells were exposed to 5 μ M MG132 for 6 h to increase expression of *NEAT1*.

Data information: Bar graphs show means $\pm$ SD ($n = 3$).

Figure 6. The prion-like domain of RNA-binding protein FUS confers the property of semi-extractability on NEAT1_2.

A Knockout of various PSPs in human HAP1 cells. Upper panel: The number of paraspeckles per cell (3.4 in WT HAP1 cell) was reduced to 1.3 in FUS KO, 2.6 in DAZAP1 KO, 2.5 in HNRNPH3 KO, and 1.7 in BRG1/BRM-double KO ($n = 20$ cells each). In EWSR1 KO cells, PSPC1 was used as a paraspeckle marker protein. Lower panel: Comparison of NEAT1_2 semi-extractability (WT: $n = 8$, FUS KO: $n = 6$, Mann–Whitney test).

B Fus KO mouse embryonic fibroblasts were used as in (A).

C Complementation of FUS KO HAP1 cells with wild-type FUS, ΔPLD FUS, or EGFP. Constructs were stably introduced into FUS KO HAP1 cells by lentiviral infection, and single clones were selected to assess NEAT1 semi-extractability as in (A) ($n = 3$, Sidak's multiple comparison test). *** $P < 0.0001$; n.s., not significant ($P = 0.09$).

Data information: Scale bars, 10 μ m. Bar graphs show means $\pm$ SD ($n = 3$) unless otherwise indicated.

Table 1 lists many semi-extractable RNAs that we did not further investigate due to their relatively low expression levels in HeLa cells. These semi-extractable RNAs may be highly expressed in other cells, in which they might function as arcRNAs. For example, SPA lncRNA (transcribed from *peri-IPW* locus) and FTX lncRNA (Table 1) are upregulated in ES cells and form nuclear foci. SPA lncRNA is expressed at a high level (one-third the amount of MALAT1) in human ES cells and sequesters several RBPs to affect splicing (Wu et al, 2016). Ftx is a conserved lncRNA upregulated in female ES cells that affects XIST RNA expression (Chureau et al, 2011).

Because lncRNA expression profiles are generally highly tissue-specific, it might be possible to discover new arcRNAs by exploratory search of abundant semi-extractable RNAs in different cell lines, tissues, and developmental stages.

The FUS PLD facilitates protein–protein interactions, leading to cellular paraspeckle formation and *in vitro* hydrogel formation (Kato et al, 2012; Burke et al, 2015; Hennig et al, 2015). Although FUS, DAZAP1, HNRNPH3, TAF15, TDP-43, and EWSR1 all have a PLD, only FUS KO cells exhibited reduced NEAT1_2 semi-extractability. This observation indicates that the FUS PLD and RNA-binding

domain are each likely to form unusually tenacious molecular interactions even in the presence of an RNA extraction reagent containing phenol and guanidinium thiocyanate. Considering (i) the entrapment of *NEAT1_2* in the protein fraction during conventional RNA extraction (Fig 1C) and (ii) the pivotal role of FUS PLD to the semi-extractability of *NEAT1_2* (Fig 6C), semi-extractability likely reflects tenacious RNA–protein and protein–protein interactions, which are the driving forces of nuclear body formation. Even though FUS binds to a variety of RNA species, among its binding partners extreme semi-extractability was specifically observed for *NEAT1_2* (Fig 2B). This may be due to a subset of PSPs that we did not test because they are required for *NEAT1* accumulation (Naganuma *et al*, 2012). These proteins include SFPQ, NONO, and RBM14, which also bear PLDs. In particular, the PLD of RBM14 is required for paraspeckle formation (Hennig *et al*, 2015). Therefore, FUS and these proteins might co-operatively contribute to the formation of tenacious protein–protein interactions in paraspeckles, giving rise to extreme semi-extractability of *NEAT1_2*. Although FUS is ubiquitously expressed in multiple types of cells (Appendix Fig S1B), the degree of *NEAT1_2* semi-extractability differed between cell lines (Fig 2F). This difference might also be due to different expression levels of other PLD-bearing PSPs in different cell lines. The dramatic increase in extraction of *NEAT1* following needle shearing suggests that *NEAT1* RNP macrocomplexes in AGPC reagent may behave as phase-separated liquid droplets, which are distorted in response to shear forces (Brangwynne *et al*, 2009).

Interestingly, semi-extractable RNAs lost this feature when they were spliced (Fig 3A, C and D). Moreover, the major nuclear isoforms of previously characterized arcRNAs are not spliced (Watanabe & Yamamoto, 1994; Prasanth *et al*, 2000; Sasaki *et al*, 2009; Biamonti & Vourc'h, 2010; Audas *et al*, 2012). Together, these reports and our data strongly suggest that, unlike mRNAs, for which splicing serves as an mRNA maturation step, subnuclear granule-forming RNAs must remain unspliced to exert their nuclear functions. We also showed that genomic deletion of the corresponding intronic region in *NEAT1* barely affected extractability, suggesting that removal of this specific sequence was not responsible for changing extractability (Fig 3E). During mRNA maturation, splicing results in remodeling of mRNP complexes to recruit factors involved in mRNA export and/or quality control (Singh *et al*, 2015). This raises the intriguing possibility that co-transcriptional arcRNP formation antagonizes association of splicing factors, allowing the primarily assembled arcRNP complex or nuclear body to become functional. Also, it should be noted that incompletely spliced mRNA precursors in general were not semi-extractable (Fig 2B), strongly suggesting that semi-extractable RNAs are not merely RNAs retained in the nucleus due to defects in splicing. *PDE3A* and *ADARB2* are protein-coding genes, but we found that substantial proportions of the transcripts from these genes were unspliced and semi-extractable, and exhibited subnuclear granule-like localizations (Figs 3A and C, and 4A). Thus, we cannot exclude the possibility that even protein-coding genes may generate transcripts that serve specific roles in the nucleus when they remain unspliced.

The semi-extractable feature of arcRNAs is achieved by RNA–protein and protein–protein interactions between arcRNA and PLD-containing RNA-binding proteins. Humans possess > 70 PLD-bearing RNA-binding proteins, and it can be postulated that these proteins may select and interact with corresponding

arcRNAs to form distinct nuclear bodies, as well as make them semi-extractable.

Materials and Methods

RNA extraction

TRI reagent (MRC) was added to cells at a ratio of 1 ml of reagent to 1×10^6 cells. Cell lysates in TRI reagent were passed 100 times through a 20-gauge needle or heated using a Thermomixer (Eppendorf) at 55°C for 10 min with 1,000 rpm agitation. Subsequently, total RNA was extracted according to the manufacturer's instructions. RNA extraction from the protein phase obtained as the intermediate insoluble fraction during standard RNA preparation with TRI reagent was carried out as follows. First, the insoluble protein phase was gently washed twice with isopropanol and then twice with chloroform. Subsequently, the protein phase samples were air-dried, resuspended in proteinase K buffer [50 mM Tris–HCl (pH 7.5), 10 mM EDTA, 0.5% SDS, 200 µg/ml proteinase K (Roche)], and digested at 37°C for 30 min; the digestion reaction was stopped by the addition of phenol–chloroform isoamyl alcohol (pH 5.2, Nacalai Tesque), and RNA was recovered by ethanol precipitation.

Cell culture

HeLa, T24, MRC5, MCF7, A549, HEK293, Huh7, HCT116, U2OS, A9, NIH3T3, Neuro2a, and MEF cells were grown in DMEM supplemented with 10% fetal bovine serum. HAP1 cells were cultured as previously described (Kawaguchi *et al*, 2015). *FUS* KO and *BRG1* KO HAP1 cells were obtained from Horizon Genomics, and *Fus* KO MEF was generated as described (Fujii & Takumi, 2005). Transcriptional inhibition experiments were performed in culture medium containing 100 µM DRB.

Knockout and complementation in HAP1 cells

CRISPR/Cas9 guide RNAs were selected from GeCKO v2 libraries (Sanjana *et al*, 2014) and cloned into the *BbsI* site of pX330 (Addgene plasmid #42230) using the DNA oligos listed in Appendix Table S1. The cloned plasmids and one-tenth quantity of pcDNA6/TR (which contains the blasticidin resistance gene) were co-transfected into HAP1 cells (Horizon Discovery) using a Nucleofector device (Lonza). For enrichment of transfectants, the cells were cultured in medium containing 20 µg/ml blasticidin (InvivoGen) for 3 days, starting 1 day after transfection. Subsequently, cells were diluted into 96-well plates for selection of single clones. The selected clones were lysed and subjected to PCR to amplify the genomic regions flanking the guide RNA target sites. To detect indels, T7 endonuclease I (NEB) cleavage assay was performed on the amplified PCR products. The indel-positive clones were further confirmed by sequencing. Finally, the absence of protein expression was confirmed by immunoblotting.

Preparation of lentiviruses expressing full-length FUS and mutant molecules is described elsewhere (West *et al*, 2016). To infect HAP1 cells, 2×10^4 cells were mixed with 1 ml of supernatant from lentivirus-producing cells and incubated for 24 h. Subsequently, cells

were diluted into 96-well plates for selection of single clones. Protein expression was confirmed by immunoblotting.

RT-qPCR

Total RNA (1 µg) was reverse-transcribed using the QuantiTect reverse transcription kit (Qiagen). RNA was reverse-transcribed with random primers, with the exception of *KCNQ1OT1*, which was reverse-transcribed using a strand-specific primer. Primers were designed using Primer3 and are listed in Appendix Table S1. Aliquots of cDNA were amplified by quantitative PCR on a Light-Cycler 480 using the SYBR Green I Master reagent (Roche). For RT-qPCR of total RNAs, 18S rRNA levels were used for normalization.

RPA

RNase protection assays were performed using the RPAIII kit (Ambion). Total RNA (5 µg) was hybridized with ³²P-labeled anti-sense RNA probes (Naganuma *et al*, 2012) synthesized using T7 RNA polymerase (TaKaRa). RNase A/T1 digestion was performed to eliminate un-hybridized single-stranded RNA probes. Samples were separated on 6% PAGE gels containing 7 M urea. Radioactive RNA bands were visualized and quantitated with an FLA-7000 analyzer (Fujifilm). For absolute quantitation of *NEAT1* or *U12*, cellular RNA extraction efficiency was determined by addition of *E. coli* total RNA before or after RNA extraction, followed by quantitation of *E. coli gapA* mRNA after extraction. Before hybridization, the number of probe RNA molecules was calculated from the radioactivity count, the manufacture date of ³²P-UTP, the date of probe transcription, and the half-life of ³²P. A fraction of ³²P-labeled RNA was electrophoresed next to the RPA samples, and band intensities were compared to calculate the number of probes that hybridized to target RNA. Probe hybridization efficiency was determined by adding or not adding a defined amount of *in vitro* transcribed ³²P-labeled sense RNA prior to hybridization. Finally, cell number, cellular RNA extraction efficiency, probe number, RPA hybridization efficiency, and RPA band intensities were used to calculate the number of target RNA molecules per cell.

NGS RNA sequencing

RNA quality was assessed using NanoDrop (Thermo) and Bio-analyzer (Agilent) instruments. First, ribosomal RNA was depleted using the Ribo-Zero Gold Kit (Epicentre). Libraries were prepared using Illumina TruSeq Stranded RNA Sample Prep Kit. Samples were subjected to 100-bp paired-end sequencing (1.7×10^8 reads per sample) on an Illumina HiSeq 2500. Fluorescence images were processed into sequences using the analysis pipeline supplied by Illumina. Adapters were trimmed using Cutadapt (ver. 1.1), and reads were mapped to the human reference genome (hg19) using TopHat (ver. 2.0.9). Transcripts were assembled with Cufflinks and Cuffmerge, and expression analysis was performed using Cuffdiff. The GEO accession number of the RNA-seq data is GSE80589.

RNA-FISH and immunofluorescence

RNA-FISH and immunofluorescence were performed as previously described (Naganuma *et al*, 2012) unless otherwise indicated.

RNA-FISH of *PDE10A*, *MYC*, and *LARP4B* intronic regions was followed by anti-DIG tyramide signal amplification (PerkinElmer). Details of probes and antibodies are described in Appendix Tables S1 and S2. Confocal images were acquired using FLUOVIEW FV1000 (Olympus). To determine the number of paraspeckles per cell, RNA-FISH and super-resolution microscopy were performed as previously described (Mito *et al*, 2016) using ELYRA PS.1 (Zeiss). One hundred cells were analyzed, and foci larger than 0.3 µm and containing signals from probes targeting the 5', middle, and 3' regions of *NEAT1* were counted as paraspeckles.

Isolation of RNA-protein complexes

Cells were trypsinized and washed three times with PBS, and 5×10^6 cells were resuspended in 500 µl of PBS in 6-well plates. Cells were irradiated on ice with 254 nm UV light (400 mJ/cm²), collected, and lysed in RIPA buffer [150 mM NaCl, 25 mM Tris-HCl (pH 7.5), 1% NP-40, 1% Na deoxycholate, 0.1% SDS, protease inhibitor cocktail (Roche) and 20 U/ml SUPERaseIn (Ambion)]. Cell lysates were sonicated for 100 s and cleared by centrifugation at 10,000 g for 10 min. FUS antibody-bound Dynabeads were incubated with cell lysate for 3 h at 4°C and washed three times with RIPA buffer.

Expanded View for this article is available online.

Acknowledgements

The authors thank C. Fujikawa and A. Kubota for technical support, the members of the Hirose laboratory for valuable discussions, and the NIKON Imaging Center for assistance with microscopy. This research was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan [to TH (26113002 and 26291001), TY (15K18474), and TT (16H06316 and 16K13110)], Grant for Joint Research Program of the Institute for Genetic Medicine, Hokkaido University (to SN and TH), and the Japan Society for the Promotion of Science fellowship (to TC).

Author contributions

TC and TH conceived of and designed this study. TC conducted most of the experiments. TY, TK, SK, and TT generated knockout cell lines. SN performed luteal cell preparation, lentiviral infection, and super-resolution microscopy. TH and TC wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, Boeckx B, Wijnhoven PW, Radaelli E, Vermi W, Leucci E, Lapouge G, Beck B, van den Oord J, Nakagawa S, Hirose T, Sablina AA, Lambrechts D, Aerts S, Blanpain C *et al* (2016) p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat Med* 22: 861–868
- Audas TE, Jacob MD, Lee S (2012) Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. *Mol Cell* 45: 147–157
- Biamonti G, Vourc'h C (2010) Nuclear stress bodies. *Cold Spring Harb Perspect Biol* 2: a000695

- Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Julicher F, Hyman AA (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324: 1729–1732
- Burke KA, Janke AM, Rhine CL, Fawzi NL (2015) Residue-by-residue view of *in vitro* FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol Cell* 60: 231–241
- Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, MacDonald TY, Fontugne J, Erho N, Vergara IA, Ghadessi M, Davicioni E, Jenkins RB, Palanisamy N, Chen Z, Nakagawa S et al (2014) The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun* 5: 5383
- Chen LL, Carmichael GG (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* 35: 467–478
- Chen Z, Li JL, Lin S, Cao C, Gimbrone NT, Yang R, Fu DA, Carper MB, Haura EB, Schabath MB, Lu J, Amelio AL, Cress WD, Kaye FJ, Wu L (2016) cAMP/CREB-regulated LINC00473 marks LKB1-inactivated lung cancer and mediates tumor growth. *J Clin Invest* 126: 2267–2279
- Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161: 404–416
- Chujo T, Yamazaki T, Hirose T (2016) Architectural RNAs (arcRNAs): a class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim Biophys Acta* 1859: 139–146
- Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, Rougeulle C (2011) Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum Mol Genet* 20: 705–718
- Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 33: 717–726
- Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI (2002) Paraspeckles: a novel nuclear domain. *Curr Biol* 12: 13–25
- Fox AH, Bond CS, Lamond AI (2005) P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol Biol Cell* 16: 5304–5315
- Fujii R, Takumi T (2005) TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *J Cell Sci* 118: 5755–5765
- Gilks N, Kedersha N, Ayodele M, Shen L, Stoecklin G, Dember LM, Anderson P (2004) Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol Biol Cell* 15: 5383–5398
- Hall LL, Carone DM, Gomez AV, Kolpa HJ, Byron M, Mehta N, Fackelmayer FO, Lawrence JB (2014) Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* 156: 907–919
- Hennig S, Kong G, Mannen T, Sadowska A, Kobelke S, Blythe A, Knott GJ, Iyer KS, Ho D, Newcombe EA, Hosoki K, Goshima N, Kawaguchi T, Hatters D, Trinkle-Mulcahy L, Hirose T, Bond CS, Fox AH (2015) Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J Cell Biol* 210: 529–539
- Hirose T, Mishima Y, Tomari Y (2014a) Elements and machinery of non-coding RNAs: toward their taxonomy. *EMBO Rep* 15: 489–507
- Hirose T, Virnecchi G, Tanigawa A, Naganuma T, Li R, Kimura H, Yokoi T, Nakagawa S, Benard M, Fox AH, Pierron G (2014b) NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* 25: 169–183
- Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K, Kato A, Kawaguchi Y, Sato H, Yoneda M, Kai C, Yada T, Suzuki Y, Yamada T, Ozawa T, Kaneki K, Inoue T, Kobayashi M, Kodama T, Wada Y et al (2014) Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol Cell* 53: 393–406
- Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149: 753–767
- Kawaguchi T, Tanigawa A, Naganuma T, Ohkawa Y, Souquere S, Pierron G, Hirose T (2015) SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies. *Proc Natl Acad Sci USA* 112: 4304–4309
- King OD, Gitler AD, Shorter J (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res* 1462: 61–80
- Liang XH, Deng WB, Liu YF, Liang YX, Fan ZM, Gu XW, Liu JL, Sha AG, Diao HL, Yang ZM (2016) Non-coding RNA LINC00473 mediates decidualization of human endometrial stromal cells in response to cAMP signaling. *Sci Rep* 6: 22744
- Mannen T, Yamashita S, Tomita K, Goshima N, Hirose T (2016) The Sam68 nuclear body is composed of two RNase-sensitive substructures joined by the adaptor HNRNPL. *J Cell Biol* 214: 45–59
- Mao YS, Sunwoo H, Zhang B, Spector DL (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* 13: 95–101
- Marques-Bonet T, Eichler EE (2009) The evolution of human segmental duplications and the core duplicon hypothesis. *Cold Spring Harb Symp Quant Biol* 74: 355–362
- McClelland ML, Mesh K, Lorenzana E, Chopra VS, Segal E, Watanabe C, Haley B, Mayba O, Yaylaoglu M, Gnad F, Firestein R (2016) CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest* 126: 639–652
- Mito M, Kawaguchi T, Hirose T, Nakagawa S (2016) Simultaneous multicolor detection of RNA and proteins using super-resolution microscopy. *Methods* 98: 158–165
- Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163: 123–133
- Naganuma T, Nakagawa S, Tanigawa A, Sasaki YF, Goshima N, Hirose T (2012) Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J* 31: 4020–4034
- Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X, Hirose T, Prasanth KV (2012) Malat1 is not an essential component of nuclear speckles in mice. *RNA* 18: 1487–1499
- Nakagawa S, Shimada M, Yanaka K, Mito M, Arai T, Takahashi E, Fujita Y, Fujimori T, Standaert L, Marine JC, Hirose T (2014) The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* 141: 4618–4627
- Ninomiya K, Kataoka N, Hagiwara M (2011) Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J Cell Biol* 195: 27–40
- Prasanth KV, Rajendra TK, Lal AK, Lakhotia SC (2000) Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in Drosophila. *J Cell Sci* 113(Pt 19): 3485–3497

- Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL (2005) Regulating gene expression through RNA nuclear retention. *Cell* 123: 249–263
- Saborio GP, Permann B, Soto C (2001) Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* 411: 810–813
- Sanjana NE, Shalem O, Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11: 783–784
- Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T (2009) MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci USA* 106: 2525–2530
- Shav-Tal Y, Blechman J, Darzacq X, Montagna C, Dye BT, Patton JG, Singer RH, Zipori D (2005) Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. *Mol Biol Cell* 16: 2395–2413
- Shevtsov SP, Dundr M (2011) Nucleation of nuclear bodies by RNA. *Nat Cell Biol* 13: 167–173
- Singh G, Pratt G, Yeo GW, Moore MJ (2015) The clothes make the mRNA: past and present trends in mRNP fashion. *Annu Rev Biochem* 84: 325–354
- Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, Nakagawa S (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci* 120: 2498–2506
- Souquere S, Beauclair G, Harper F, Fox A, Pierron G (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Mol Biol Cell* 21: 4020–4027
- Standaert L, Adriaens C, Radaelli E, Van Keymeulen A, Blanpain C, Hirose T, Nakagawa S, Marine JC (2014) The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* 20: 1844–1849
- Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 19: 347–359
- Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 39: 925–938
- Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC, Essig J, Otto GM, O'Sullivan MG, Largaespada DA, Schwertfeger KL, Marahrens Y, Kawakami Y, Bagchi A (2014) PVT1 dependence in cancer with MYC copy-number increase. *Nature* 512: 82–86
- Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA (2005) SVA elements: a hominid-specific retroposon family. *J Mol Biol* 354: 994–1007
- Watanabe Y, Yamamoto M (1994) S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell* 78: 487–498
- 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
- West JA, Mito M, Kurosaka S, Takumi T, Tanegashima C, Chujo T, Yanaka K, Kingston RE, Hirose T, Bond C, Fox A, Nakagawa S (2016) Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J Cell Biol* 214: 817–830
- Wu H, Yin QF, Luo Z, Yao RW, Zheng CC, Zhang J, Xiang JF, Yang L, Chen LL (2016) Unusual processing generates spa lncRNAs that sequester multiple RNA binding proteins. *Mol Cell* 64: 534–548
- Yamazaki T, Hirose T (2015) The building process of the functional paraspeckle with long non-coding RNAs. *Front Biosci (Elite Ed)* 7: 1–41