<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions for use</td>
<td>Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles</td>
</tr>
<tr>
<td>Title</td>
<td>Hirose, Tetsuro; Yamazaki, Tomohiro; Nakagawa, Shinichi</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Wiley interdisciplinary reviews, RNA, 10(6), e1545</td>
</tr>
<tr>
<td>Citation</td>
<td><a href="https://doi.org/10.1002/wrna.1545">https://doi.org/10.1002/wrna.1545</a></td>
</tr>
<tr>
<td>Issue Date</td>
<td>2019-11</td>
</tr>
<tr>
<td>Doc URL</td>
<td><a href="http://hdl.handle.net/2115/79649">http://hdl.handle.net/2115/79649</a></td>
</tr>
<tr>
<td>Rights</td>
<td>This is the peer reviewed version of the following article: Tetsuro Hirose Tomohiro Yamazaki Shinichi Nakagawa. (2019) Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles. Wiley interdisciplinary reviews. RNA: 10(6): e1545., which has been published in final form at <a href="https://doi.org/10.1002/wrna.1545">https://doi.org/10.1002/wrna.1545</a>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.</td>
</tr>
<tr>
<td>Type</td>
<td>article (author version)</td>
</tr>
<tr>
<td>File Information</td>
<td>Wiley Interdiscip Rev RNA_10(6)_e1545.pdf</td>
</tr>
</tbody>
</table>

Hokkaido University Collection of Scholarly and Academic Papers: HUSCAP
Article Title:
Molecular anatomy of the architectural NEAT1 noncoding RNA: the domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles

Article Type:

Authors:

First author
Tetsuro Hirose*, ORCID iD: https://orcid.org/0000-0003-1068-5464, Affiliation: Institute for Genetic Medicine, Hokkaido University, E-mail address: hirose@igm.hokudai.ac.jp, no conflict of interest

Second author
Tomohiro Yamazaki, ORCID iD: https://orcid.org/0000-0003-0411-3590, Affiliation: Institute for Genetic Medicine, Hokkaido University, E-mail address: tomohiro_yamazaki@igm.hokudai.ac.jp, no conflict of interest

Third author
Shinichi Nakagawa, ORCID iD: https://orcid.org/0000-0002-6806-7493, Faculty of Pharmaceutical Sciences, Hokkaido University, E-mail address: nakagawas@pharm.hokudai.ac.jp, no conflict of interest
Abstract

Long noncoding RNAs (lncRNAs) are extremely diverse and have various significant physiological functions. LncRNAs generally associate with specific sets of RNA-binding proteins (RBPs) to form functional ribonucleoprotein (RNP) complexes. NEAT1 is a highly abundant IncRNA in the mammalian cell nucleus that associates with specific RBPs to form NEAT1 RNPs. Intriguingly, cellular NEAT1 RNPs are extraordinarily large and can be detected using an optical microscope. These gigantic RNPs, so-called paraspeckles, are a type of membraneless nuclear body. Paraspeckles contain approximately 50 NEAT1 RNA molecules together with characteristic RBPs possessing aggregation-prone prion-like domains. Paraspeckle formation proceeds on the nascent NEAT1 transcript in conjunction with NEAT1 biogenesis, which exhibits various features that differ from those exhibited by mRNA biogenesis, including a lack of introns, noncanonical 3’ end formation, and nuclear retention. These unique features may be required for the mechanism of paraspeckle formation. NEAT1 possesses three distinct RNA domains (A, B, and C), which function in stabilization (A), isoform switching (B), and paraspeckle assembly (C). In particular, the central C domain contains smaller subdomains that are high-affinity binding sites for the essential paraspeckle proteins (NONO and SFPQ) that subsequently polymerize along NEAT1. Subsequent recruitment of additional essential PSPs (FUS and RBM14) induces liquid-liquid phase separation to build a massive paraspeckle structure. Thus, the molecular anatomy of the NEAT1 arcRNA provides an ideal model to understand how lncRNAs form the functional RNP machinery.

Graphical/Visual Abstract and Caption

The NEAT1 architectural RNA (arcRNA) is a structural scaffold of membraneless paraspeckle nuclear bodies. NEAT1 possesses functional RNA domains that locally concentrate prion-like proteins to induce liquid-liquid phase separation and form paraspeckles as gigantic RNP complexes.

Introduction

Like proteins, long noncoding RNAs (lncRNAs) are extremely diverse and their specific functions are determined by unique features (1, 2). The majority of lncRNAs are transcribed by RNA polymerase II (RNAPII) and they form ribonucleoprotein (RNP) complexes with multiple abundant nuclear proteins (3-6). LncRNAs are broadly categorized into those that regulate gene expression in cis and those that perform functions in trans (2). Particularly, the trans-acting lncRNAs form functional RNP complexes by binding to specific regulatory factors that affect their function (7-10). Unlike mRNAs, which contain an open reading frame that is defined by the genetic code, lncRNAs contain distinct RNA domains that define their function. At least some of these RNA domains are recognized by specific RNA-binding proteins (RBPs), which usually recognize short sequence stretches, specific structures, or chemical modifications of the RNA domains.

Architectural lncRNAs (arcRNAs) act as structural scaffolds for subcellular structures or RNA granules (11-13). The mammalian cell nucleus is a highly structured compartment that plays important roles in coordinating the spatiotemporal regulation of genomic functions. Nuclear bodies, including the nucleolus, Cajal body, histone locus body, promyelocytic leukemia body, and nuclear speckle, are
approximately 0.2 to 2 µm in diameter and contain multiple nuclear regulatory factors, such as DNA-binding proteins and/or RBPs, involved in various stages of gene expression, including transcription, RNA processing, and export and storage of these factors (14, 15). Most nuclear bodies exhibit liquid droplet-like features, are phase-separated from the surrounding nucleoplasm, and can fuse to form larger droplets (16, 17). The droplets are assembled via liquid-liquid phase separation (LLPS), which is accomplished by molecules that form multiple intermolecular multivalent interactions (18-20). Nuclear bodies contain multiple RBPs with intrinsically disordered regions (IDRs) that enable RNA binding and the induction of LLPS, and are hence the driving force of non-membranous nuclear body formation (18-20). ArcRNAs sequestrate multiple IDR-containing RBPs and enable the induction of the LLPS that is essential for formation of the nuclear bodies. In this review, we focus on NEAT1 as the arcRNA involved in nuclear paraspeckle formation, and describe how it enables construction of a massive RNP complex. This review introduces recent findings regarding the molecular anatomy of the NEAT1 arcRNA, including its unique biogenesis pathway and functional domain composition.

**OVERVIEW OF PARASPECKLE STRUCTURE AND FUNCTION**

Paraspeckles are non-membranous nuclear bodies that were first identified by Visa et al. (21) as the interchromatin granule-associated zone, an electron-dense structure detected by electron microscopy. In 2002, a study identified paraspeckles as subnuclear bodies located in interchromosomal regions adjacent to nuclear speckles and notably a specific paraspeckle marker protein, paraspeckle component 1 (PSPC1) (22). In many cultured cells, paraspeckles are observed as 2–20 discrete nuclear foci with a mean diameter of 0.36 µm (23) (Figure 1A). Members of the Drosophila behavior human splicing (DBHS) family of RBPs, which includes non-POU domain-containing octamer-binding protein (NONO), splicing factor proline- and glutamine-rich (SFPQ), and PSPC1, are specifically enriched in paraspeckles (24, 25). Notably, paraspeckles are RNase-sensitive structures that depend on transcription by RNAPII, suggesting a requirement of RNAs for their maintenance (26). In 2009, four groups independently discovered that the NEAT1 (or MENε/β) lncRNA, originally named Nuclear Enriched Abundant Transcript 1, which was subsequently changed to Nuclear Paraspeckle Assembly Transcript 1, is an essential architectural component of paraspeckles by showing that depletion of NEAT1 results in the disintegration of these structures (Figure 1A) (27-30). Hence, NEAT1 appears to be essential for the formation and maintenance of paraspeckles.

Paraspeckles are massive RNP complexes composed of NEAT1 arcRNAs and multiple RBPs. Both NEAT1 and paraspeckle proteins (PSPs) are spatially arranged to form the ordered structure (Figure 1B). Transcription of NEAT1 is upregulated by various conditions, resulting in an increase in the size and number of paraspeckles, and subsequent sequestration of specific RBPs and/or RNAs away from the nucleoplasm to control gene expression (31, 32). Physiologically, NEAT1 is required for the development of specific tissues involved in female reproduction in mice and the progression of various cancers (33-36). Particularly, a subset of female Neat1 knockout mice were reported with impaired formation of the ovarian corpus luteum, a structure involved in the production of progesterone during pregnancy where Neat1 is prominently expressed. Lack of a formation of the corpus luteum during pregnancy leads to infertility and/or fewer viable pregnancies in knockout females (34). Thus, NEAT1 arcRNA is an exceptionally well characterized lncRNA and the features of
the RNA molecule and its intracellular RNP machinery, as well as its molecular and physiological functions, are well known.

**COMPONENTS REQUIRED TO BUILD A PARASPECKLE**

**NEAT1 arcRNA**

Paraspeckles can be considered as massive RNP complexes containing NEAT1 arcRNA scaffolds. The longer NEAT1\_2 isoform (22.7 kb in human) is an essential component of paraspeckles, whereas the shorter NEAT1\_1 isoform (3.7 kb in human) is dispensable (Figure 1C) (24, 28). Electron and super-resolution microscopic observations revealed that NEAT1\_2 folds into a U-shaped configuration with its 5’ and 3’ termini located in the peripheral region and its middle part located in the central core of the paraspeckle (Figure 1B) (37). A quantitative analysis revealed that 122 and 994 molecules of NEAT1\_1 and NEAT1\_2, respectively, are present in each HeLa cell; however, these numbers vary across different cell lines (38). NEAT1 expression is often induced by various internal and external stimuli. In adult mice, Neat1\_2 expression that enables paraspeckle formation is restricted to specific organs such as the corpus luteum (25, 39). Many mouse tissues express only the Neat1\_1 isoform, and paraspeckle formation is barely detectable since the expression level of the architectural Neat1\_2 isoform is too low to detect these structures (39). A single paraspeckle is estimated to contain approximately 50 NEAT1\_2 molecules (38). The detailed features of the biogenesis pathway and domain structure of NEAT1 are described in the subsequent sections.

**PSPs**

Initial studies identified a number of PSPs, including NONO, SFPQ, PSPC1, RNA-binding protein 14 (RBM14), and cleavage and polyadenylation specificity factor subunit 6 (CPSF6) (22, 40, 41). An additional 32 PSPs were identified by colocalization screening using a Venus-tagged cDNA library to recognize clones with products that colocalized with paraspeckle markers (6). Most PSPs contain canonical RNA-binding domains such as RNA recognition motifs, RGG boxes, and zinc fingers, as well as IDRs. Extensive RNAi-mediated knockdown experiments enabled the separation of PSPs into three categories according to their dependency on paraspeckle formation (6). Seven RBPs were categorized as essential proteins (Category 1) required for paraspeckle formation. Intriguingly, four of the Category 1 proteins (NONO, SFPQ, RBM14 and heterogeneous nuclear ribonucleoprotein K (HNRNPK); known as Category 1A proteins) are required for accumulation of the NEAT1\_2 isoform and paraspeckle formation (Figure 2) (6). NONO, SFPQ, and RBM14 stabilize NEAT1\_2, and HNRNPK accelerates NEAT1\_2 synthesis by inhibiting polyadenylation of the truncated short NEAT1\_1 isoform. The remaining Category 1 proteins (fused in sarcoma (FUS), DAZ-associated protein 1, and heterogeneous nuclear ribonucleoprotein H3; also known as Category 1B) are required for paraspeckle assembly but do not affect NEAT1\_2 expression (Figure 2) (6). In addition, SWI/SNF chromatin remodeling complexes, part of which localize in paraspeckles, interact with multiple Category 1 proteins and NEAT1, and are essential for paraspeckle formation (Figure 2). Intriguingly, the function of the SWI/SNF complexes in paraspeckle formation does not require canonical ATP-dependent chromatin remodeling activity, suggesting that the complexes may facilitate paraspeckle assembly by integrating paraspeckle components (42). Overall, paraspeckle formation proceeds via two independent steps: synthesis of an individual NEAT1\_2 RNP complex, followed by assembly of
approximately 50 NEAT1_2 RNP s into a massive paraspeckle structure formed in conjunction with NEAT1_2 biogenesis (Figure 2) (24, 43).

**NEAT1 arcRNA BIOGENESIS**

Synthesis of the NEAT1 arcRNA is the primary event that initiates paraspeckle assembly. NEAT1 biogenesis exhibits several features that differ from those of mRNA biogenesis. First, the NEAT1_2 isoform lacks a canonical poly(A) tail; second, both NEAT1 isoforms lack introns; and third, both NEAT1 isoforms are retained in the nucleus. In this section, we focus on the unique features of the NEAT1 biogenesis pathway and discuss the possible interplay between NEAT1 biogenesis and paraspeckle assembly.

**Transcription**

A subpopulation of paraspeckles is usually detected at the NEAT1 genomic locus (27, 44). Using a genomically integrated reporter system, Mao et al. showed that transcription of NEAT1 from its genomic loci by RNAPII can induce de novo paraspeckle formation at the transcription site (44). On the other hand, artificially tethered NEAT1 arcRNA cannot induce paraspeckle formation, suggesting that paraspeckles are co-transcriptionally formed with nascent NEAT1 at its genomic locus (45). Rapid assembly and disassembly of paraspeckles is stimulated by the induction and inhibition of global RNAPII transcription, respectively (26, 28, 30). This phenomenon may be related to the short half-life of NEAT1 (approximately 1 hour) (46) and is consistent with the liquid-like nature of paraspeckles. In addition, tethering of PSPs, rather than NEAT1, cannot induce de novo paraspeckle formation, counteracting the self-assembly mechanism of nuclear bodies such as Cajal bodies (44). Therefore, paraspeckle formation is thought to fit a two-step model, in which initial non-random NEAT1 arcRNA seeding events are followed by self-assembly with the core PSPs (14). This model is consistent with the process of paraspeckle assembly proposed based on functional analyses of the essential PSPs described above. NEAT1 transcription is modulated by various internal and external conditions, pathogens and chemicals (24, 30, 31, 33, 35, 47). Recently, genome wide screening of NEAT1 regulators revealed cross-regulation between paraspeckles and mitochondria, which is subject to regulation by NEAT1, a factor under the transcriptional control of ATF2 (48).

**3' end processing**

NEAT1_2 lacks a canonical poly(A) tail and possesses a characteristic triple helix structure at its 3’ end (3’TH) (Figure 1C) (30, 49, 50). A tRNA-like structure located downstream of 3’TH acts as a target for endonucleolytic cleavage by tRNA processing enzymes for the creation of the 3’ end of NEAT1_2 (30). The 3’TH is critical for stabilization of NEAT1_2 in vitro and in vivo (49-51); however, the significance of the 3’TH to NEAT1 function is not yet known. Cells transfected with a plasmid expressing mouse Neat1_2 containing a poly(A) tail instead of the 3’TH formed paraspeckle-like foci, indicating that the 3’TH is not essential for paraspeckle assembly (6). The 3’TH and its immediate downstream tRNA-like structure are conserved in NEAT1 orthologues in various mammalian species, including opossum in the marsupial lineage (52). A similar TH is found at the 3’ end of the MALAT1 IncRNA, supporting the possibility that MALAT1 and NEAT1 evolved from a common ancestor (53). Recently, using an RNA sequence- and structure-based covariance model, hundreds of genomic loci
were shown to contain the 3’TH in vertebrate, plant, fungi and viral genomes (54-56). Considering the function of noncanonical 3’ ends of histone mRNAs in cell cycle-dependent expression, the 3’THs likely play regulatory roles under certain conditions to maintain the proper functions of NEAT1 and MALAT1, as well as those of other IncRNAs.

**Avoiding RNA splicing and export**

Lack of introns may be an important feature of NEAT1_2. Considering the normal structure of protein-coding genes, the single exon 22.7 kb gene encoding NEAT1_2 is fairly unusual. In HeLa cells, a minor population of NEAT1_2 is spliced. We found that the spliced NEAT1 isoform is not enriched in the high density paraspeckle fraction, but is found mainly in the lower density fraction (unpublished data). This finding suggests that the spliced NEAT1 isoform is unable to localize to paraspeckles. Recently, we reported that NEAT1_2 exhibits an unusual semi-extractable feature following conventional RNA extraction that is likely linked to the arcRNA function of NEAT1_2 (38). We found that the spliced isoform was more extractable than the unspliced NEAT1_2 isoform. CRISPR-Cas9-mediated genomic deletion of the intronic region that is absent from the minor NEAT1 isoform preserved semi-extractability, suggesting that the splicing event rather than the removed sequence results in defective arcRNA function (38). Splicing events facilitate remodeling of RNP complexes to remove the pre-associated nuclear factors and recruit factors for mRNA export and/or RNA quality control (57). It is possible that co-transcriptional formation of NEAT1 RNPs antagonizes the association of splicing factors, allowing the primarily assembled RNP complexes to further assemble into functional paraspeckles. Nuclear retention of NEAT1 may also be related to the intronless structure of the gene and less efficient recruitment of the export adaptor complex. The export adaptor complex is also recruited via the 5’ cap bound to the nuclear cap-binding complex (CBC) (58). Although the presence of CBC at the 5’ end of NEAT1 remains to be proven, it is possible that a specific mechanism prevents recruitment of the export factor onto NEAT1. Recently a large R-loop was found to extend from the 5’ terminus to 1700 nt downstream of NEAT1, raising an intriguing possibility that the R-loop may play significant roles in NEAT1 biogenesis to facilitate assembly of the potent RNPs (59). The co-transcriptional assembly of NEAT1 RNPs may be controlled by the elongation rate of RNAPII, which usually affects the pattern of alternative splicing and the fate of pre-mRNAs (60). It would be interesting to measure precisely the dynamics of RNAPII at the NEAT1 genomic locus to determine its role in the unique biogenesis of NEAT1.

**NEAT1 arcRNA DOMAINS**

Analogous to polypeptides, IncRNAs likely possess functional domains that comprise specific sequences, local structures, and/or chemical modifications (12). However, the functional domains of IncRNAs have not been investigated extensively. The domains involved in the architectural function of NEAT1_2 have been studied using CRISPR-Cas9-mediated deletion analyses. As shown in Figure 3, human NEAT1_2 contains three functionally distinct domains (A, B, and C), all of which are required for paraspeckle formation (51). Intriguingly, the role of each domain can superimpose the steps of paraspeckle formation delineated from the functions of the essential PSPs.

**NEAT1 stabilization by the A domain**
The A domains located at 5' and 3' termini are required for stabilization of the NEAT1 arcRNA (51). The essential function of the 3'TH in paraspeckle formation through stabilization of NEAT1_2 has been validated experimentally. Notably, the 5' terminal 1kb region was identified as the other functional A domain required for NEAT1 stabilization and paraspeckle formation (51) (Figure 3 upper). RNA half-life measurements revealed that the 5' A domain stabilizes the NEAT1_1 and NEAT1_2 isoforms, even though the structures of their 3' termini differ, suggesting that the 5' A domain protects NEAT1 from the RNA degradation pathway that targets the 5' terminus (51). Super-resolution microscopic observations suggested that the 5' terminal regions of several NEAT1 molecules bundle and occupy distinct parts of the peripheral shells of paraspeckles (37). The bundled 5' domains may form specific RNP complexes that protect NEAT1 against degradation. It would be intriguing to explore the A domain-binding protein(s) identified from the eCLIP data summarized in our recent review (25). Previous knockdown experiments indicated that NONO and SFPQ act to stabilize NEAT1_2 without affecting the level of NEAT1_1 (28). Considering that the A domain stabilizes both NEAT1 isoforms, NONO and SFPQ may exert their stabilizing effects via a mechanism distinct from that involving the A domain.

**NEAT1 isoform switching by the B domain**

The B domains of NEAT1 regulate isoform switching by inhibiting polyadenylation of NEAT1_1 to promote synthesis of the architectural NEAT1_2 isoform. Two regions upstream (2.1–2.8 kb) and downstream (4.0–5.1 kb) of the polyadenylation site (PAS) of NEAT1_1 promote the expression of NEAT1_2 and suppress the expression of NEAT1_1 (Figure 3 upper). The B domains appear to be functional only when the PAS is intact, suggesting that they suppress PAS-dependent polyadenylation to promote NEAT1_2 synthesis. The regulation of NEAT1_1 polyadenylation reportedly occurs via a distinct mechanism in which CPSF6/Nudt21 binds to the CUGA sequence cluster located close to the PAS to promote polyadenylation of NEAT1_1. HNRNPK interferes with polyadenylation by binding to a pyrimidine stretch located between the CUGA cluster and the PAS (6). The relationship between the role of B domains and the HNRNPK-dependent mechanism is still unclear. We recently observed that the functionality of HNRNPK in NEAT1_2 synthesis is cell type-dependent (unpublished data), suggesting that isoform switching of NEAT1 is controlled by at least two distinct mechanisms, each of which can depend on the specific cell type.

**Control of paraspeckle assembly by the C domain**

The C domain is a relatively large middle region (8–16.6 kb) of NEAT1_2 that is required and sufficient for paraspeckle assembly (51) (Figure 3 upper). Deletion of the C domain impairs paraspeckle formation, and a NEAT1_2 deletion mutant (mini-NEAT1) possessing the C domain, 5' A domain, and 3'TH (Figure 3 lower) is able to form a structurally ordered paraspeckle (51). Further deletion analyses revealed that the C domain contains at least three functional subdomains (51) (C1: 9.8–12 kb, C2: 12–13 kb, and C3: 15.4–16.6 kb). An analysis of various deletion mutants derived from mini-NEAT1 showed that deletion of C1 and C2 abolishes paraspeckle assembly, whereas deletion of either of these subdomains alone does not, suggesting functional redundancy of C1 and C2. The C3 subdomain may be functionally distinct because deletion of this region alone also abolishes paraspeckle assembly (51). Similar redundancy in function is also exerted by regions outside of the C domain. Thus, there appear to be complex multiple functional redundancies between the C domain and the outside region, as well as within the C domain itself. The C domain contains eight distinct
sequences derived from long interspersed nuclear element 1 (LINE1). The C1 and C3 subdomains are largely occupied by the LINE1 sequences, raising the intriguing possibility that the RNA sequences derived from repetitive genetic elements play critical roles in paraspeckle assembly. Each of the functional C subdomains likely interacts with the essential PSPs to dictate function (51). Artificial tethering of NONO, SFPQ, and FUS, but not RBM14, to a functionally defective NEAT1 mutant lacking the C1 and C2 subdomains rescues paraspeckle assembly (51). In addition, in vitro binding experiments showed that NONO and SFPQ preferentially bind to RNAs derived from C1 and C2 subdomains, strongly suggesting that the role of these subdomains is to recruit the essential NONO and SFPQ onto NEAT1_2 to initiate paraspeckle assembly (51) (see below).

**NEAT1_1**

The NEAT1_1 isoform is dispensable for paraspeckle formation, and its role remains poorly understood. The gene structure in which NEAT1_1 and NEAT1_2 share a common 5’ terminus has made the functional relationship between the two isoforms difficult to dissect. Initially, it was reported that the amount of NEAT1_1 in HeLa cells is approximately 2–3-fold higher than the amount of NEAT1_2 (28). However, it was recently found that NEAT1_2 is less extractable with the canonical RNA extraction reagent than NEAT1_1, and the actual amount of NEAT1_1 in HeLa cells is 10-fold lower than that of NEAT1_2, indicating that it is not a major component of paraspeckles (38). Genome editing-mediated depletion of each NEAT1 isoform revealed similar evidence and additionally found that NEAT1_1 localizes in numerous non-paraspeckle foci termed as “microspeckles,” which may have paraspeckle-independent functions (61). The ratio of NEAT1_1 and NEAT1_2 is highly variable in different cell lines and under different external conditions, raising the intriguing possibility that paraspeckles sequestrate NEAT1_1 RNPs in the peripheral shell (38, 39, 61).

**Pseudo-miRNA**

NEAT1_2 harbors a “pseudo-miRNA” (miR-612) at its 3’ terminal region that is poorly processed into mature miRNA (62). The pseudo-miRNA functions to attract the microprocessor that processes pri-miRNA to pre-miRNA in the nucleus. The major PSPs, NONO/SFPQ, form extensive interactions with expressed pri-miRNAs. Furthermore, the presence of a certain pri-miRNA, such as pri-miR-1, was detected at paraspeckles in differentiated C2C12 cells. Based on these data, it seems that NEAT1 with the pseudo-miRNA and the multiple interaction regions of NONO/SFPQ forms paraspeckles that facilitate pri-miRNA processing. We recently observed that CRISPR/Cas9 mediated deletion of the pseudo-miRNA containing region of NEAT1_2 did not affect paraspeckle structure (unpublished data). It is noteworthy that rodent Neat1 lacks the pseudo-miRNA; therefore, the contribution of this element to pri-miRNA localization and processing requires further investigation.

**HOW TO ASSEMBLE A MASSIVE PARASPECKLE STRUCTURE**

This section focuses on the molecular interactions between the NEAT1 functional domains and the interacting proteins, and the mechanism that enables construction of paraspeckles as gigantic RNP complexes with arcRNA and prion-like domain proteins.

**NONO/SFPQ polymerization**
Functional interaction of NONO and SFPQ with the C1 and C2 subdomains of NEAT1_2 initiates paraspeckle assembly (Figure 5) (51). NONO and SFPQ form a major heterodimer in cells, and the characteristic NOPS domain in NONO is responsible for its dimerization with DBHS family proteins (63). When artificially tethered to a functionally defective NEAT1_2 mutant, a NONO mutant lacking the NOPS domain lost its ability to rescue paraspeckle assembly, suggesting that heterodimerization of NONO and SFPQ is required to initiate assembly of these structures. The primarily associated NONO/SFPQ dimers can subsequently induce polymerization via the coiled-coil (CC) domain, which may eventually cover the whole NEAT1_2 arcRNA. PAR-CLIP data indicated that PSP-binding sites broadly cover the whole NEAT1_2 region, and the three subdomains contain multiple prominent binding peaks for NONO and SFPQ (51). In addition, a previous study using transmission electron microscopy showed that SFPQ can coat DNA to form higher-order complexes by polymerization through the CC domain (64). Based on this idea, we propose that the role of the C1 and C2 subdomains is to act as primary binding site(s) for the NONO/SFPQ dimer(s), to enable subsequent polymerization and form the foundation of the NEAT1_2 RNP. Similar cooperative RNA binding was proposed for HNRNPA1, which primarily binds exonic splicing silencers as well as the adjacent sequences (65), suggesting the occurrence with multiple RBPs. To prove the hypothesized two-directional polymerization from the C1 and C2 subdomains, mapping of NONO\(\Delta_{\text{CC}}\) mutant binding sites on NEAT1_2 in the absence of endogenous NONO should be performed. Such experiments would also enable discrimination of the high-affinity binding sites of NONO and SFPQ from the secondary binding sites generated by polymerization.

**LLPS**

LLPS is accomplished by proteins possessing low complexity domains (LCDs) (Figure 4A). LCDs lack a defined three-dimensional structure and provide the basis for multivalent weakly adhesive intermolecular interactions, such as electrostatic, pi stacking, and hydrophobic interactions (13, 18, 20). Above a threshold concentration, proteins that form multivalent interactions can self-assemble and undergo LLPS to enable formation of massive membraneless bodies such as paraspeckles. Therefore, there is a strong likelihood that the NEAT1 arcRNA provides multiple binding sites for LCD-containing proteins to promote their local concentration, and eventually induces LLPS near its own transcription site. Most PSPs contain LCDs that are categorized as prion-like domains (PLDs), which efficiently induce LLPS (24, 51, 66). The PLDs of two essential PSPs, FUS and RBM14, cause in vitro hydrogel formation and are required for in vivo paraspeckle formation (Figure 4B) (66). Like LLPS-formed Cajal bodies, paraspeckles are sensitive to aliphatic alcohol 1,6-hexanediol treatment (51). PSPs rapidly move into and out of the paraspeckle, and thus the paraspeckle is highly dynamic, which is one of the characteristics of phase-separated bodies. Furthermore, the formation of large bead aggregates that may mimic LLPS has been observed in vitro using magnetic beads conjugated with RNA fragments derived from the C2 subdomain of NEAT1_2, suggesting that this subdomain enables LLPS induction (51). Bead aggregation occurs in association with selective binding of NONO and SFPQ, and subsequent binding of the PLD proteins, FUS and RBM14, which depends on the presence of NONO and SFPQ. Association of the C2 subdomain with NONO and SFPQ may subsequently recruit FUS and RBM14 to trigger LLPS and eventually form a massive paraspeckle structure (Figure 5). This proposal is supported by recent reports showing that the NEAT1 RNA fragment promotes in vitro nucleation of FUS droplets buffered by nonspecific RNAs (67). Intriguingly, the unusual semi-extractable feature of NEAT1_2 described above is diminished by
either depletion of FUS or deletion of the NEAT1_2 C domain, suggesting that an interaction between FUS and NEAT1_2 RNA contributes to the acquisition of the semi-extractability of NEAT1_2 (38). The semi-extractable feature of NEAT1_2 may relate to its architectural function through the induction of LLPS.

RNA-RNA interactions

Recent evidence suggested that intermolecular RNA-RNA interactions play a role in forming RNP granules (68-70). For example, repeat expansion of CTG or CAG DNA creates repetitive RNAs that form multivalent base-pairing, causing purified RNA to undergo a sol-gel transition in vitro. The evidence that repetitive RNAs form distinct nuclear body-like foci raises the intriguing possibility that paraspeckle formation may be instigated by intermolecular RNA-RNA interactions. Based on the volume of a paraspeckle and the number of NEAT1 RNA molecules it contains, the concentration of NEAT1 is estimated to be high enough to induce RNA-based self-assembly under physiological conditions (69). Treatment with 1,6-hexanediol causes dissociation of NONO from paraspeckles, while weaker NEAT1 foci remain detectable, suggesting the involvement of protein-independent RNA-RNA interactions of NEAT1 in paraspeckle formation (51). Indeed, recent RNA structural mappings have identified numerous intramolecular RNA-RNA interactions within NEAT1_2 (71), which may occur intermolecularly at the site of transcription or within paraspeckles where NEAT1 is highly concentrated.

How to fix the NEAT1 configuration

A CRISPR-Cas9-mediated deletion analysis revealed the competence of mini-NEAT1 in paraspeckle assembly; however, mini-NEAT1 cannot form a U-shape with proper positioning of the 5’ and 3’ termini in the shell (51). Mini-NEAT1 largely lacks the 5’ and 3’ regions, suggesting that they contain an unidentified functional domain required for proper NEAT1_2 configuration within a paraspeckle. RNA structural mapping and bioinformatics analyses predicted possible intramolecular interactions between the 5’ and 3’ regions of NEAT1_2 and suggested the involvement of these interactions in the U-shape configuration (71). Pinpoint mutational analyses that disrupt and/or compensate the possible intramolecular interactions should be able to verify this possibility. It is thought that a number of NEAT1_2 bundles and their 5’ and 3’ parts occupy distinct regions in the peripheral shell of paraspeckles, suggesting the presence of shell factors that facilitate bundling of NEAT1_2 to form the fixed configuration.

GENERALITY OF arcRNA FUNCTION

The generality of arcRNA functions could be addressed by exploration and identification of novel arcRNAs and subsequent comparison of their mechanisms of action. Based on the distinct characteristics of NEAT1_2, an arcRNA can be categorized according to the following criteria: 1) it is localized and enriched in a specific nuclear body, and 2) it constructs and stabilizes the body structure. The latter could be confirmed by RNA depletion to disrupt the body or artificial RNA tethering to construct the body. At present, eight IncRNAs can be classified as arcRNAs involved in
building nuclear bodies, including NEAT1_2 lncRNA in the paraspeckle, intergenic spacer lncRNA in the amyloid body (72), high-copy satellite II lncRNA in Cancer-Associated Satellite Transcript bodies (73), high-copy satellite III lncRNA in the nuclear stress body (74), pyrimidine-rich noncoding transcript in the perinucleolar compartment (75), tumor-associated NBL2 transcript (TNBL) in TNBL aggregates (76), Drosophila heat shock RNA omega in the omega speckle (77), and fission yeast meiRNA in the Mei2 dot (78). Many of these arcRNAs are relatively abundant transcripts that contain multiple repetitive sequences or are transcribed from repetitive genomic regions such as satellite repeats. These features of arcRNAs may be critical for their common architectural function by local sequestration of specific PLD proteins to concentrate and trigger LLPS. This proposal is consistent with the evidence that abnormal expansion of repetitive sequence in some RNAs, which often results in neurodegenerative diseases such as myotonic dystrophy type 1 and spinocerebellar ataxia type 31, can confer an architectural role to RNA (79). Some arcRNAs are primate-specific, suggesting that they were independently selected for massive complex formation during evolution. In addition to the established arcRNAs described above, unidentified arcRNAs are presumed to build several other nuclear bodies, including the Sam68 nuclear body and DBC1 body, which are RNase-sensitive and built with specific RBPs (80). As described above, we recently developed a method for transcriptome-wide screening of arcRNA candidates according to their semi-extractable features. Using transcriptome-wide analyses of HeLa cells, 45 semi-extractable RNAs were identified, of which the top 10 most abundant RNAs all showed predominant localization in nuclear foci (38). These findings suggest the presence of unidentified arcRNAs in the human transcriptome.

Figures and Tables

**Figure 1.** The NEAT1 arcRNA in paraspeckle nuclear bodies. A. NEAT1 is essential for paraspeckle integrity. The paraspeckles visualized by immunostaining of NONO (green signals in WT) are abolished by NEAT1 depletion (NEAT1 KO). Scale bar 5 µm. B. The upper image shows the fine structure of mouse paraspeckles determined by super-resolution microscopy. Paraspeckle substructures were detected by RNA-FISH with NEAT1 antisense RNA probes (NEAT1_5’ and NEAT1_3’: green; NEAT1_m: magenta). The lower panel shows a schematic model of the core-shell arrangement of paraspeckle components. Scale bar 500 nm. C. NEAT1 isoforms synthesized by alternative 3’ end processing. NEAT1_2 possesses non-polyadenylated 3’TH. The positions of RNA-FISH probes used in B (5’, m, and 3’) are shown.

**Figure 2.** Paraspeckle formation proceeds via two distinct steps. The essential factors for paraspeckle formation are shown.

**Figure 3.** The modular domain structure of the human NEAT1_2 arcRNA. The three distinct domains (A, B, and C) and three subdomains within the C domain are shown, and their respective functions are indicated. The position of each domain relative to the 5’ end (kb) is also shown. The structure of mini-NEAT1 is shown below. The dashed lines represent the deleted regions.

**Figure 4.** LLPS is a driving force of paraspeckle formation. A. The basic concept of LLPS. LLPS is driven by multivalent interactions between proteins containing intrinsically disordered PLDs. B. The PLDs of FUS and RBM14 are required for in vitro phase separation and in vivo paraspeckle formation. The
domain structures of FUS and RBM are shown at the top. The consensus sequence in each PLD is shown in the middle.

**Figure 5.** An updated model of paraspeckle formation. The NONO/SFPQ dimer primarily associates with the C1 and C2 subdomains of NEAT1_2 and subsequently polymerizes via the CC extended domains to form the NEAT1_2 RNP platform. PLD proteins (FUS and RBM14) associate with this platform to form the NEAT1_2 RNP complex. NEAT1_2 RNAs bundle to form larger complexes in which LLPS is induced to sequester more complexes to eventually form an intact paraspeckle.

**Table 1.** Factors essential for paraspeckle formation

<table>
<thead>
<tr>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>This review focused on the molecular features of the NEAT1 arcRNA identified via molecular anatomical analyses of its biogenesis and functional domains. The unique NEAT1 biogenesis pathway is thought to be required for paraspeckle formation. NEAT1_2 possesses three distinct domains that function in stabilization, isoform switching, and paraspeckle assembly through interaction with PSPs to induce LLPS. Thus, we argue that the major role of arcRNAs is to induce LLPS to locally enrich specific regulatory proteins. Further exploration of arcRNA elements located in the functional domains would provide additional insights into the mechanisms of arcRNA function. Analyses of numerous arcRNAs prevalently synthesized from eukaryotic genomes would advance current understanding of the fundamental rules underlying the RNA-instructed intracellular architecture.</td>
</tr>
</tbody>
</table>

**Funding Information**

This work was supported by MEXT KAKENHI Grants (to T.H [26113002], T.Y.[26891001], and to S.N. [26113005]), by JSPS KAKENHI Grants (to T.H. [17H03630, 17K19335 and 16H06279], T.Y. [17K15058] and S.N [17H03604, 16H06279 and 16H06276] ), by the Tokyo Biochemical Research Foundation to T.H., by the Mochida Memorial Foundation for Medical and Pharmaceutical Research to T.Y., by a Toray Science and Technology Grant to S.N., and by a grant from the Joint Research Program of IGM, Hokkaido University to S.N. and T.H..

**Acknowledgments**

The authors thank the members of the Hirose laboratory and the Nakagawa laboratory for valuable discussions.

**References**

2. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding
14. Fox AH, Nakagawa S, Hirose T. Molecular dissection of nuclear paraspeckles: towards...
understanding the emerging world of the RNP milieu. Open Biol. 2018;8(10).
61. Li R, Harvey AR, Hodgetts SI, Fox AH. Functional dissection of NEAT1 using genome editing reveals substantial localization of the NEAT1_1 isoform outside paraspeckles. RNA. 2017;23(6):872-81.


71. Lin Y, Schmidt BF, Bruchez MP, McManus CJ. Structural analyses of NEAT1 IncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. Nucleic acids research. 2018;46(7):3742-52.


NEAT1_2 arcRNA biogenesis
no splicing, no polyadenylation, no export

Paraspeckle

NEAT1_2 RNP

NEAT1 bundling

LLPS

Prion-like RBPs

NONO/SFPQ
**Figure A**

Comparison between WT and NEAT1 KO cells. Scale bars indicate measurement units.

**Figure B**

Immunofluorescence of NEAT1 expression in cells. Scale bars indicate measurement units.

**Figure C**

Genetic map of NEAT1 and its surrounding genes. Key:
- **NEAT1_1**: 3.7 kb
- **NEAT1_2**: 22.7 kb
- **NEAT1** (Chr. 11)
- **TH**

**Legend**
- **Core**: Neat1_5', Neat1_3', Tardbp
- **Shell**: Neat1_5', Neat1_3', Sfpq, Nonm, Pspc1, Fus, Rbm14, Brg1
- **Patch**: Neat1_5', Neat1_3'
NEAT1 stability isoform switch Paraspeckle assembly NEAT1_2 stability

mini-NEAT1
**A**

In vitro phase separation

**B**

**FUS PLD**

**RBM14 PLD**

In vivo paraspeckle formation
Formation of NEAT1_2 RNP complex

NEAT1 bundling

LLPS

Paraspeckle formation
<table>
<thead>
<tr>
<th>Essential factors</th>
<th>PSP category</th>
<th>Function in paraspeckle formation</th>
<th>Molecular function</th>
<th>Presence of PLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEAT1_2</td>
<td>n/a</td>
<td>Structural scaffold</td>
<td>arcRNA</td>
<td>n/a</td>
</tr>
<tr>
<td>RNA polymerase II</td>
<td>n/a</td>
<td>NEAT1 transcription</td>
<td>RNA polymerase</td>
<td>(+)</td>
</tr>
<tr>
<td>NONO</td>
<td>1A</td>
<td>NEAT1_2 stability</td>
<td>RNA-binding protein</td>
<td>+</td>
</tr>
<tr>
<td>SFPQ</td>
<td>1A</td>
<td>NEAT1_2 stability</td>
<td>RNA-binding protein</td>
<td>+</td>
</tr>
<tr>
<td>RBM14</td>
<td>1A</td>
<td>NEAT1_2 stability</td>
<td>RNA-binding protein</td>
<td>+</td>
</tr>
<tr>
<td>HNRNPK</td>
<td>1A</td>
<td>Blocking polyadenylation of NEAT1_1</td>
<td>RNA-binding protein</td>
<td>-</td>
</tr>
<tr>
<td>FUS</td>
<td>1B</td>
<td>Paraspeckle assembly</td>
<td>RNA-binding protein</td>
<td>+</td>
</tr>
<tr>
<td>DZAP1</td>
<td>1B</td>
<td>Paraspeckle assembly</td>
<td>RNA-binding protein</td>
<td>+</td>
</tr>
<tr>
<td>HNRNPH3</td>
<td>1B</td>
<td>Paraspeckle assembly</td>
<td>RNA-binding protein</td>
<td>+</td>
</tr>
<tr>
<td>SWI/SNF complexes</td>
<td>1B</td>
<td>Paraspeckle assembly</td>
<td>Chromatin remodeling</td>
<td>(+)</td>
</tr>
</tbody>
</table>