Functional dissection of NEAT1 using genome editing reveals substantial localization of the NEAT1_1 isoform outside paraspeckles

RUOHAN LI,[1,2] ALAN R. HARVEY,[2,3] STUART I. HODGETTS,[2,3] and ARCHA H. FOX[1,2]

[1]Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical Research, The University of Western Australia, Nedlands, Western Australia 6009, Australia
[2]School of Human Sciences, The University of Western Australia, Crawley, Western Australia 6009, Australia
[3]Western Australian Neuroscience Research Institute, Nedlands, Western Australia 6009, Australia

ABSTRACT

Large numbers of long noncoding RNAs have been discovered in recent years, but only a few have been characterized. NEAT1 (nuclear paraspeckle assembly transcript 1) is a mammalian long noncoding RNA that is important for the reproductive physiology of mice, cancer development, and the formation of subnuclear bodies termed paraspeckles. The two major isoforms of NEAT1 (3.7 kb NEAT1_1 and 23 kb NEAT1_2 in human) are generated from a common promoter and are produced through the use of alternative transcription termination sites. This gene structure has made the functional relationship between the two isoforms difficult to dissect. Here we used CRISPR-Cas9 genome editing to create several different cell lines: total NEAT1 knockout cells, cells that only express the short form NEAT1_1, and cells with twofold more NEAT1_2. Using these reagents, we obtained evidence that NEAT1_1 is not a major component of paraspeckles. In addition, our data suggest NEAT1_1 localizes in numerous nonparaspeckle foci we termed "microspeckles," which may carry paraspeckle-independent functions. This study highlights the complexity of lncRNA and showcases how genome editing tools are useful in dissecting the structural and functional roles of overlapping transcripts.

Keywords: long noncoding RNAs; NEAT1; paraspeckles; CRISPR genome editing; DBHS proteins

INTRODUCTION

Large-scale transcriptomic and epigenetic analyses have systematically demonstrated regulatory potential for the majority of our genome; in particular, pervasive transcription occurring largely in the form of long noncoding RNAs (lncRNAs) with diverse and enigmatic functions (The ENCODE Project Consortium 2012; The FANTOM Consortium and the RIKEN PMI and CLST [DGT] 2014). While the majority of lncRNAs are expressed at low levels with tissue-specific and developmental stage-specific expression (for review, see Mercer et al. 2009; Mattick 2011), the mammalian-specific lncRNA NEAT1 (nuclear paraspeckle assembly transcript 1) is ubiquitous and abundant, suggesting a more generic biological role. Importantly, NEAT1 is one of the few lncRNAs known to have a robust phenotype when knocked out in mice, in which the secretory function and the development of critical tissues relating to female reproduction are compromised (Nakagawa et al. 2014; Standaert et al. 2014).

NEAT1 transcription can be significantly influenced by p53 and is often up-regulated by stresses, such as hypoxia, viral infection/products, and genotoxic agents (Saha et al. 2006; Choudhry et al. 2014; Imamura et al. 2014; Adriaens et al. 2016). Unsurprisingly, there is also a strong link between high NEAT1 expression and more aggressive forms of cancer (Chakravarty et al. 2014; Chai et al. 2016; Chen et al. 2016; Fu et al. 2016; Ma et al. 2016; Sun et al. 2016; Wang et al. 2016). In neurological systems, up-regulated expression of NEAT1 has also been observed in the early phase of amyotrophic lateral sclerosis and Huntington's disease, and more interestingly, its expression could also be acutely down-regulated in response to neuronal activity (Nishimoto et al. 2013; Sunwoo et al. 2016; Barry et al. 2017). At the molecular level, NEAT1 is important for the formation of paraspeckles, a type of mammalian nuclear RNA–protein body found in close proximity to nuclear speckles (Fox et al. 2002; Hutchinson et al. 2007). Paraspeckles modulate gene expression by sequestering mRNAs and transcription factors (TFs)

(Prasanth et al. 2005; Chen and Carmichael 2009; Naganuma et al. 2012; Choudhry et al. 2014; Hirose et al. 2014; Imamura et al. 2014; West et al. 2016), and their formation occurs when specific architectural proteins, including those of the DBHS (*Drosophila* behavior/human splicing) family, bind to nascent NEAT1 transcripts (Mao et al. 2011). Many other protein and RNA factors localize to paraspeckles, some with potential roles in paraspeckle function and others not required for function but potentially regulated by paraspeckle sequestration (Prasanth et al. 2005; Chen and Carmichael 2009; Naganuma et al. 2012; Choudhry et al. 2014). Currently, a working definition of a paraspeckle is a nuclear body in which NEAT1 RNA and DBHS proteins are colocalized (Naganuma et al. 2012).

The *NEAT1* gene is transcribed as two major isoforms that overlap completely at their 5′ ends: the shorter, polyadenylated NEAT1_1 (also called MENε, 3.7 kb in human) and the longer NEAT1_2 isoform (also called MENβ, 23 kb in human) (Sasaki et al. 2009; Sunwoo et al. 2009). This overlapping nature has caused great difficulties in differentiating the molecular binding targets of each. For example, recent CHART (capture hybridization analysis of RNA targets) experiments using reagents common to both NEAT1 isoforms have shown NEAT1 enrichment at chromatin containing active histone marks, and localization to the transcriptional start sites (TSS) and termination sites (TTS) of active genes (Chakravarty et al. 2014; West et al. 2014). However, due to the CHART probes binding both isoforms, it is not clear if NEAT1_1, NEAT1_2, or both isoforms are involved in this chromatin binding. On the other hand, it is established that only NEAT1_2 forms the essential RNA backbone for paraspeckle formation (Clemson et al. 2009; Sasaki et al. 2009; Sunwoo et al. 2009; Mao et al. 2011; Naganuma et al. 2012). The shorter NEAT1_1 isoform, albeit highly abundant, is not sufficient for paraspeckle formation and its involvement with the function of paraspeckles is still unclear (Mao et al. 2011). Moreover, several functional studies reported that overexpression of NEAT1_1 may promote tumor growth and metastasis (Chakravarty et al. 2014; Chen et al. 2016; Sun et al. 2016), and other studies in different cancer subtypes have indicated that high NEAT1_2 expression is linked to worse cancer prognosis via the activation of ATR (ataxia telangiectasia and Rad3-related) pathways (Adriaens et al. 2016). These reports seem to imply that the two NEAT1 isoforms may be functionally distinct, or even opposite, entities, thus rendering the relationship between the two NEAT1 isoforms and paraspeckles enigmatic.

To dissect the functional relationship between the two isoforms, cell models that selectively express each isoform individually would be required. However, due to the fact that the two NEAT1 isoforms overlap completely at the 5′ end, transient specific knockdown of NEAT1_1 is impossible to achieve without perturbing NEAT1_2 levels. In addition, transient overexpression of the 23 kb NEAT1_2 is also very technically challenging and does not recapitulate the local

chromatin environment upon which paraspeckle formation depends (Mao et al. 2011; Naganuma et al. 2012). Interestingly, the biogenesis of NEAT1_1 relies on efficient early polyadenylation of NEAT1_1 transcripts at several classical poly(A) sites, with the production of NEAT1_2 resulting from the occasional inhibition of this process by hnRNP-K (Naganuma et al. 2012).

In this study, we created several genome edited stable U2OS osteosarcoma cell lines to dissect the functional relationship between NEAT1 isoforms and paraspeckles. Our microscopic data suggest that NEAT1_1 may not be a significant component of paraspeckles. As a result, our data challenge a long-standing assumption in the field, raising the possibility that NEAT1_1 may indeed carry paraspeckle-independent functions.

RESULTS AND DISCUSSION

Generation of total NEAT1 depleted cells and NEAT1_2 specifically depleted cells using a gene structure disruption strategy

NEAT1 is one of the most abundant lncRNA transcripts (Gibb et al. 2011). Transient reduction in NEAT1 levels via siRNAs or antisense oligos has not allowed the investigation of long-term effects of NEAT1 depletion at a cellular level. Thus we reasoned a stable NEAT1 knockout human cell line would provide a useful platform to conduct such studies. Using CRISPR-Cas9 genome editing, we applied a strategy to disrupt the transcription start site (TSS) of the human *NEAT1* gene using a fluorescent protein expression cassette, a similar strategy to that taken by Gutschner and colleagues to create MALAT1 (metastasis associated lung adenocarcinoma transcript 1) lncRNA stable knockout in human cell lines (Fig. 1A; Gutschner et al. 2011). As successfully edited cells would express YFP, we enriched this population by three rounds of FACS (fluorescent-activated cell sorting), then seeded single cells, and subsequently screened the genomic DNA to identify lines in which both *NEAT1* alleles were mutated. Thus we generated the first human osteosarcoma U2OS cell lines in which expression of both NEAT1 isoforms was stably reduced by at least 98% in a homozygous knockout setting, far exceeding what is normally achieved with transient methods (Fig. 1B; Sasaki et al. 2009; Hirose et al. 2014). As expected, the stable total NEAT1$^{-/-}$ U2OS cells showed no sign of paraspeckles, as defined by colocalization of NEAT1 with the essential DBHS paraspeckle protein NONO (non-POU domain containing, octamer-binding) (Fig. 1C; Fox et al. 2005).

We next sought to investigate the relationship between the two NEAT1 isoforms as NEAT1_1 is much more abundant than NEAT1_2, yet is not structurally important for paraspeckle formation. Until this point the only way to dissect these two isoforms was to transiently ablate NEAT1_2 with siRNA or ASOs, leaving NEAT1_1 intact (Sunwoo et al.
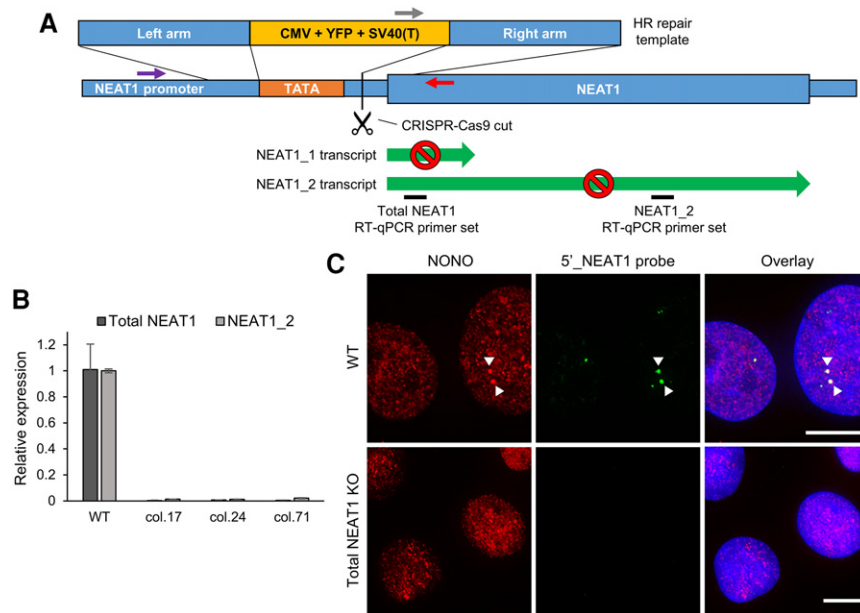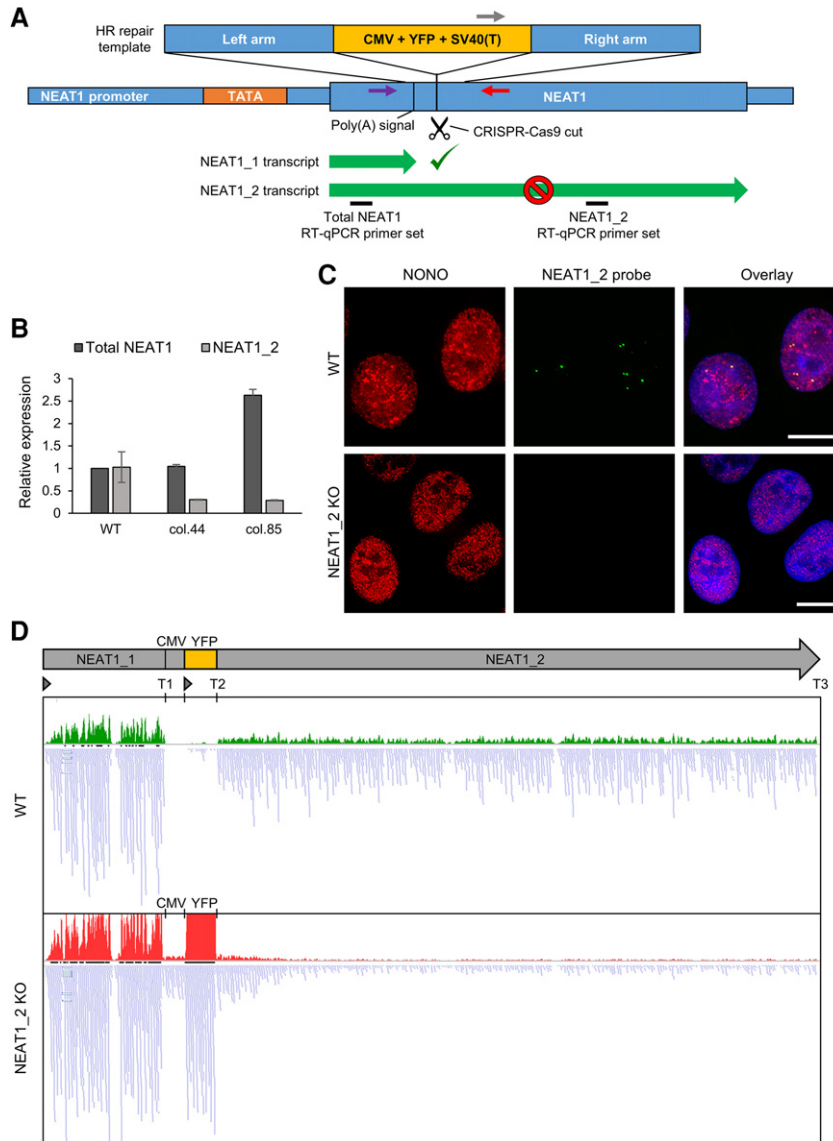
FIGURE 1. Generation of U2OS cell lines lacking expression of NEAT1. (*A*) A schematic representation of the total NEAT1 knockout strategy through CRISPR-Cas9-mediated YFP knock-in. A Cas9-mediated DSB, guided by either NEAT1 promoter-targeting sgRNA_1 or sgRNA_2, was introduced to the region between the TATA box and the transcription start site (TSS) of NEAT1. Each cut was to be repaired by homologous recombination, which would insert one YFP expression cassette for downstream selection and also remove the TATA box. Homozygous clones were screened using PCR primer sets indicated by gray, purple, and red arrows. (*B*) NEAT1 knockout efficiency detected by the two RT-qPCR primer sets (black bars), of three representative homozygous clonal lines (col.17 and 71 made with sgRNA_1, and col.24 made with sgRNA_2). (*C*) Immunofluorescence of NONO and RNA-FISH using probes directed to the 5′ end of NEAT1 showed complete loss of NEAT1 RNA and any paraspeckle-like foci (arrows) of NONO in the total NEAT1$^{-/-}$ cells. DAPI counterstaining is shown in blue. Scale bar = 10 µm. RT-qPCR results were one representative measurement plotted as mean ± SD of two technical replicates normalized to U6.

2009; Imamura et al. 2014; Adriaens et al. 2016). To make a permanent NEAT1_2 knockout line, here we applied a variation on the YFP knock-in strategy used for promoter disruption: We targeted the genomic region closely downstream from the polyadenylation signal of the NEAT1_1 isoform to selectively disrupt NEAT1_2, leaving NEAT1_1 intact (Fig. 2A). We reasoned that NEAT1_1 would be made unaffected, but that the transcriptional terminator (SV40-T) in the YFP cassette following NEAT1_1 would lead to a significant reduction of the continued transcription of NEAT1_2. In addition, the presence of the embedded cassette would alter the NEAT1_2 primary sequence/structure in the event of any transcription reading through the SV40-T, which should render the recombinant transcript nonfunctional in forming paraspeckles. Indeed, homozygous NEAT1_2$^{-/-}$ cell lines made with this strategy were devoid of paraspeckles, marked by both the loss of typical paraspeckle-like localization of NONO, and the loss of NEAT1_2 RNA-FISH (RNA fluorescent in situ hybridization) signal (Fig. 2C). This suggested that the primary NEAT1_2 transcript was successfully knocked out, and the remnant recombinant transcripts with the YFP coding region and SV40-T embedded within them did not retain para-

speckle scaffolding ability. At the RNA level, using the RT-qPCR primers that selectively amplify NEAT1_2, we found ∼70% reduction of NEAT1_2 levels in homozygous NEAT1_2$^{-/-}$ cell lines (Fig. 2B). It is likely that the remaining 30% RNA detected is potentially due to leaky transcription of the nonfunctional YFP-cassette-NEAT1_2 fusion transcript. Interestingly, NEAT1_1 levels were either unaltered or slightly increased in some NEAT1_2$^{-/-}$ lines (Fig. 2B). This indicated a possible feedback mechanism in which cells might increase the transcription of the *NEAT1* gene to restore the lost NEAT1_2. Lastly, we performed RNA-seq analysis of both WT and NEAT1_2$^{-/-}$ cells to confirm that the biogenesis of NEAT1_1 was still normal in the engineered cells (Fig. 2D). The abundant reads mapping to the NEAT1_1 sequence indicate that the vast majority of the NEAT1_1 transcripts terminate properly at the natural NEAT1_1 termination site (T1, Fig. 2D). In addition, the abundant reads mapping to the YFP sequence indicate that the vast majority of the YFP transcripts terminate properly at the YFP termination site (T2, Fig. 2D). Interestingly, a small number of reads mapped to sequences downstream from both termination sites, i.e., within the CMV promoter, and along NEAT1_2, indicating some incomplete transcriptional termination at both T1 and T2. This small pool of incompletely terminated signal demonstrates the likely existence of one or more artificial NEAT1-YFP fusion transcripts that are nevertheless not capable of forming paraspeckles. Taken together, these results show the utility of such approaches for permanently reducing levels of abundant transcripts and confirm earlier studies indicating that NEAT1_1 per se is not sufficient to form paraspeckles. In the future, these NEAT1_2$^{-/-}$ lines could be extremely useful reagents for studying any potential paraspeckle independent function of NEAT1_1 without the interference of NEAT1_2.

NEAT1_1 lacks interaction with DBHS proteins and localizes to nonparaspeckle nuclear sites

NEAT1_1 is known to be the more abundant NEAT1 isoform, but its localization has not been clearly defined. As mentioned above, any attempt to visualize NEAT1_1 inevitably also targets the 5′ end of NEAT1_2. However, in cells in which NEAT1_2 can no longer be generated, such an issue
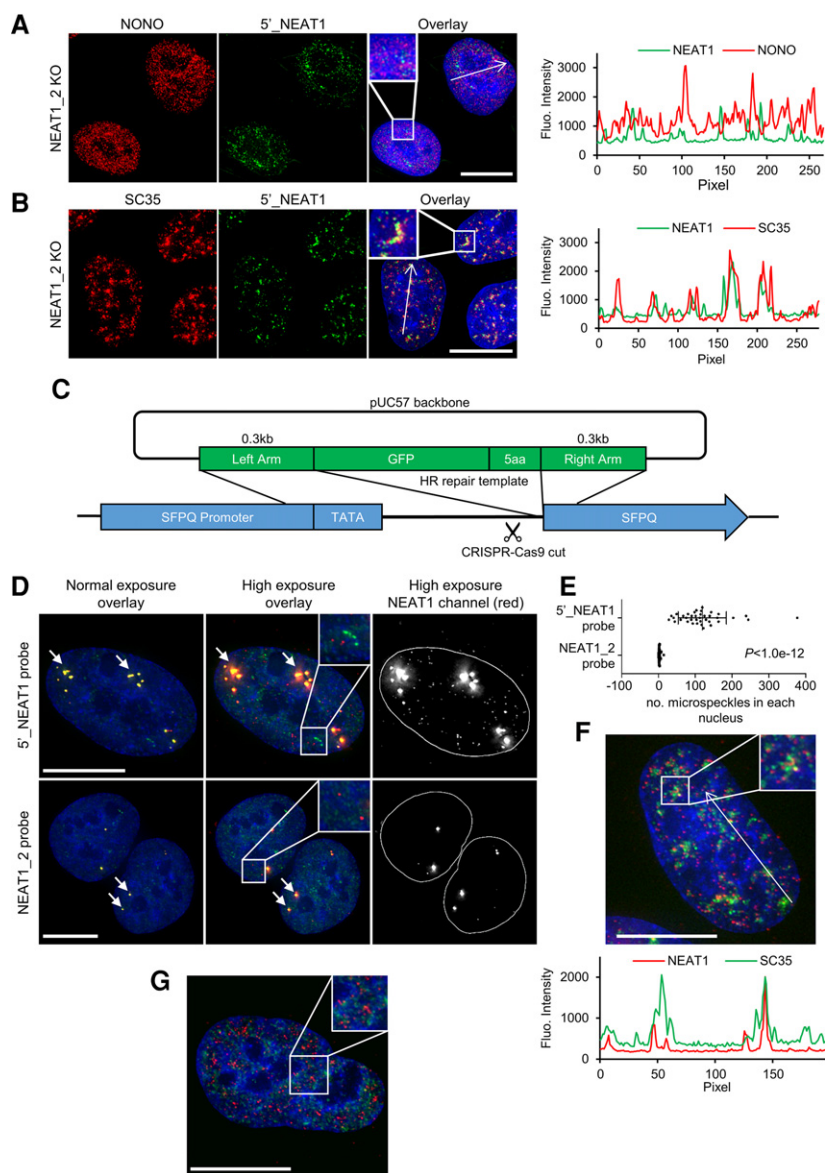
FIGURE 2. Generation of U2OS cell lines lacking expression of NEAT1_2. (*A*) A schematic representation of the CRISPR-Cas9 mediated NEAT1_2 knockout strategy. A YFP expression cassette was inserted downstream from the poly(A) signal after the NEAT1_1 isoform to disrupt the gene structure of NEAT1. sgRNA_3 was used to guide the Cas9 to create the DSB required for the homologous recombination and subsequent repair with the YFP cassette. Homozygous clones were again screened using two PCR primer sets indicated by gray, purple, and red arrows. (*B*) NEAT1_2 knockout efficiency in two representative homozygous clonal lines (col. 44, and 85) as measured by RT-qPCR. (*C*) Immunofluorescence of paraspeckle protein NONO (red) and RNA-FISH of NEAT1_2 (green) showed complete loss of NEAT1_2 RNA signal as well as any typical paraspeckle-like localization of NONO following the NEAT1_2 knockout. DAPI counterstaining is shown in blue. Scale bar = 10 μm. RT-qPCR results were one representative measurement plotted as mean ± SD of two technical replicates normalized to U6. (*D*) RNA-seq tracks of mapped reads (blue) to the edited *NEAT1* gene with YFP insertion indicated proper termination of NEAT1_1 and YFP transcripts in the NEAT1_2 KO cells. Transcription start sites (arrows) and terminations sites (T1–T3) are annotated. Histograms represent number of the overlapped reads for each nucleotide on a linear scale from 0 to 300. A small number of mismapped reads in the YFP cassette can be seen in the WT track that is supposed to have no YFP insertion.

targeted to the NEAT1 5′-end (which can only detect NEAT1_1 in this case where functional NEAT1_2 is lacking), we observed nuclear puncta that were smaller and more uniformly distributed throughout the nucleus than typical paraspeckles, with an absence of NONO colocalization to indicate they were not bona fide paraspeckles (Fig. 3A). While it cannot be ruled out that some of the minor pool of artificial NEAT-YFP fusion transcripts might be detected by this probe set and contribute to this signal, the relative overabundance of NEAT1_1 to any of these artificial fusions (Fig. 2D) argues strongly that these foci largely consist of NEAT1_1. The NEAT1_1 foci, hereafter referred to as "microspeckles," are located throughout the nucleus, and in some instances reside close to nuclear speckles, marked by SC35 (Fig. 3B). This apparently nonparaspeckle localization of NEAT1_1 is reminiscent of the relationship between hsr-ω-c transcript and Omega speckles in the *Drosophila* nucleus. In *Drosophila* nuclei, hsr-ω-c transcripts completely overlap the 5′ end of hsr-ω-n transcripts, which, like NEAT1_2, is the essential component for forming nuclear Omega speckles. However, hsr-ω-c transcripts localize to the cytoplasm (Garbe and Pardue 1986; Fini et al. 1989).

We next investigated the possibility that NEAT1_1 nonparaspeckle foci might also exist in normal WT U2OS cells. To facilitate easier observation of paraspeckles, we generated U2OS cells with the endogenous essential DBHS paraspeckle protein SFPQ (splicing factor proline/glutamine-rich) tagged with GFP at its N terminus, using another variant of a CRISPR-Cas9 genome editing strategy (Fig. 3C). This method involved inducing a double-stranded break (DSB) proximal to the start ATG of *SFPQ*, while at the same time supplying a repair plasmid containing cDNA encoding AcGFP with arms of the homologous sequence matched to either side of the cut site. With this strategy, only successfully engineered cells exhibited green fluorescence driven by the *SFPQ* promoter, therefore a pooled heterogeneously labeled population was easily selected by successive rounds of FACS for GFP fluorescence.

can be easily resolved. Taking advantage of our NEAT1_2$^{-/-}$ cells, our examination of the NEAT1_1 transcript localization yielded an interesting observation: Using RNA-FISH probes

FIGURE 3. NEAT1_1 localizes to numerous microspeckles in the absence of NEAT1_2. (*A,B*) Immunofluorescence and 5′_NEAT1 FISH on U2OS NEAT1_2 knockout cell lines showing that NEAT1_1 (green) no longer colocalizes with NONO (red) (*A*), but has a close spatial relationship with SC35 (red) (*B*). A line graph of both fluorescence intensities was provided for the entire length of each long arrow to show the localization pattern. (*C*) A graphic representation of the CRISPR double-nickase genome editing strategy for the generation of endogenous SFPQ tagged with GFP at the N terminus in U2OS cells. A GFP open reading frame with 5aa linker was incorporated via HR into the space between the TATA box of SFPQ and its first exon. (*D*) Endogenously tagged SFPQ with GFP (green) colocalized with NEAT1 as shown by the RNA-FISH probes (red) targeting either the 5′ region of NEAT1 (*top*, labels, both NEAT1_1 and NEAT1_2) or the NEAT1_2 probe (*bottom*). Both probes show colocalization in paraspeckles (arrows). The *middle* and *right* panels show that increasing the brightness on both images reveals that the 5′ NEAT1 probe is frequently found in many foci outside paraspeckles. These additional mini foci lack colocalization with SFPQ. (*E*) Number of microspeckles in each nucleus using different probes; error bars represent mean ± SD from more than 35 cell nucleus; *P*-value was from two-tailed Student *t*-test. DAPI counterstaining is shown in blue. (*F*) NEAT1_1 (red) colocalizes with SC35 (green) in U2OS GFP-SFPQ cells (GFP fluorescence was not included in the image and SC35 was pseudo-colored to green to enhance the viewer's visual discrimination between the NEAT1 and SC35 signals). A line graph of both fluorescence intensities was provided for the entire length of the long arrow to show the colocalization pattern. (*G*) NEAT1_1 probe (red) showed different localization to the GFP-SFPQ (green) in cells with no detectable paraspeckles. DAPI counterstaining is shown in blue. Scale bar = 10 μm.

We next performed RNA-FISH on the U2OS-GFP-SFPQ cells with probes that either recognize both NEAT1 isoforms (NEAT1 5′ probe set, Fig. 3D, top), or solely the NEAT1_2 isoform (Fig. 3D, bottom). When imaged normally, and with the brightness scaled to contrast paraspeckles nicely with the background, both probe sets showed a similar colocalization of NEAT1 with GFP-SFPQ inside typical paraspeckles with no obvious nonparaspeckle nuclear signal (Fig. 3D, left images). However, when these same images were subject to increased brightness scaling for the red FISH channel, numerous microspeckles were observed only with the 5′ NEAT1 probe, but not with the NEAT1_2 specific probe (Fig. 3D, middle and right images). Counting with nonsubjective strict fluorescence thresholding parameters confirmed the presence of many hundreds of microspeckles in cells labeled with the NEAT1 5′ probe that were absent in cells labeled with the NEAT1_2 specific probe (Fig. 3E). In addition, as with the NEAT1_2$^{-/-}$ lines, a proportion of microspeckles appear to reside closely next to nuclear speckles (Fig. 3F), and numerous microspeckles could still be found in cells that naturally have no obvious paraspeckles (Fig. 3G). These microspeckles thus appear to be primarily composed of NEAT1_1, whereas NEAT1_2 is restricted to paraspeckles. Thus, these observations in WT cells demonstrate that microspeckles are independent of NEAT1_2 KO and represent natural NEAT1_1 localization. Lastly, we performed absolute qPCR and RT-qPCR using a human NEAT1_1 plasmid, coupled with primers that amplified both overlapping endogenous NEAT1 transcripts, or NEAT1_2 alone, to estimate the amount of NEAT1_1 and NEAT1_2 in WT cells. The result showed there are approximately 157 ± 16 (mean ± SD) NEAT1_1 molecules and about 26 NEAT1_2 molecules per U2OS cell (data not shown). Given there are approximately 110 microspeckles per nucleus (Fig. 3E), this suggests that a large proportion of the NEAT1_1 likely exists in the form of microspeckles, not paraspeckles. Further, some microspeckles

might only be a single NEAT1_1 molecule, given that the Stellaris RNA-FISH probe sets are an array of up to 48 probes for each RNA target that achieves both high specificity and sensitivity capable of detecting a single RNA transcript without interference from weak off-target signals (Raj et al. 2008). In summary, the creation of the NEAT1_2$^{-/-}$ lines and microscopic analyses revealed the localization of the NEAT1_1 isoform outside paraspeckles with no obvious relationship with DBHS proteins, but instead occasionally associated with SC35 domains. It is intriguing to hypothesize that these microspeckles could be potential sites for new paraspeckle-independent functions of NEAT1_1. Interestingly, nuclear speckles are also known to be the localization sites for splicing factors, MALAT1, certain actively transcribed genes, and polyadenylated transcripts (Carter et al. 1991; Visa et al. 1993; Xing et al. 1995; Smith et al. 1999; Hutchinson et al. 2007). Future experiments could address the possibility that these NEAT1_1 microspeckles are related to the CHART data that showed MALAT1 and NEAT1 both binding to chromatin of actively transcribed genes (West et al. 2014). It would also be illuminating in future experiments to examine the protein binding partners of microspeckles in the NEAT1_2$^{-/-}$ cells in order to reveal potential functions.

At this stage, our observation of NEAT1_1 throughout the nucleoplasm of both WT and NEAT1_2$^{-/-}$ cells revealed failure of NEAT1_1 to colocalize with DBHS proteins that are essential for paraspeckle formation. Thus, NEAT1_1 is not capable of supporting the formation of paraspeckles. Furthermore, NEAT1_1 may also be incapable of binding NONO and SFPQ at a level that can be easily observed through microscopy. In this regard, an earlier study of cross-linking and immunoprecipitation of NONO and SFPQ also showed considerably reduced signal corresponding to the 5′ end of NEAT1_2 that overlaps NEAT1_1, compared to strong binding for the rest of NEAT1_2 (Sasaki et al. 2009). Given that NEAT1_1 is much more abundant than NEAT1_2 in cells, this probably suggests selective binding of DBHS proteins to NEAT1_2, not NEAT1_1 inside paraspeckles. In other words, there is a possibility that NEAT1_1 may not be a major component of paraspeckles.

NEAT1_1 is insignificant in controlling paraspeckles' number and size

As transient overexpression of the 23 kb NEAT1_2 is technically challenging and does not recapitulate the local chromatin environment upon which paraspeckle formation depends (Mao et al. 2011; Naganuma et al. 2012), we next wanted to selectively increase the ratio of NEAT1_2 compared to NEAT1_1 in WT cells; therefore we carried out a fourth type of genome editing strategy. We used CRISPR-Cas9 double nickase with homologous recombination to mutate several previously identified canonical polyadenylation and cleavage signals (CFIm sites) required for NEAT1_1 production, aiming to switch from the production of the naturally

more favored NEAT1_1, to the less favored NEAT1_2 (Fig. 4A; Naganuma et al. 2012). With only a transient fluorescence selection component, we found isolation of successfully engineered lines challenging, but nevertheless succeeded in generating two heterozygous clones (NEAT1 isoform switched lines—NEAT1_IS$^{+/-}$) from WT U2OS cells. Although only heterozygous mutants, these lines nevertheless demonstrated two- to threefold increased NEAT1_2 expression (Fig. 4C), while the total NEAT1 level decreased nonsignificantly (Fig. 4B). In addition, there was a proportional increase in the percentage of total NEAT1 that was derived from NEAT1_2 (∼10% in the WT cells, increasing to 26%–37% in the NEAT1_IS$^{+/-}$ lines, Fig. 4D). It is also interesting to note that NEAT1_IS$^{+/-}$ cells did not show more than a 50% "switch" from NEAT1_1 to NEAT1_2 (Fig. 4D), perhaps indicating the presence of alternative mechanisms of determining NEAT1_1 versus NEAT1_2 ratios in addition to the canonical polyadenylation process (Naganuma et al. 2012). We assessed if this reduction in the percentage of NEAT1_1 transcripts generated from the locus affected paraspeckle levels. We observed a small, yet statistically significant, increase in the number of paraspeckles and a marginal increase in their size (Fig. 4E,F). These increases in paraspeckle numbers reinforce the notion that NEAT1_1 does not play a role in directing paraspeckle number and size and support the proven role of NEAT1_2 in this function.

In conclusion, our data, combined with previous observations, suggest that NEAT1_1 is very likely not a major component of paraspeckles. First, the paraspeckle proteins NONO and SFPQ do not clearly colocalize with microspeckles, suggesting a lack of robust interaction with NEAT1_1, at least in the absence of the scaffolding and paraspeckle-enriching molecule, NEAT1_2 (Fig. 3A,D). Second, previous immunoprecipitation studies demonstrated preferential binding of NONO and SFPQ to the unique NEAT1_2 sequence, compared to the overlapping NEAT1_1 and NEAT1_2 sequence (Sasaki et al. 2009). Given these two observations, we conclude that the likelihood is that DBHS proteins bind to NEAT1_2 in paraspeckles with an absence of high levels of NEAT1_1 in paraspeckles. Lastly, NEAT1_1 lacks the ability to control the number and size of paraspeckles, in contrast to NEAT1_2, again implying that NEAT1_1 is not likely a major paraspeckle component. In this study, we also describe NEAT1_1 labeled "microspeckles" that can closely associate with nuclear speckles, potentially representing a previously unknown paraspeckle-independent functional location of NEAT1_1. Thus, by taking advantages of various genome editing strategies, this study highlights the different biological roles of the two major isoforms of NEAT1. We would encourage future studies in the field to treat these two isoforms of NEAT1 more separately to understand the broader biological significance of the *NEAT1* locus. This will be particularly important given the current interest in potentially developing agents to target NEAT1 in cancer.
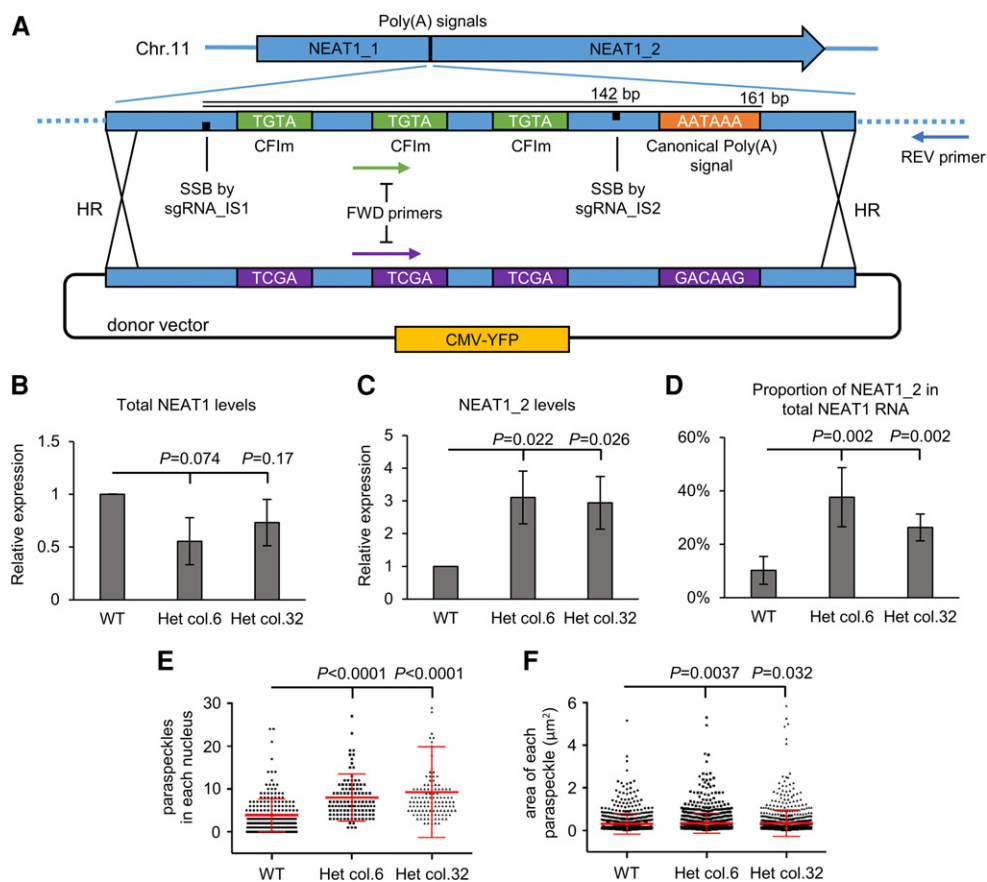
FIGURE 4. Switching between the production of NEAT1 isoforms through genomic editing. (*A*) A schematic representation of the CRISPR editing strategy to switch the transcription of NEAT1 toward the production of NEAT1_2 isoform instead of NEAT1_1. All known signals [CFIm and canonical poly(A) signal] that could contribute to the polyadenylation of NEAT1_1 were identified and mutated using a double nickase Cas9 system (sgRNA_IS1 and IS2) alongside donor repair template in a YFP expression vector. Primers used for genotyping screening are indicated by arrows. Only two heterozygous and no homozygous clones were identified. The graphic representation is not scaled. (*B–D*) RT-qPCRs comparing two heterozygous NEAT1 isoform switched lines (NEAT1_IS lines) with the WT lines. (*B*) Total NEAT1 levels, (*C*) NEAT1_2 levels, and (*D*) proportion of NEAT1_2 in total NEAT1 levels. All RT-qPCR were normalized to housekeeping gene RPLP0. *P*-values are one-tailed Student *t*-test from four biological replicates. Error bars represent mean ± SD. (*E*) Number of paraspeckles per cells and (*F*) their areas were counted and analyzed using nonparametric one-way ANOVA with multiple testing, $n > 100$ cells in each cell type, paraspeckle count in each cell was plotted as scatter plots, and mean ± SD were plotted.

MATERIALS AND METHODS

Cell culture and treatments

U2OS osteosarcoma cells were grown in DMEM (Invitrogen) and supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin. All cells were cultured in a 37°C incubator supplied with 5% CO_2.

Molecular cloning

For genome engineering, Phusion DNA polymerases (New England Biolabs) were used for all PCRs to create repair templates. Gibson cloning (New England Biolabs) was used for incorporating repair templates into either pEGFP-C1 (Clontech) or pUC57 plasmids and transformed into DH5α *E. coli*. For screening cell clones with genomic editing, genomic DNA of clones were extracted using the Purelink Genomic DNA Mini Kit (Invitrogen) and genotyping of

PCRs was performed using MyTaq DNA polymerase (Bioline). Details of primers and repair templates are listed in Supplemental Table S1.

RT-qPCR

RNA was first extracted using TRIsure (Bioline), and 100 ng of extracted RNA was converted into cDNA using the QuantiTect Reverse Transcription Kit (Qiagen) with the instructions provided. Quantitative real-time PCR was performed using the SensiMix SYBR No-ROX kit (Bioline) and Rotor-Gene Q Real-time PCR cycler. Fluorescence was detected at a temperature 3°C lower than the T_m of the specific product to increase the specificity of detection. Cycle threshold (Ct) values of each product were determined under a constant normalized fluorescence threshold of 0.075. Relative expression of genes was calculated using the $2^{-\Delta\Delta Ct}$ method against housekeeping genes (Livak and Schmittgen 2001). Quantitation of the proportion of NEAT1_2 transcripts in total NEAT1 transcripts

was performed using the "Comparative Quantitation" software supplied by Corbett Research for the Rotergene, as previously described (Warton et al. 2004). RT-qPCR primers are listed in the Supplemental Table.

RNA-FISH and immunofluorescence

Cells grown on coverslips were fixed using 4% paraformaldehyde and permeabilized by 70% ethanol overnight. For RNA-FISH, Stellaris RNA-FISH probes labeled with Quasar 570 Dye for 5′ NEAT1 (SMF-2036-1) and NEAT1_2 (SMF-2037-1) (1:100, Biosearch Technologies) were used and the procedure was carried out according to the instructions provided. For subsequent immunofluorescence staining, NONO antibody (1:500, housemade in Souquere et al. 2010) and SC35 antibody (1:500, Sigma-Aldrich, S4045) were used, and to be stained with Alexafluor 647 antibody (1:250, Thermofisher Scientific, A-21237). Finally, cells were counterstained with 1:15,000 DAPI (4′,6-diamidino-2-phenylindole) in water for 2 min at room temperature and then mounted onto slides with VectaShield (Vector Laboratories) mounting media. Fluorescence signals were imaged using the DeltaVision Elite Imaging System and Softworx software (GE Healthcare). Images were Z-stacks of 0.2 μm sections, subjected to deconvolution, and projected with maximum intensity. For counting paraspeckles, images were taken using a 60× objective lens using the same exposure settings. Paraspeckles were defined as NEAT1_2 RNA-FISH signals that are colocalizing with NONO and are larger than 6 pixels. For counting microspeckles, images were taken using a 100× objective lens using the same exposure settings. Microspeckles were considered as NEAT1 RNA-FISH foci that are between 2 and 10 pixels.

CRISPR genomic engineering

PX330 and pX335 CRISPR-Cas9 WT or nickase plasmids were obtained as a kind gift from Feng Zhang's laboratory (Cambridge, Massachusetts) via Addgene for the cloning of the guide RNAs (sgRNA) (Cong et al. 2013). sgRNAs were designed using a CRISPR design tool available on http://crispr.mit.edu (Hsu et al. 2013). Annealed double-stranded DNA oligonucleotides for sgRNAs were ligated into the BbsI restriction site in the plasmids. For total NEAT1 KO and NEAT1_2 KO editing, we amplified homologous sequences each of approximately 500 bp from the genomic DNA of healthy human sample, and cloned each pair of homologous repair arms into the AseI and MluI sites flanking each side of the fluorescent protein expression cassette in the pEGFP-C1 backbone (Clontech) using Gibson cloning (New England Biolabs). For NEAT1_IS editing, only the MluI site was used for repair template cloning. The vector backbone also carried an in-house generated mutation that turned the original EGFP to a YFP. For tagging GFP to SFPQ, repair templates that contain AcGFP were incorporated into a pUC57 vector for propagation by supplier (Genscript). To transfect cells, plasmids containing CRISPR components and repair template plasmids were cotransfected with a 1:1 ratio using Lipofectamine 3000 (Invitrogen) into the cells, which were seeded at 1×10^5 cells per well of a 12-well plate. Transfected cells were passaged for 1 wk and then sorted by FACS. For single cell clones, fluorescence positive cells were seeded into 96-well plates 2 wk after the initial transfection. Approximately 2–3 wk after the seeding, genomic DNA of each colony was extracted

and tested by PCR to screen for genomic editing events. Cell clones that were transfected but did not undergo targeted genome editing were taken as WT controls. Details of sgRNAs, repair templates, and PCR screening primers are listed in the Supplemental Table.

RNA sequencing and analysis

U2OS WT and NEAT1_2$^{-/-}$ cells had their transcriptome analyzed. Total RNA was prepared from cultured cells using TRIsure (Bioline). Ribosomal RNA was depleted in all samples before library preparation and sequencing, which was carried out by the Australian Genome Research Facility (AGRF). Using the Illumina TruSeq Stranded Total Library Prep Kit with the Ribo-Zero Gold Kit for the removal of ribosomal RNA, the entire transcriptome was sequenced on the Illumina HiSeq 2000 platform to generate 25–28 million 50-bp single-end reads for each sample, and all samples were sequenced in a single lane. Sequenced reads files were uploaded onto the Galaxy platform (http://galaxyproject.org), converted to FASTQ format by FASTQ Groomer (Blankenberg et al. 2010), and then aligned to the human Hg19 reference genome with an YFP expression cassette inserted into the designated genomic location of NEAT1_2 using Tophat2 (version 0.6) with the default settings for a minimum of 97% reads being mapped (Kim et al. 2013). Mapped reads were visualized using IGV (version 2.3.88) using default settings (Robinson et al. 2011; Thorvaldsdóttir et al. 2013).

REFERENCES

Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, Boeckx B, Wijnhoven PWG, Radaelli E, Vermi W, et al. 2016. p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat Med* **22:** 861–868.

Barry G, Briggs JA, Hwang DW, Nayler SP, Fortuna PRJ, Jonkhout N, Dachet F, Maag JLV, Mestdagh P, Singh EM, et al. 2017. The long non-coding RNA NEAT1 is responsive to neuronal activity and is associated with hyperexcitability states. *Sci Rep* **7:** 40127.

Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* **89:** 19.10.1–19.10.21.

Carter KC, Taneja KL, Lawrence JB. 1991. Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. *J Cell Biol* **115:** 1191–1202.

Chai Y, Liu J, Zhang Z, Liu L. 2016. HuR-regulated lncRNA NEAT1 stability in tumorigenesis and progression of ovarian cancer. *Cancer Med* 5: 1588–1598.

Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, 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 X, Hao Z, Zhou J, Fan S, Luo G, Zhang L, Zhang K, Zhang Y, Liang C. 2016. Promoting progression and clinicopathological significance of NEAT1 over-expression in bladder cancer. *Oncotarget* doi: 10.18632/oncotarget.10084.

Choudhry H, Albukhari A, Morotti M, Hider S, Moralli D, Smythies J, Schödel J, Green CM, Camps C, Buffa F, et al. 2014. Tumor hypoxia induces nuclear paraspeckle formation through HIF-2a dependent transcriptional activation of NEAT1 leading to cancer cell survival. *Oncogene* 34: 4482–4490.

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.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819–823.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74.

The FANTOM Consortium and the RIKEN PMI and CLST (DGT). 2014. A promoter-level mammalian expression atlas. *Nature* 507: 462–470.

Fini ME, Bendena WG, Pardue ML. 1989. Unusual behavior of the cytoplasmic transcript of hsr omega: an abundant, stress-inducible RNA that is translated but yields no detectable protein product. *J Cell Biol* 108: 2045–2057.

Fox AH, Lam YW, Leung AKL, 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.

Fu JW, Kong Y, Sun X. 2016. Long noncoding RNA NEAT1 is an unfavorable prognostic factor and regulates migration and invasion in gastric cancer. *J Cancer Res Clin Oncol* 142: 1571–1579.

Garbe JC, Pardue ML. 1986. Heat shock locus 93D of *Drosophila melanogaster*: a spliced RNA most strongly conserved in the intron sequence. *Proc Natl Acad Sci* 83: 1812–1816.

Gibb EA, Vucic EA, Enfield KSS, Stewart GL, Lonergan KM, Kennett JY, Becker-Santos DD, MacAulay CE, Lam S, Brown CJ, et al. 2011. Human cancer long non-coding RNA transcriptomes. *PLoS One* 6: e25915.

Gutschner T, Baas M, Diederichs S. 2011. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res* 21: 1944–1954.

Hirose T, Virnicchi G, Tanigawa A, Naganuma T, Li R, Kimura H, Yokoi T, Nakagawa S, Bénard M, Fox AH, et al. 2014. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* 25: 169–183.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech* 31: 827–832.

Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 8: 39.

Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K, Kato A, Kawaguchi Y, Sato H, Yoneda M, 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.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14: R36.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* 25: 402–408.

Ma Y, Liu L, Yan F, Wei W, Deng J, Sun J. 2016. Enhanced expression of long non-coding RNA NEAT1 is associated with the progression of gastric adenocarcinomas. *World J Surg Oncol* 14: 41.

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.

Mattick JS. 2011. The central role of RNA in human development and cognition. *FEBS Lett* 585: 1600–1616.

Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10: 155–159.

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, Shimada M, Yanaka K, Mito M, Arai T, Takahashi E, Fujita Y, Fujimori T, Standaert L, Marine JC, et al. 2014. The lncRNA *Neat1* is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* 141: 4618–4627.

Nishimoto Y, Nakagawa S, Hirose T, Okano HJ, Takao M, Shibata S, Suyama S, Kuwako KI, Imai T, Murayama S, et al. 2013. The long non-coding RNA nuclear-enriched abundant transcript 1_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. *Mol Brain* 6: 31.

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.

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5: 877–879.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotech* 29: 24–26.

Saha S, Murthy S, Rangarajan PN. 2006. Identification and characterization of a virus-inducible non-coding RNA in mouse brain. *J Gen Virol* 87: 1991–1995.

Sasaki YTF, Ideue T, Sano M, Mituyama T, Hirose T. 2009. MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci* 106: 2525–2530.

Smith KP, Moen PT, Wydner KL, Coleman JR, Lawrence JB. 1999. Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J Cell Biol* 144: 617–629.

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.

Sun C, Li S, Zhang F, Xi Y, Wang L, Bi Y, Li D. 2016. Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of miR-377-3p-E2F3 pathway. *Oncotarget* 7: 51784–51814.

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. 2009. *MEN* ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 19: 347–359.

Sunwoo JS, Lee ST, Im W, Lee M, Byun JI, Jung KH, Park KI, Jung KY, Lee SK, Chu K, et al. 2016. Altered expression of the long noncoding RNA NEAT1 in Huntington's disease. *Mol Neurobiol* 54: 1577–1586.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192.

Visa N, Puvion-Dutilleul F, Harper F, Bachellerie JP, Puvion E. 1993. Intranuclear distribution of poly(A) RNA determined by electron microscope in situ hybridization. *Exp Cell Res* **208:** 19–34.

Wang P, Wu T, Zhou H, Jin Q, He G, Yu H, Xuan L, Wang X, Tian L, Sun Y. 2016. Long noncoding RNA NEAT1 promotes laryngeal squamous cell cancer through regulating miR-107/CDK6 pathway. *J Exp Clin Cancer Res* **35:** 22.

Warton K, Foster NC, Gold WA, Stanley KK. 2004. A novel gene family induced by acute inflammation in endothelial cells. *Gene* **342:** 85–95.

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, et al. 2016. Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J Cell Biol* **214:** 817–830.

Xing Y, Johnson CV, Moen PT, McNeil JA, Lawrence J. 1995. Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J Cell Biol* **131:** 1635–1647.