Science Advances
AAAS

Manuscript Template

**Title:** Integrator restrains paraspeckles assembly by promoting isoform switching of the lncRNA *NEAT1*

**Running Title:** Mechanism of paraspeckles assembly and chemosensitivity

**Authors**

Jasmine Barra[1,2,7,8,10], Gabriel S Gaidosh[3], Ezra Blumenthal[3], Felipe Beckedorff[3], Mina M Tayari[3], Nina Kirstein[3], Tobias K Karakach[4,5], Torben Heick Jensen[6], Francis Impens[7,8,9], Kris Gevaert[7,8], Eleonora Leucci[10,] †, Ramin Shiekhattar[3,] † & Jean-Christophe Marine[1,2,] †,*

**Affiliations**

[1]Laboratory for Molecular Cancer Biology, Center for Cancer Biology, VIB, Leuven, Belgium. [2]Laboratory for Molecular Cancer Biology, Department of Oncology, KULeuven, Leuven, Belgium.

[3]University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Department of Human Genetics, 1501 NW 10th Avenue, Miami, FL 33136.

[4]Bioinformatics Core Laboratory, Children's Hospital Research Institute of Manitoba (CHRIM), Winnipeg, Manitoba, Canada.

[5]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada.

[6]Department of Molecular Biology and Genetics, Aarhus University, Denmark.

[7]VIB Center for Medical Biotechnology, VIB, 9000, Ghent, Belgium.

[8]Department of Biomolecular Medicine, Ghent University, 9000 Ghent, Belgium.

[9]VIB Proteomics Core, 9000 Ghent, Belgium.

[10]Laboratory for RNA Cancer Biology, Department of Oncology, LKI, KULeuven.

†These authors contributed equally to this work

* Correspondence should be addressed to jeanchristophe.marine@kuleuven.vib.be

## Abstract

Alternative RNA 3'-end processing provides an important source of transcriptome diversification, which impacts various biological processes and the etiology of diseases, including cancer. A prime example of this is the transcript isoform switch that leads to the read-through expression of the long non-coding RNA *NEAT1_2*, at the expense of the shorter polyadenylated transcript *NEAT1_1*. Expression of *NEAT1_2* is required for the formation of paraspeckles (PS), nuclear bodies that protect cancer cells from oncogene-induced replication stress and chemotherapy. Searching for proteins that modulate this isoform switching event we identified factors involved in the 3'-end processing of polyadenylated ($pA^+$) RNA as well as several components of the Integrator complex. Perturbation experiments established that, by promoting the cleavage of *NEAT1_2*, Integrator forces *NEAT1_2* to *NEAT1_1* isoform switching and, thereby, restrains PS assembly. Consistently, low expression levels of several Integrator subunits correlated with poorer prognosis of cancer patients exposed to chemotherapeutics. Our study identifies Integrator as a key regulator of PS biogenesis and establish a link between Integrator, cancer biology and chemosensitivity, which may be exploited therapeutically.


## Teaser

Integrator is a key regulator of *NEAT1* isoform switching and, thereby, paraspeckles biogenesis and chemosensitivity.

## MAIN TEXT

### Introduction

Most human genes have multiple sites at which RNA 3′-end cleavage and polyadenylation can occur (*1*). Alternative 3'-end cleavage gives rise to transcript isoforms that differ either in their coding sequences (CDSs) or their 3′-UTRs and, thus, contribute to transcriptome diversification (*1*, *2*). Remodeling of 3′-UTRs can have particularly profound phenotypic consequences, as such transcript isoforms may differ in their relative stability, localization, translation rate, and/or function (*2*). Although it is well-known that RNA 3′-end processing can be finely regulated depending on the cellular needs, the factors involved in alternative 3'-end processing are only partially characterized.

The core pre-mRNA 3′-end processing complex consists of four subcomplexes, namely cleavage and polyadenylation factor (CPSF), cleavage stimulation factor (CSTF), cleavage factor I (CFI) and CFII. A few other proteins, including Symplekin and poly(A) polymerase (PAP), are also involved in completing the 3'-end formation of polyadenylated ($pA^+$) RNA (*3*). In metazoans, sites of pre-mRNA polyadenylation are primarily defined by the canonical poly(A) signal AAUAAA, which is positioned ~21 nucleotides upstream of the cleavage site (*3*). This hexamer is recognized by the co-transcriptionally recruited CPSF subcomplex, which carries out the endonucleolytic cleavage event followed by the addition of a poly(A) tail to the 5' cleavage product by PAP. Deregulation of protein expression levels and/or activity of core 3′-end processing factors can obviously contribute to 3'-end processing rewiring either globally or specifically, and thereby affect transcriptome diversification in response to specific environmental cues. Moreover, many other RNA binding proteins (RBPs) can influence RNA 3'-end processing, often depending on the binding positions within mRNA target 3'-UTRs (*4*).

Other protein complexes are involved in the 3'-end processing of non-polyadenylated RNA species. For instance, the Integrator complex, which binds to the C-terminal domain (CTD) of the RNA polymerase II, is responsible for the RNA 3'-end processing of UsnRNA (*5*). This complex has been shown to control termination of transcription and 3'-end processing at enhancer RNA (eRNA) and replication-dependent histones loci (*6*, *7*). Furthermore, Integrator binding to the proximal promoter region of polyadenylated genes negatively regulates their expression (*7*).

Alternative 3'-end processing-dependent transcriptome diversification plays key roles in various important biological processes (*4*). Individual 3'-end processing events have also been implicated in pathological conditions, including autoimmune disorders and cancer (*4*). Consistent with a reported general association between the expression of short RNA 3′-UTRs and a proliferative

cellular state (*8*), most cancers express transcripts with shorter 3′-UTRs than those expressed in corresponding normal tissues (*4*). Some studies have attributed cancer-related 3'-end RNA patterns to the deregulated activity of specific 3′-end processing factors, such as CSTF2 (*9*) and CFIm25 (*10*). However, the motifs recognized by these core 3′-end processing factors do not explain the observed quantitative changes in poly(A) site usage between tumor and normal tissue samples from The Cancer Genome Atlas (*11*), indicating that other unknown modulators also contribute.

Deregulation of RNA 3'-end processing at specific loci may also contribute to tumor growth as illustrated in recent findings implicating the lncRNA locus *NEAT1* in cancer development (*12*). This locus produces two lncRNA isoforms (*13*). The shorter isoform, *NEAT1_1* (3700 nt in length), contains a functional polyA site. The long *NEAT1_2* isoform (22700 nt in length), which is not polyadenylated, is produced as a read-through transcript when the 3'-end processing of *NEAT1_1* is inefficient (*14*). The mechanisms underlying *NEAT1* isoform switching remain poorly understood. The ubiquitous nucleic acid binding protein hnRNPK has been implicated in this process, by competing with CPSF6 for the binding of NUDT21 and impairing *NEAT1_1* polyadenylation (*14*). Moreover TDP-43 enhances *NEAT1_1* polyadenylation in pluripotent cells (*15*). Whereas the function of *NEAT1_1* still needs to be established (*16*, *17*), *NEAT1_2* is an essential architectural component of paraspeckles (PS) (*18*), which are highly ordered and phase-separated nuclear stress bodies (*19*). Thus, PS assembly critically depends on the poorly understood *NEAT1* isoform switch. Expression of *NEAT1_2,* and thereby PS assembly, can only be detected under specific physiological conditions (i.e. lactating mammary glands) and in response to various forms of stresses, including oncogenic stress (*12*, *20–22*). Accordingly, PS appear in over 65% of human epithelial cancers (*12*), where they predict poor prognosis (*23*) and are either completely absent, or only sporadically detectable, in normal tissues (*12*, *24*). In a classical two-stage chemically-induced skin cancer mouse model, PS are induced in skin epidermal cells exposed to oncogenic stress, while genetic ablation of *NEAT1* dramatically impairs tumor initiation and progression into aggressive and invasive lesions (*12*). However, mouse skin that lacks only the short *Neat1_1* isoform does not exhibit these protective properties (*17*). Critically, specific downregulation of *NEAT1_2* using antisense oligonucleotides sensitized a series of epithelial cancer cell lines to various clinically-relevant anti-cancer therapeutics (*12*). Hence, these studies identified *NEAT1_2*, and by extension PS, as promising cell-specific therapeutic targets for the chemo-sensitization of a wide range of epithelial cancers. We therefore reasoned that a better understanding of pathways and factors/enzymes involved in the molecular mechanisms underlying *NEAT1* isoform switching and, thereby, PS biogenesis may lead to the identification of targets that are amenable to conventional therapeutics.

## Results

### Identification of Integrator as a novel *NEAT1* RNA interactor

To identify proteins that modulate *NEAT1_2* expression, and consequently PS biogenesis, we adapted an RNA Antisense Purification (RAP) protocol that we had previously used to identify interactors of the melanoma-specific lncRNA *SAMMSON* (*25*). Complexes directly bound to the endogenous *NEAT1* transcript were purified from freshly isolated nuclei of MCF-7 cells exposed to UV crosslinking, using tiling DNA-based biotinylated oligonucleotides, targeting the 5' portion of *NEAT1* (N1_5'). In parallel, control probes (Ctrl) were designed against the melanoma-specific *LINC00698* transcript, which is not expressed in the human breast adenocarcinoma MCF-7 cells (**Fig. 1A**). The quality of the nuclear isolation was verified by RT-qPCR, assessing the cytoplasmic RNA encoding the 40S ribosomal protein *S14*, as well as *NEAT1* and *MALAT1*, both of which are exclusively nuclear transcripts (**Fig. 1B**). The efficiency and specificity of the N1_5' pull-down was confirmed by RT-qPCR. Whereas a robust signal was detected for total *NEAT1* (NEAT1) and *NEAT1_2* transcripts in the isolated RAP extracts, neither the housekeeping *TBP* and *HPRT1* mRNAs nor the lncRNA *MALAT1*, used as negative controls, were detectable (**Fig. 1C**).

RAP experiments were performed in biological triplicates and purified proteins were analyzed by label-free mass spectrometry (MS). Principal component analysis (PCA) of the replicates confirmed the clustering of the samples into two groups: the control (Ctrl) and the RAP pull-down performed with *NEAT1* specific probes (N1_5') (fig. S1A). We identified 34 proteins, which were significantly enriched by the N1_5' probes (t-test $p$ value<0.05 and FC>1.6), as high-confidence *NEAT1* interactors (table S1). Two of these were known PS proteins (**Fig. 1D**). Gene Ontology (GO) Analysis using the Search Tool for Recurring Instances of Neighboring Genes (STRING; https://string-db.org), indicated that the remaining *NEAT1* interactors are mainly involved in key aspects of RNA biogenesis and processing (table S2). Among these were multiple subunits of the Integrator and mRNA 3'-end processing complexes, including INTS1, INTS3, INTS6, CSTF1, CSTF2, CSTF2T, CSTF3, CPSF1, WDR33, SYMPK and FIP1L1 (**Fig. 1E**, top panels). In fact, one third (11 out of 34) of all high-confidence interactors belonged to these two multi-component protein complexes. Three additional previously unknown *NEAT1* interactors were also identified, namely the F-box protein FBXO11 and its binding partner CUL1, as well as the transcription factor TCF7L2 (**Fig. 1E**, panels below).

These findings were next validated by RAP-western blotting experiments. Using the N1_5' probes, the interactions between *NEAT1* and FBXO11, TCF7L2, CPSF2 (component of the mRNA 3'-end processing machinery), INTS3, INTS11 (catalytic subunit of Integrator complex), INIP (auxiliary

component of the complex), H3 (used as negative control) and the PS proteins PSF, PSPC1, NONO, TDP-43 (used as a positive control) was confirmed (**Fig. 1F**).

**Integrator limits PS biogenesis by promoting *NEAT1* isoform switching**

The identification of several components of the Integrator complex as high-confidence *NEAT1* RNA interactors, raises the possibility that the complex contributes to the regulation of *NEAT1* isoform switching. To test this hypothesis, we first checked for interaction between *NEAT1* transcript and INTS11, the catalytic subunit of Integrator. To this end, enhanced cross-linking and immunoprecipitation (eCLIP), a well-established and comprehensive procedure for the identification of RNA-binding protein targets (*26*), was performed in HeLa cells using two distinct INTS11 specific antibodies (**Fig. 2A**). Quantification of the eCLIP signal (size-matched coverage) relative to the control IgG showed an enrichment of INTS11 binding to the *NEAT1_1* transcript (**Fig. 2B**). *RNU11* and other non-coding RNAs are shown as positive and negative controls respectively. Note that the binding between integrator and *NEAT1_1* could also be validated in MCF-7 cells by RIP qPCR (**Fig. 2C;** fig. S1B). In agreement with previous findings (*7*, *27*), eCLIP detects two major peaks at the 5'-end of the transcript that may be implicated in premature transcriptional termination. Interestingly, another peak is detected immediately upstream of the 3'-end of *NEAT1_1* (**Fig. 2D).** To determine whether Integrator contributes to the 3'-end maturation of *NEAT1_1*, we performed RNA-sequencing (RNA-seq) in HeLa cells expressing a doxycycline-inducible shRNA construct targeting INTS11 or GFP as a control (shCtrl) (**Fig. 2E**). We observed an accumulation of the long *NEAT1_2* isoform in the INTS11 knockdown cells (**Fig. 2F** and fig. S1C). In agreement with the total RNA sequencing data, small-RNA sequencing (smRNA-seq) analysis, which captures cleavage products of INTS11 catalysis (RNA species smaller than 75nt), indicated a decrease in the small-RNA cleavage product of *NEAT1_1* (**Fig. 2F**, panel below). Likewise, 3' mRNA sequencing (3' mRNA-seq), which detects selectively polyadenylated transcripts, confirmed a decrease in the *NEAT1_1* 3'-ends upon integrator KD (fig. S1D).

Importantly, this phenotype was not an off-target effect as it was rescued by concomitant expression of a wild-type form of INTS11 (WT), but not a catalytic dead mutant (E203Q) (**Fig. 2G;** fig. S1E), indicating that the *NEAT1* isoform switching is dependent on INTS11 enzymatic activity. Similar results were obtained for the snRNA *RNU11*, a well-known integrator target (fig. S1F-G).

The phenotype was not cell type-specific as it could be recapitulated in the breast cancer cell line MCF-7 in which INTS11 was silenced by siRNA (**Fig. 2H;** fig. S1H). As expected, the levels of the 3'-end extended product of two well-established Integrator targets, *RNU11* and *RNU12*, as indicated by the percentage of long transcript relative to the gene body (**Fig. 2H**). Together, these data further

supported a direct contribution of Integrator in the regulation of *NEAT1* isoform switching and indicated that the catalytic activity of Integrator is required for the correct processing of the *NEAT1* transcript.

The observed increase in *NEAT1_2* levels upon silencing of INTS11 raised the possibility that Integrator activity limits the formation of PS. Consistent with this possibility, an increase in the number and size of *NEAT1_2* foci was observed by RNA FISH in MCF-7 cells depleted for INTS11 (**Fig. 2I**). Importantly, RNA FISH coupled to immunofluorescence revealed that the foci co-localized with the PS-specific protein PSPC1, thus demonstrating an increase in PS assembly in INT11-depleted cells (**Fig. 2J**; fig. S1I). This observation indicated that Integrator restrains the formation of PS nuclear bodies by promoting *NEAT1_1* expression, at the detriment of *NEAT1_2*, in steady-state conditions.

## Stress does not disrupt *NEAT1*-Integrator interaction and promotes accumulation of Integrator at PS

Various forms of stress stimulate PS formation (*28*, *29*). It was recently shown that exposure of cells to Hydroxy Urea (HU), which inhibits deoxyribonucleotide synthesis and thus causes DNA replication stress, induces the formation of large PS (*17*). Accordingly, increased PS formation was detected by RNA FISH and RNA FISH combined to immunofluorescence for PSPC1 in HU-exposed MCF-7 cells (**Fig. 3A**). The increase in PS formation was the result of the transcriptional upregulation of the *NEAT1* locus (particularly *NEAT1_2*) as demonstrated by the fact that treatment with Actinomycin D (RNA Pol II inhibitor) abolished HU-induced *NEAT1* upregulation (**Fig. 3B**). We reasoned that stress-induced *NEAT1_2* expression and PS formation may be caused, at least in part, by a decrease in the recruitment of Integrator to the *NEAT1* transcript. To test this hypothesis, we performed RAP-MS experiments on freshly isolated nuclei of MCF-7 cells exposed to HU (fig. S2A-B). As expected, the recovery of *NEAT1_2* RNA in this assay was higher in HU- than in DMSO-treated cells (fig. S2A). Surprisingly our RAP-MS data revealed that most of the previously enriched candidates from unstimulated cells were also recovered in stimulated cells (**Fig. 3C**, fig. S2C, and table S3). Integrator subunits and mRNA 3'-end processing factors were also efficiently pulled-down by the RAP N1_5' probes in these experimental conditions. These interactions were further validated by RAP-western blotting analysis (**Fig. 3D**). Immunofluorescence for the PS protein PSPC1, in intact cells exposed to HU, was performed to confirm the co-localization of CPSF1 and CPSF2 (as well as FBXO11 and TCF7L2) with PSPC1 (fig. S2D) and thus their recruitment to PS. Moreover, STochastic Optical Reconstruction Microscopy (STORM) showed a significant co-localization of the Integrator subunit INTS1 with the PS protein NONO in cells exposed to HU (**Fig. 3E**). Taken together

these data demonstrated that DNA damage-induced stress is not sufficient to disrupt the interaction between *NEAT1* and Integrator. Moreover, these experiments also showed that Integrator accumulates to PS nuclear bodies in DNA-damaged cells.

DNA-damage-induced PS formation is, at least partly, a consequence of activation of the p53 transcription factor (fig. S2B), which in turn enhances *NEAT1* promoter activity (*12*, *22*). Accordingly, exposure to the MDM2 antagonist Nutlin-3a, which causes stabilization of p53 without inducing cellular stress responses, was sufficient to increase *NEAT1_2* transcription and enlarged PS (**Fig. 4A-B**). Consistently, just like HU, exposure to Nutlin-3a did not disrupt *NEAT1*-Integrator association as demonstrated by RAP-MS (**Fig. 4C**; fig. S2E and table S4). PCA analysis of all the RAP-MS experiments under control and stressed conditions confirmed the consistency of these results (fig. S2F). RAP-western blotting (**Fig. 4D**), confocal microscopy and super resolution microscopy (**Fig. 4E**, and fig. S2G) further confirmed that activation of p53 by Nutlin-3a is not sufficient to disrupt the interaction between *NEAT1* and Integrator, and that Integrator accumulates to PS nuclear bodies in these experimental conditions.

Note that exposure to HU or Nutlin-3a did not significantly alter the expression levels of various Integrator subunits (**Fig. 5A**). Importantly, the processing of several well-known Integrator targets, including histones (*7*), was severely compromised upon induction of DNA damage or p53 activation (**Fig. 5B-C**). Further evidence of impaired Integrator activity was also obtained in MCF-7 treated with Nutlin-3a, with HU, or transfected with siINT11, by using a reporter construct that directs expression of GFP upon read-through of RNU7 (30) (fig. S3A-B-C). These data indicated that Integrator activity is compromised in cells exposed to stress, possibly as a consequence of its recruitment to PS. Together with the observation that stress does not disrupt the interaction between *NEAT1* and Integrator, these data favor a model in which upregulation of *NEAT1_2* levels and PS formation in stressed cells (exhibiting elevated transcriptional rates of *NEAT1*) occurs because the amount of functional Integrator available to process *NEAT1* transcripts becomes rate-liming. Consistent with such model, overexpression of exogenous INTS11 (fig. S3D) abolished stress-induced upregulation of *NEAT1_2* (**Fig. 5D-E**), decreased paraspeckles assembly (**Fig. 5F**) and phenocopied the decrease in p53 activation and increase in levels of DNA damage observed following *NEAT1_2* KD (**Fig. 5G-H**).

**Low levels of Integrator components correlate with poorer survival and response to chemotherapy**

We previously established a genetic link between PS formation and tumorigenesis and demonstrated that PS can be detected in about 65% of the human carcinomas analyzed, including skin SCC and

ovarian carcinomas. Importantly, we also showed that expression of *NEAT1_2,* but not *NEAT1_1,* reliably predicts the response of ovarian cancer to platinum-based chemotherapy (*12*). Given that Integrator modulates *NEAT1_2* expression and PS biogenesis, we therefore assessed whether correlations between (altered) expression of Integrator subunits and overall patient survival (OS) may exist. Analysis of patients that underwent chemotherapy in the ovarian cancer cohort (GSE30161) analyzed in our previous study (*12*), confirmed that lower levels of INTS10 and INTS11 significantly correlated with worse OS (**Fig. 6A**). Importantly in this cohort the differential expression levels of INT11 and INT10 exhibit an inverse relationship with that of *NEAT1_2* as shown in **Fig. 6B**. Analysis was then expanded to publicly available TCGA datasets corresponding to 11 epithelial cancer cohorts. In addition to gene expression levels, the results were adjusted for the effect of other risk factors (covariates) such as age, race, stage and gender by performing a multivariate analysis using the Cox proportional hazards model (table S5). Moreover, within these studies only subjects who underwent treatment with various chemotherapeutic agents were retained. Consistently, patients with lower levels of INTS6, INTS7, INTS8, INTS10, INTS11 and INTS12 exhibited poorer OS (**Fig. 6C** and fig. S4). Notably, the most striking effect was observed with the catalytic subunit of Integrator, INTS11. We subsequently performed a similar analysis using multiple Affymetrix gene expression cancer datasets, including two Colorectal Cancer cohorts (GSE33113 and GSE39582), two breast cancer cohorts, one of which is split into 2 GEO submissions (GSE9195, GSE6532.1 and GSE6532.2). Again a significant correlation between low expression levels of Integrator subunits and a poorer OS was observed (table S5). Together these data support a model in which decreased levels/activity of the Integrator complex may affect chemotherapeutic response via modulation of the biogenesis of *NEAT1_2* and PS.

## Discussion

Using an unbiased proteomics screen, we have identified known and previously unknown *NEAT1* RNA binding partners, such as the transcription factor TCF7L2 and a member of the F-box protein family, FBXO11, which were both subsequently validated as *bona fide NEAT1* interactors and novel PS proteins. This study therefore provides a new list of factors that may modulate *NEAT1* and PS biology. Remarkably, a large proportion of the identified *NEAT1* interactors belongs to two functionally related protein complexes, namely the core 3'-end processing and Integrator complexes. The Integrator complex contains two essential Integrator subunits, INTS11 and INTS9, which are homologous of CPSF73 (alias CPSF3) and CPSF100 (alias CPSF1), respectively. Integrator interacts with the C-terminal domain (CTD) of RNA Pol II and processes newly transcribed RNA molecules, mainly non-polyadenylated transcripts and uridylate-rich small-

nuclear RNA transcripts (UsnRNAs). Integrator has also been recently implicated in the modulation of gene expression via regulation of protein-coding gene transcription initiation and premature termination (*7*, *27*), in RNA Pol II pause-release (*31*, *32*), and in the biogenesis of enhancer RNAs (*6*). Here, we provide functional evidence for an unexpected role of Integrator in the regulation of the isoform switching of the lncRNA *NEAT1*. Our data are compatible with a model in which Integrator is recruited to the *NEAT1* transcript and participates in the cleavage and subsequent processing of the polyadenylated *NEAT1_1* isoform (**Fig. 6D**). Although eCLIP data indicate that Integrator is also recruited to the 5'-end of the *NEAT1* transcript, our data highlight a role for Integrator in the processing of the 3'-end of *NEAT1_1* to restrain the expression of the long isoform and, thereby, PS formation. Moreover, our previous observation that *NEAT1_1* is constantly made and degraded by the exosome (*17*) raise the possibility that processing of this 3'-end site by Integrator is a critical step in this degradation process. Interestingly, previous data have already implicated the 3'-end cleavage factor Im (CFIm) complex in the processing of *NEAT1_1*. Whether Integrator and the core 3'-end-processing machinery cooperate to process polyadenylated transcripts such as *NEAT1_1* or work independently on different pools of transcripts remains to be addressed.

The interaction between *NEAT1* and Integrator is not disrupted in cells exposed to stress (i.e. HU-induced replication stress) or in cells in which we artificially increased the transcription rate of *NEAT1* (i.e. upon Nutlin-3a exposure). These data therefore favor a model in which bypassing *NEAT1* cleavage may occur because the pool of Integrator available is not sufficient to process the high amounts of *NEAT1* transcripts being produced in cells exposed to stress (or in which the transcriptional rate of *NEAT1* is artificially elevated). The ratio between the rate of *NEAT1* transcription and overall expression levels of the Integrator complex may therefore determine whether *NEAT1_2* remains expressed and whether PS are being assembled (**Fig. 6D**). This model is further supported by the fact that *NEAT1* is an unusually abundant lncRNA (in fact, the "A" in *NEAT1* refers to "Abundant"), being expressed at levels that rival highly expressed housekeeping genes, such as *GAPDH*.

The observation that the *NEAT1*-Integrator association is not disrupted in stressed cells may have important functional implications. We indeed showed that several components of Integrator colocalize with PS in stressed cells. PS assembly is thought to phase separate its content from the nucleoplasm (*19*, *28*) and thus Integrator recruitment to PS may affect its recruitment and activity at other loci. Paralleling this possibility, a comparable cross-regulation between TDP-43 and *NEAT1*/PS was recently shown to promote pluripotency-differentiation transition (*15*). In addition to repressing the formation of PS by enhancing the maturation of *NEAT1_1*, TDP-43 also regulates alternative 3'-

end processing of transcripts encoding pluripotency factors, such as SOX2. PS sequesters TDP-43, just like Integrator, and thereby reduce its binding to polyadenylated RNAs to promote exit from pluripotency (*15*). In a similar way, sequestration of Integrator to PS may contribute to an overall decrease in the processing of small and/or enhancer RNAs in stressed cells and thereby cause an overall downregulation of gene expression and/or rewiring towards a "stress" transcriptome that help cells cope with (chemotherapy-induced) stress. In support of this hypothesis, our data show that the processing of two well-known Integrator targets *RNU11* and *RNU12* is compromised in cells exposed to HU and Nutlin-3a (**Fig. 5B**). Similarly, the 3'-end processing of the replication-dependent histones, previously shown to be affected by silencing of INTS3 (*7*) was also compromised in these experimental conditions (**Fig. 5C**). On the other end, overexpression of INT11 in condition of stress abolished the increase of *NEAT1_2* and PS, and increased DNA damage, thus phenocopying the effects observed upon *NEAT1_2* KD (**Fig. 5G-H**). This model is compatible with the switch from cell cycle arrest/dormancy to apoptotic cell death we observed in cancer cells exposed to chemotherapy following *NEAT1_2* silencing (and PS disruption) and suggests a role for PS as key modulators of 3'-end RNA processing.

Finally, our data also establish an important mechanistic link between Integrator and PS biology. Given the recently recognized role of PS as a modulator of cancer development and sensitivity to cancer therapy, our work therefore highlights the importance of studying Integrator in a cancer biology context. In keeping, we provide evidence that decreased expression of various components of the Integrator complex, and in particular its catalytic subunit INTS11, correlates with poorer clinical outcome for patients exposed to chemotherapy. These observations may ultimately bear important therapeutic implications. Indeed, agents that may increase either the half-life or the recruitment of Integrator to the *NEAT1* locus, or stimulate INTS11 catalytic activity, would be expected to impair PS formation and thereby increase chemosensitivity.

## Materials and Methods

### Cell culture and cloning

All cell lines were acquired from the LCG ATCC Cell Biology collection and kept in culture at 37°C and 5% $CO_2$ in medium supplemented with 1% penicillin and streptomycin (Invitrogen) and 10% FBS (Invitrogen). All cell lines tested negative for mycoplasma contamination. MCF-7 breast cancer cell line was grown in RPMI 1640-glutamax (Gibco, Invitrogen) supplemented with 10µg/ml of insulin (Sigma, I9278).

INTS11 and GFP inducible knockdown clones (HeLa cells) were established as previously described in (*6*). HeLa rescue cells were established by cloning the same shINTS11 sequence into Tet-pLKO-neo vector (Addgene) and single clones were selected with G418 (500µg/ml). shRNA-resistant N terminal Flag-tagged WT or E203Q mutant INTS11 cDNA (*5*) were cloned into Cumate-pLenti-Cloning-2A-GFP vector (ABM Inc.), and transfected into a shINTS11-Tet-pLKO-neo single clone. Stable cell lines were maintained in puromycin (2µg/ml) and G418 (200µg/ml) containing DMEM medium. Knockdowns were induced by adding of 1µg/ml doxycycline (Dox) into the culture medium daily for 3 days.

WT INTS11 cDNA was cloned into a VP16 plasmid (Addgene) to transiently overexpress (OE) INTS11 in MCF-7 cells to perform rescue experiments.

### Cell transfections

For transient knockdown experiments MCF-7 cells were seeded in 6 well plates (200,000 cells/well) and transfected with Lipofectamine RNAiMax (Thermo Fisher Scientific) according to the manufacturer instructions, using 30nM of siNEAT1 or siNEAT1_2 siPOOLs (siTOOLS Biotech), or 35nM of ON-TARGET plus siCPSF3L (siINTS11 - Dharmacon). Transient transfections with the plasmid of interest were performed in 6 well plates (120,000 cells/well) using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer instructions. We transfected either 10µg of DNA for the pVP16-INTS11 overexpression construct, or 60µg of DNA for the U7-GFP reporter construct (*30*). Cell media was refreshed after 8h from transfection and treatments started 24h post transfection.

### Cell treatments

MCF-7 cells were treated with 5µM of Nutlin-3a (Sellekchem) for 24 hours, or with 1mM hydroxyurea (HU) (Sigma) for 44 hours. For Actinomycin D experiments in **Fig. 3B** and **Fig. 4B**, MCF-7 cells were seeded in 6 well plates (180,000 cells/well) and exposed to 1 hour pulse of 3µM

Actinomycin D (Sigma) 24 hours after seeding. After two washes in PBS cells were treated with either DMSO (vehicle), 5μM Nutlin-3a, or 1mM HU for 24 hours. RNA was extracted with TRIzol lysis reagent (Qiagen), according to the manufacturer instructions, and DNAse treated to measure the transcript levels of the ribosomal RNA *16S*, the lncRNA *NEAT1* (*NEAT1*) and its long form specifically (*NEAT1_2*), and *SRSF1* (used here as positive control) by RT-qPCR.
In the rescue experiments cells were first transfected with either INTS11 overexpressing construct, or with siN1_2 siPooLs, and then continuously treated with DMSO, Nutlin-3a (5μM) or HU (1mM) for 24 hours (rescue with siN1_2), or 72 hours (rescue with INTS11 OE plasmid). For the RNA read-through experiments of **Fig. 5B-C**, MCF-7 cells were either transfected with 35nM of ON-TARGET plus against CPSF3L (siINTS11) or exposed to stress for 108h (5μM Nutlin-3a or 1mM HU).

**Cell fractionation**

Nuclear and cytoplasmic extracts were prepared from 15-cm plates using the Nuclei EZ prep kit (Sigma-Aldrich) according to the manufacturer's instructions. The quality of the nuclear isolation was verified by RT-qPCR, assessing the cytoplasmic RNA encoding the 40S ribosomal protein *S14*, the exclusively nuclear non-coding RNAs *NEAT1* and *MALAT1*.

**RNA Antisense Purification (RAP) and quantitative label-free Mass Spectrometry (MS)**

Briefly, for antisense purification of the protein interactors of *NEAT1*, 100μg of Streptavidin Sepharose High Performance beads (GE Healthcare) were coupled overnight at 4°C to 800pmol of biotinylated RAP probes against the 5' portion of the *NEAT1* transcript (N1_5' - Biosearch Technologies) or RAP probes designed against the melanoma-specific *LINC00698* (Ctrl - Biosearch Technologies). MCF-7 breast cancer cells ($1.5 \times 10^7$ cells per treatment) were washed twice in PBS and UV crosslinked dry at 400mj/cm$^2$ with a CL-1000 Crosslinker (254 nm lamp). After performing cell fractionation as indicated above, nuclei were lysed in pull-down buffer (20mM Tris-HCl pH 8.0, 200mM NaCl, 2.5mM $MgCl_2$, 0.05% Triton X-100 in DEPC water) supplemented with a cocktail of protease inhibitors (Halt Protease and Phosphatase Inhibitor Single-Use Cocktail (100X) – Thermo Fisher), 1mM dithiothreitol and 60U/ml of SUPERase• In™ RNase Inhibitor (Life Technologies). Lysates were incubated with the beads coupled to the RAP probes at 4°C for 3 hours. Beads were rinsed 3 times with pull-down buffer and twice with DEPC-treated water. For mass spectrometry analysis proteins were rinsed in trypsin digestion buffer (20mM Tris-HCl pH 8.0, 2mM $CaCl_2$) and eluted by on-beads digestion with 1μg of trypsin (Promega) over night at 37°C. Peptides were purified with OMIX tips (C18

resin) and dried to be stored till MS analysis (see LC-MS/MS analysis paragraph). For western blot, proteins were directly eluted in 30µl of Laemmli-buffer supplemented with TCEP, boiled for 15 minutes at 95°C and stored at -80°C. For RNA elution samples were first decrosslinked at 56°C in decrosslinking buffer (100mM Tris-HCl pH 7.5, 50mM NaCl, 10mM EDTA, 0.5% SDS) with Proteinase K (Roche) to a final working concentration of 2 mg/ml for 30-40 minutes, and then extracted in TRIzol – clorophorm and precipitated overnight at -80°C in 1/10$^{th}$ (v/v) NaCl and EtOH 100%. The purified RNA was treated with DNAse, measured with a nanodrop, and stored at -80°C.

**LC−MS/MS analysis**

The cleaned peptide mixtures were dried completely and re-suspended in 20µl loading solvent (0.1% TFA in water/acetonitrile, 2/98 (v/v)). 2µl of the peptide mixtures were analyzed by LC−MS/MS on an Ultimate 3000 RSLC nano LC (Thermo Fisher Scientific, Bremen, Germany) in-line connected to a Q Exactive mass spectrometer (Thermo Fisher Scientific). Peptides were separated with a linear gradient at 300nl/min from 98% solvent A (0.1% formic acid in water) to 55% solvent B (0.1% formic acid in water/acetonitrile, 20/80 (v/v)) in 120 min before ultimately reaching 99% solvent B. The mass spectrometer was operated in data-dependent, positive ionization mode, automatically switching between MS and MS/MS acquisition for the 10 most abundant peaks in a given MS spectrum.

**Proteomics data analysis**

Data analysis was performed with MaxQuant (version 1.5.4.1) using the Andromeda search engine with default search settings including a false discovery rate set at 1% on both the peptide and protein level. Spectra were searched against human proteins in the UniProt/Swiss-Prot database (database release version of August 2016 containing 20,210 human protein sequences, www.uniprot.org). The mass tolerance for precursor and fragment ions was set to 20 and 4.5 ppm, respectively, during the main search. Enzyme specificity was set to C-terminal to arginine and lysine, also allowing cleavage at arginine/lysine-proline bonds with a maximum of two missed cleavages. Variable modifications were set to oxidation of methionine (to sulfoxides) and acetylation of protein N-termini. A minimum of one peptide was required for protein identification. We allowed for matching between runs using a 1 minute match time window and a 20 minute alignment time window. Proteins were quantified by the MaxLFQ algorithm integrated in the MaxQuant software. A minimum ratio count of two unique or razor peptides was required for quantification.

Further data analysis was performed with the Perseus software (version 1.5.5.3) loading the protein groups file from MaxQuant. First, proteins only identified by site, reverse database hits and

potential contaminants were removed. The LFQ intensities were Log2 transformed, the replicate samples were grouped and protein groups with less than 3 valid values in at least one group were removed. Missing values were then imputed with values from the lower part of the normal distribution representing the detection limit, leading to a list of 1063 reliably quantified proteins. Moreover, we filtered out proteins identified by less than 3 peptides (n=995). Then, a t-test was performed (FDR=0.05) to compare the RAP N1_5' with the RAP Ctrl samples and generate the volcano plots depicted in **Fig. 1E**, **Fig. 3C**, **Fig. 4C** and fig. S2C and fig. S2E. Of the 995 quantified protein candidates, 698 candidates were enriched by N1_5' RAP probes. Significantly enriched proteins (*p* value<0.05) with a N1_5'/Ctrl fold change >1.6 (arbitrary cut-off) were considered as highly confident *NEAT1* interaction partners (table S1, S3, S4).

PCA analysis: ellipses represent 95% confidence intervals, around each cluster's centroid, calculated using Hotelling's $T^2$ statistics. The axes are the respective first and second principal components (PCs) with the percent variance captured by each PC in the parentheses. The figure was generated in R using the "factoextra" package.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015158.

**RIP**

RNA ImmunoPrecipitation (RIP) was performed on freshly isolated nuclei from MCF-7 cells (2.5 $10^7$ cells/sample) after UV-crosslinking with 0.4 J $cm^{-2}$ of $UV_{254\ nm}$. Nuclei were lysed with polysome buffer (20mM Tris-HCl pH 8.0, 200mM NaCl, 2.5 mM $MgCl_2$, 1% Triton X-100, DEPC water) supplemented with a cocktail of protease inhibitors (Halt Protease and Phosphatase Inhibitor Single-Use Cocktail (100X) - Thermo Fisher), 1mM dithiothreitol and 60U/ml of SUPERase• In™ RNase Inhibitor (Life Technologies), and precleared with protein A beads for 1h at 4°. RIP was performed overnight at 4 °C on a rotating wheel using 5µg of the specific antibody INTS11 (Sigma, A107128) or Normal rabbit IgG (Millipore, 12-370) used as control. On the following day 50µl of protein A Dynabeads (Invitrogen) were coupled to the antibody for 3h at 4°. The beads were rinsed 5 times with polysome buffer and split in two, to either elute proteins or RNA (see RAP protocol for elution steps).

**RT–qPCR**

Total RNA was extracted with TRIzol lysis reagent (Qiagen) according to the manufacturer instructions, DNAse treated and reverse transcribed using the High-Capacity complementary DNA Reverse Transcription Kit (Thermo Fisher Scientific). RNA expression levels were measured by

qPCR on a LightCycler 480 (Roche). Data was analyzed in qbase + 3.0 (Biogazelle) using *HPRT1*, *TBP* and *GAPDH* as reference genes. For the sequences of the RT qPCR primers see table S6. Primers for *HIST* transcripts (**Fig. 5C**) were taken from (*7*).

**RAP (and RIP) analysis**

The RAP (and RIP) efficiency was estimated by RT-qPCR starting from 0.2µg of RNA/sample. The enrichment of the gene of interest for the RAP (RIP) experiment (*NEAT1* and *NEAT1_2* primers) was calculated applying the $\Delta(\Delta Ct)$ method. In brief, the $C_t$ value of the RAP (RIP) elution was subtracted from the $C_t$ value of the input for every gene, thus obtaining the $\Delta C_t$ for each gene in the RAP (RIP) sample. From the RAP (RIP) $\Delta C_t$ was subtracted the $\Delta Ct$ of the RAP control (Ctrl probes, targeting the melanoma-specific *LINC00698*), or of the Normal IgG (RIP), for every gene, thus obtaining the $\Delta(\Delta Ct)$. The equation "fold enrichment $= 2^{-\Delta(\Delta Ct)}$" was used to calculate the fold change for each gene and was plotted as such.

**Immunoblotting**

Cells were scraped on ice in RIPA buffer containing protease and phosphatase inhibitor cocktails (Thermo Fisher). The cell lysates were pushed through a 22-gauge needle with syringe five times and vortexed, incubated on ice for 10 minutes and then centrifuged at 21,000 × *g* for 15 minutes at 4°C. 30µg or 20µg of total protein lysate were loaded on NuPAGE Novex 4–12% Bis-Tris Protein Gels (Invitrogen) and probed with primary antibodies at 4°C overnight (see Antibodies paragraph below).

**eCLIP Assay**

eCLIP was performed in HeLa cells in duplicates as previously described in (*26*). In brief, 2 x $10^7$ cells were crosslinked by UV-C irradiation (254 nm, 400 mJ/cm$^2$) and lysed on ice followed by sonication. Antibodies (INTS11: Abcam ab75276 or Sigma Prestige HPA029025) were incubated with Dynabeads™ M-280 Sheep Anti-Rabbit IgG (Invitrogen, 11204D) for 1 hour. After limiting RNase I (Ambion) digest in presence of DNase, the lysate was subjected to immunoprecipitation at 4˚C for 16 hours. In the following, 2% of the lysate was removed for size-matched input control. IP efficiency and specificity were verified by immunoblot using 20% of the IP material. Co-immunoprecipitated RNA was dephosphorylated, followed by 3'RNA adapter ligation using T4 RNA Ligase (NEB). Input and IgG controls and INTS11-RNA complexes were run on a NuPAGE™ 4-12% Bis-Tris Gel, transferred to nitrocellulose and cut from the membrane between 65-145kDa. Protein-bound RNA was released from the membrane by Urea/ Proteinase K digest,

followed by acid Phenol/ Chloroform/ Isoamyl alcohol RNA extraction and purification using RNA Clean & Concentrator (Zymo Research). After reverse transcription (AffinityScript reverse transcriptase, Agilent), RNA was treated with exonuclease (ExoSAP-IT, Affymetrix) and removed by combined NaOH/ HCl treatment. A 3'Linker was ligated to the cDNA, and the resulting library was PCR amplified using Q5 Polymerase (NEB), purified and size selected for sequencing. Single-end (SE100) sequencing was performed to an average of 40 million reads per sample using Illumina HiSeq 3000 sequencer. Data was processed according to (*26*), including removal of repetitive sequences prior to mapping against the human genome version hg19. eCLIP sequencing coverage of non-coding RNAs (*MALAT1*, *RN7SL*, *TUG1*, *CRNDE*, *RNU11*, *NEAT1_1*, *NEAT1_2 only*) was quantified using bigWigAverageOverBed (*33*). Mean eCLIP signal per transcript was normalized to the expression levels of the lncRNA based on total RNA-seq (with the *NEAT1_2* transcript arbitrarily set to 1). Significant INTS11 binding compared to input was determined using the CLIPper tool with a threshold of Log2>3.7 and a *p* value<$10^{-26}$ (*34*).

**RNA-sequencing**

~$3x10^7$ cells were used for total RNA extraction using TRIzol reagent (Thermo Fisher Scientific, #15596026) according to the manufacturer's instructions. Genomic DNA was removed by Turbo DNAse treatment (Invitrogen, #AM1907). Total RNA-seq libraries were produced using Truseq Stranded Total RNA library prep kit (Illumina, #20020596) with 500ng of DNAse-treated Input RNA. Genome-wide experiments were performed as two independent biological replicates. To avoid a batch effect in library preparation and sequencing flow cell, these replicates were processed together. Raw fastq RNA-seq data were processed with Trimmomatic v0.32 (*35*) and aligned to the human genome (hg19 version) using STAR aligner v2.5.3a (*36*) with default parameters. For visualization on the UCSC Genome Browser, all tracks were CPM (count per million) normalized against the total number of usable reads in that data set using deepTools2 (*37*).

**Small RNA analysis**

~$3x10^8$ cells were used for nuclear fractionation and RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, #15596026) according to the manufacturer's instructions. Genomic DNA was removed by Turbo DNAse treatment (Invitrogen, #AM1907). Small RNA libraries were prepared using the SMARTer smRNA-seq kit (Takara, #635030) with 750ng of Nuclear-enriched total RNA and the experiments were performed as two independent biological replicates. Raw fastq reads were then adapter trimmed (AAAAAAA) as recommend by SMARTer smRNA-seq kit (Takara, #635030) protocol using Cutadapt (v1.14) and reads less than 17bp were discarded. First,

we aligned the reads against human elements in RepBase (v23.08) with STAR (v2.5.3a) (*36*)*4*), repeat-mapping reads were removed, all others were then mapped against the full human genome (hg19 version) and we keep all unique aligned reads. For visualization on the UCSC Genome Browser, all tracks were CPM (count per million) normalized against the total number of usable reads in that data set using deepTools2 (*37*).

**3' end RNA- seq (3' Quant-seq) and data analysis**

Total RNA was extracted and treated with TURBO DNase for 60min at 37°C. We used QuantSeq 3' mRNA-Seq Library Prep Kit REV (Lexogen) to prepare 3' end libraries. 3'Quant-seq was performed on NEXTSeq 500 machine with single-end 75bp sequencing. For the data analysis we followed the Lexogen protocol. Briefly raw fastq data was processed with BBMap (https://sourceforge.net/projects/bbmap/) to remove the adapter contamination, polyA read through and low quality tails, and aligned to the human genome (hg19 version) using STAR aligner v2.5.3a (*36*) with the following parameters (– outFilterType BySJout – outFilterMultimapNmax 20 – alignSJoverhangMin 8 – alignSJDBoverhangMin 1 – outFilterMismatchNmax 999 – outFilterMismatchNoverLmax 0.1 – alignIntronMin 20 – alignIntronMax 1000000 – alignMatesGapMax 1000000 – outSAMattributes NH HI NM MD). For visualization on the UCSC Genome Browser, all tracks were CPM (count per million) normalized against the total number of usable reads in that data set using deepTools2 (*37*).

**RNA–fluorescence *in situ* hybridization**

RNA–fluorescence *in situ* hybridization (FISH) was performed using Stellaris FISH probes (Biosearch Technologies) for human *NEAT1*: SMF-2036-1 for NEAT1_5 and VSMF-2251-5 for NEAT1_m. FISH was performed according to the manufacturer's protocol. In brief: cells were grown on slides (round cover glasses - VWR), fixed in 4% formaldehyde and permeabilized in EtOH 70% over night at 4°C. Cells can be stained within the following 2 weeks maximum. Cells were washed twice in PBS and incubated for 5 minutes in FISH washing buffer (2xSSC and 10% formamide). Hybridization of FISH probes was carried out overnight at 37°C in 2× SSC, 10% formamide and 10% dextran, in a dark humid chamber. After 3 washes with FISH washing buffer, slides were mounted in ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific) and images acquired on a confocal microscope Nikon C2**.** Imaging panels were prepared using Imaris 7.2.3 and ImageJ (plugins such as Interactive 3D surface plot and JACoP were used respectively to produce the 3D plots to show colocalization, and to quantify fluoresce and signal colocalization).

607

608 **Immunofluorescence**

609 Cells were grown on slides, fixed in 4% formaldehyde and permeabilized in EtOH 70% over night
610 at 4°C. Cells were washed twice in PBS and blocked for 1 hour in 3% BSA (Sigma-Aldrich), 10%
611 goat serum (DAKO), 0.2% Triton X-100 (Sigma-Aldrich). Slides were incubated with primary
612 antibodies at room temperature for 1 hour, washed 3 times in PBS and incubated with secondary
613 antibodies, either anti-rabbit or anti-mouse AlexaFluor-488 or AlexaFluor-555 (Life
614 Technologies) at room temperature for 45 minutes. After 3 washes in PBS, slides were mounted in
615 ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific). Images were acquired
616 on a confocal microscope Nikon C2**.** Imaging panels were prepared using Imaris 7.2.3 and ImageJ.
617

618 **Immunofluorescence combined to RNA FISH**

619 Cells were grown on slides and fixed in 4% PFA, permeabilized in EtOH 70% over night at 4°C
620 and stained within the following 2 weeks maximum. The protocol for RNA FISH was performed
621 first by incubation over night at 37°C with FISH probes (NEAT1_5 Quasar 560 and NEAT1_m
622 Quasar 670). The following day cells were incubated 30 minutes in FISH wash buffer at 37°C,
623 washed twice in PBS and fixed again at room temperature for 15 minutes in 2% PFA. After two
624 washes in PBS cells were blocked for 1 hour in IF buffer: 3% BSA (Sigma-Aldrich), 10% goat
625 serum (DAKO), 0.2% Triton X-100 (Sigma-Aldrich) and 60 U/ml of SUPERase• In™ RNase
626 Inhibitor (Life Technologies), used also for further washes and antibody-incubations. Slides were
627 incubated with the primary antibody for 1 hour at room temperature in the dark. After 3 washes,
628 cells were incubated for 45 minutes in secondary antibody anti-mouse or anti-rabbit AlexaFluor-
629 488 (Life Technologies), washed again in PBS prior of mounting the slides with ProLong Gold
630 Antifade Mountant with DAPI (Thermo Fisher Scientific). Images were acquired on a confocal
631 microscope Nikon C2**.** Imaging panels were prepared using Imaris software 7.2.3 and ImageJ (the
632 plugin Interactive 3D surface plot was used to produce the 3D plots to show colocalization; JACoP
633 was used to quantify fluoresce and signal colocalization).
634

635 **Antibodies**

636 Western blotting experiments and/or IF and IF combined to FISH were performed using the
637 following primary antibodies: PSPC1 (Sigma, SAB4200503), PSF (Sigma, P2860), FBXO11
638 (Novus Biologicals, NB100-59826), TCF7L2 (Cell Signaling, 2565), CPSF1 (Santa Cruz, sc-
639 166281), CPSF2 (Santa Cruz, sc-165983), GAPDH (Abcam, ab9485), NONO (Bethyl, A300-
640 587A), TDP-43 (Proteintech, 12892-1 ap), CPSF3L or INTS11 (Sigma, A107128), INTS1

(Millipore), INIP (C9orf80 (E-12), Santa Cruz, SC-137357), INTS3 (Bethyl, A300-427A), INTS6 (Bethyl, A301-658A), INTS8 (Bethyl, A300-269A), p53-DOI (Santa Cruz, sc-126), p21 (Santa Cruz, sc-6246), phospho-γH2AX (Cell Signaling, 2577), Laminin A+C (Abcam, ab108922), vinculin (Sigma, V9131), GFP (Clontech, 632375), H3 (Abcam, ab1791).

**STochastic Optical Reconstruction Microscopy (STORM)**

Cells were seeded in ibidi µ slide 4 wells and, after the indicated treatments, fixed in 4% formaldehyde for 10 minutes at room temperature. Cells were permeabilized in 0.5% Triton X-100 for 15 minutes and incubated overnight at -20°C with the primary antibody (recombinant Anti-nmt55 / p54nrb antibody, (Abcam, ab133574) at a dilution of 1:1500, and with INTS1 antibody (Millipore-sigma, MABS1984) at a dilution of 1:100. Finally, samples were incubated for 30 minutes at room temperature with the secondary antibodies diluted 1:500, JF646 anti-rabbit (Novus biologicals, NB7156JF646) and Alexa568 anti-mouse (Thermo-Fisher, A-11004).

Imaging experiments were carried out with a Nikon eclipse Ti2 microscope equipped with Nikon Instruments (NSTORM). For two color dSTORM imaging, Janelia 646, and Alexa568 secondary antibodies were used with MEA STORM imaging buffer and were imaged continuously with 5000 frames collected per filter range at a frequency of 20 ms. Images were acquired using a 100x, 1.49 NA objective, and imaged onto a Hamamatsu C11440 ORCA-flash 4.0 camera. Storm localization analysis was carried out with Image J, thunderstorm plugin (1.3-2014-11-08). Molecule list files were then exported from Image J to be further analyzed using Coloc-Tesseler. Cluster analysis, specifically Voronoi function, was carried out after manually selecting regions of interest. For quantification, the whole image was compared versus the selected Region of interest (ROI, 1.5µm x 1.5µm) area of paraspeckle (high NONO signal) for all the treatment groups (at least 13 samples per group, we used 16 for HU). A two-tailed t-test was performed using Graphpad prism 5. More detail on the analysis method have been published previously (*38*, *39*).

***In silico* survival analysis**

Gene expression data and the corresponding clinical information from 11 cancers were downloaded from TGCA repository using the *GDCquery* function of the TCGAbiolinks R package (*40*). These data were: breast cancer (BRCA), lung adenocarcinoma (LUAD), colon adenocarcinoma (COAD), glioblastoma (GBM), low grade glioma (LGG), liver hepatocellular carcinoma (LIHC), kidney renal cell carcinoma (KIRC), Adrenocortical carcinoma (ACC), skin cutaneous melanoma (SKCM) and ovarian cancer (OVC). All the expression and clinical data were then merged and analyzed together excluding genes whose expression level was zero (0) in 50% of the samples. The data were

subsequently normalized using the voom normalization (*41*) and partitioned the data into "normal" and "tumor" samples based on the information provided with the clinical data.

Differential expression was calculated based on Z-scores calculated on the basis of the difference between the means of the normal samples and those from tumor samples correcting for sample heterogeneity using the standard deviation across all genes. With this approach, we assigned differential expression to correspond to genes whose Z-scores were greater or less than ± 1.96 respectively in line with classical z-statistical theory. We then used a univariate Kaplan-Meier survival analysis was conducted to test the effect of the levels of expression of each of the genes of interest (*INTS6, INTS7, INTS8, INTS10, INTS11* and *INTS12*) on the overall survival of the patients that underwent chemotherapy treatment. In addition, we performed a multivariate Cox proportional hazards analysis to model the effects of Age, Stage, Race and Gender along with the expression levels of the genes on the overall survival.

DNA microarray data (Affymetrix dataset) were downloaded from Gene Expression Omnibus (GEO) database repository. The data downloaded were GSE33113, GSE39582, GSE9195 and GSE6532 and GSE30161. They were pre-processed using standard tools for microarray data normalization available through the Affy Package in R (*42*). Given the lack of "normal" samples in these data, differential expression was calculated based, again, on the z-score assuming the global mean to represent the expression levels of "normal" samples. Kaplan-Meier survival analysis was conducted in a similar way to that described for TCGA data.

The bar plots shown in **Fig. 6B** represent the Log2 fold changes of INTS10 and INTS11, and the corresponding fold changes for *NEAT1_2* among respective samples. INTS10 and INTS11 "high" represents the samples for which the log fold changes of the INT subunit is greater than 1.96 (panel on the left), while INTS10 and INTS11 "low" represents log fold less than 1.96 (panel on the right). Fold changes were calculated using z-scores (as described above) to represent the classic null hypothesis of no overall fold change in the mean of all samples. A weighted average of all the 5 *NEAT1_2* Affymetrix probes was taken to represent the expression values for *NEAT1_2*.

## Statistical analysis

The significance between means was determined by two-tailed paired Student′s *t*-test, or with a two-way analysis of variation (ANOVA) test. All *p* values are represented as follows: ns (not significant; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. All statistical analyses were performed with GraphPad Prism v7.0a.

## Graphical output

Figure panels have been generated using Adobe Illustrator 22.1; scientific illustrations were created with the online web-based software BioRender (https://biorender.com/) and iStock (https://www.istockphoto.com/).

## References and Notes


1. B. Tian, J. Hu, H. Zhang, C. S. Lutz, A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* (2005), doi:10.1093/nar/gki158.
2. R. Elkon, A. P. Ugalde, R. Agami, Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* (2013), , doi:10.1038/nrg3482.
3. A. J. Gruber, R. Schmidt, A. R. Gruber, G. Martin, S. Ghosh, M. Belmadani, W. Keller, M. Zavolan, A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.* (2016), doi:10.1101/gr.202432.115.
4. A. J. Gruber, M. Zavolan, Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* (2019), doi:10.1038/s41576-019-0145-z.
5. D. Baillat, M. A. Hakimi, A. M. Näär, A. Shilatifard, N. Cooch, R. Shiekhattar, Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*. **123**, 265–276 (2005).
6. F. Lai, A. Gardini, A. Zhang, R. Shiekhattar, Integrator mediates the biogenesis of enhancer RNAs. *Nature*. **525**, 399–403 (2015).
7. J. R. Skaar, A. L. Ferris, X. Wu, A. Saraf, K. K. Khanna, L. Florens, M. P. Washburn, S. H. Hughes, M. Pagano, The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* (2015), doi:10.1038/cr.2015.19.
8. R. Sandberg, J. R. Neilson, A. Sarma, P. A. Sharp, C. B. Burge, Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science (80-. ).* (2008), doi:10.1126/science.1155390.
9. Z. Xia, L. A. Donehower, T. A. Cooper, J. R. Neilson, D. A. Wheeler, E. J. Wagner, W. Li, Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'2-UTR landscape across seven tumour types. *Nat. Commun.* (2014), doi:10.1038/ncomms6274.
10. C. P. Masamha, Z. Xia, J. Yang, T. R. Albrecht, M. Li, A. Bin Shyu, W. Li, E. J. Wagner, CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* (2014), doi:10.1038/nature13261.
11. A. J. Gruber, R. Schmidt, S. Ghosh, G. Martin, A. R. Gruber, E. van Nimwegen, M. Zavolan, Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. *Genome Biol.* (2018), doi:10.1186/s13059-018-1415-3.
12. C. Adriaens, L. Standaert, J. Barra, M. Latil, A. Verfaillie, P. Kalev, B. Boeckx, P. W. G. Wijnhoven, E. Radaelli, W. Vermi, E. Leucci, G. Lapouge, B. Beck, J. Van Den Oord, S. Nakagawa, T. Hirose, A. A. Sablina, D. Lambrechts, S. Aerts, C. Blanpain, J. C. Marine, P53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* **22**, 861–868 (2016).
13. Y. T. F. Sasaki, T. Ideue, M. Sano, T. Mituyama, T. Hirose, MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci.* (2009), doi:10.1073/pnas.0807899106.
14. T. Naganuma, S. Nakagawa, A. Tanigawa, Y. F. Sasaki, N. Goshima, T. Hirose, Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–4034 (2012).
15. M. Modic, M. Grosch, G. Rot, S. Schirge, T. Lepko, T. Yamazaki, F. C. Y. Lee, E. Rusha, D. Shaposhnikov, M. Palo, J. Merl-Pham, D. Cacchiarelli, B. Rogelj, S. M. Hauck, C. von Mering, A. Meissner, H. Lickert, T. Hirose, J. Ule, M. Drukker, Cross-Regulation between TDP-43 and Paraspeckles Promotes Pluripotency-Differentiation Transition. *Mol. Cell* (2019), doi:10.1016/j.molcel.2019.03.041.
16. M. Isobe, H. Toya, M. Mito, T. Chiba, H. Asahara, Forced isoform switching of Neat1 _ 1 to Neat1 _ 2 leads to the hyperformation of paraspeckles but does not affect the development (2019).

17. C. Adriaens, F. Rambow, G. Bervoets, T. Silla, M. Mito, T. Chiba, H. Asahara, T. Hirose, S. Nakagawa, T. H. Jensen, J. C. Marine, The long noncoding RNA NEAT1_1 is seemingly dispensable for normal tissue homeostasis and cancer cell growth. *RNA* (2019), doi:10.1261/rna.071456.119.
18. C. M. Clemson, J. N. Hutchinson, S. A. Sara, A. W. Ensminger, A. H. Fox, A. Chess, J. B. Lawrence, An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell*. **33**, 717–726 (2009).
19. T. Yamazaki, S. Souquere, T. Chujo, S. Kobelke, Y. S. Chong, A. H. Fox, C. S. Bond, S. Nakagawa, G. Pierron, T. Hirose, Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol. Cell*. **70**, 1038-1053.e7 (2018).
20. L. Standaert, C. Adriaens, E. Radaelli, A. Van Keymeulen, C. Blanpain, T. Hirose, S. Nakagawa, J.-C. Marine, The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA*. **20**, 1844–1849 (2014).
21. S. Nakagawa, M. Shimada, K. Yanaka, M. Mito, T. Arai, E. Takahashi, Y. Fujita, T. Fujimori, L. Standaert, J.-C. Marine, T. Hirose, The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development*. **141**, 4618–4627 (2014).
22. S. S. Mello, C. Sinow, N. Raj, P. K. Mazur, K. Bieging-Rolett, D. K. Broz, J. F. C. Imam, H. Vogel, L. D. Wood, J. Sage, T. Hirose, S. Nakagawa, J. Rinn, L. D. Attardi, Neat1 is a p53-inducible lincRNA essential for transformation suppression. *Genes Dev.* **31**, 1095–1108 (2017).
23. S. Li, J. Li, C. Chen, R. Zhang, K. Wang, Pan-cancer analysis of long non-coding RNA NEAT1 in various cancers. *Genes Dis.* **5**, 27–35 (2018).
24. S. Nakagawa, T. Naganuma, G. Shioi, T. Hirose, Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* **193**, 31–39 (2011).
25. E. Leucci, R. Vendramin, M. Spinazzi, P. Laurette, M. Fiers, J. Wouters, E. Radaelli, S. Eyckerman, C. Leonelli, K. Vanderheyden, A. Rogiers, E. Hermans, P. Baatsen, S. Aerts, F. Amant, S. Van Aelst, J. Van Den Oord, B. De Strooper, I. Davidson, D. L. J. Lafontaine, K. Gevaert, J. Vandesompele, P. Mestdagh, J. C. Marine, Melanoma addiction to the long non-coding RNA SAMMSON. *Nature*. **531**, 518–522 (2016).
26. E. L. Van Nostrand, G. a Pratt, A. a Shishkin, C. Gelboin-Burkhart, M. Y. Fang, B. Sundararaman, S. M. Blue, T. B. Nguyen, C. Surka, K. Elkins, R. Stanton, F. Rigo, M. Guttman, G. W. Yeo, Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*. **13**, 1–9 (2016).
27. D. C. Tatomer, N. D. Elrod, D. Liang, M. S. Xiao, J. Z. Jiang, M. Jonathan, K. L. Huang, E. J. Wagner, S. Cherry, J. E. Wilusz, The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* (2019), doi:10.1101/gad.330167.119.
28. T. Hirose, G. Virnicchi, A. Tanigawa, T. Naganuma, R. Li, H. Kimura, T. Yokoi, S. Nakagawa, M. Bénard, A. H. Fox, G. Pierron, NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell*. **25**, 169–83 (2014).
29. Y. Wang, S.-B. Hu, M.-R. Wang, R.-W. Yao, D. Wu, L. Yang, L.-L. Chen, Genome-wide screening of NEAT1 regulators reveals cross-regulation between paraspeckles and mitochondria. *Nat. Cell Biol.* **20**, 1 (2018).
30. T. R. Albrecht, S. P. Shevtsov, Y. Wu, L. G. Mascibroda, N. J. Peart, K. L. Huang, I. A. Sawyer, L. Tong, M. Dundr, E. J. Wagner, Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res.* (2018), doi:10.1093/nar/gky100.
31. A. Gardini, D. Baillat, M. Cesaroni, D. Hu, J. M. Marinis, E. J. Wagner, M. A. Lazar, A.

Shilatifard, R. Shiekhattar, Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell*. **56**, 128–139 (2014).
32. N. D. Elrod, T. Henriques, K.-L. Huang, D. C. Tatomer, J. E. Wilusz, E. J. Wagner, K. Adelman, The Integrator complex terminates promoter-proximal transcription at protein-coding genes. *bioRxiv* (2019), doi:10.1101/725507.
33. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics* (2010), doi:10.1093/bioinformatics/btq351.
34. M. T. Lovci, D. Ghanem, H. Marr, J. Arnold, S. Gee, M. Parra, T. Y. Liang, T. J. Stark, L. T. Gehman, S. Hoon, K. B. Massirer, G. A. Pratt, D. L. Black, J. W. Gray, J. G. Conboy, G. W. Yeo, Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* (2013), doi:10.1038/nsmb.2699.
35. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014), doi:10.1093/bioinformatics/btu170.
36. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013), doi:10.1093/bioinformatics/bts635.
37. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* (2016), doi:10.1093/nar/gkw257.
38. M. Ovesný, P. Křížek, J. Borkovec, Z. Švindrych, G. M. Hagen, ThunderSTORM: A comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* (2014), doi:10.1093/bioinformatics/btu202.
39. F. Levet, G. Julien, R. Galland, C. Butler, A. Beghin, A. Chazeau, P. Hoess, J. Ries, G. Giannone, J. B. Sibarita, A tessellation-based colocalization analysis approach for single-molecule localization microscopy. *Nat. Commun.* (2019), doi:10.1038/s41467-019-10007-4.
40. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2016), doi:10.1093/nar/gkv1507.
41. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* (2014), doi:10.1186/gb-2014-15-2-r29.
42. L. Gautier, L. Cope, B. M. Bolstad, R. A. Irizarry, Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* (2004), doi:10.1093/bioinformatics/btg405.
43. T. Yamazaki, T. Hirose, The building process of the functional paraspeckle with long non-coding RNAs. *Front. Biosci. (Elite Ed).* **7**, 1–41 (2015).

## Acknowledgments

### General

We thank the Proteomics core facility at VIB/UGent for performing label-free mass spectrometry, the VIB Bio Imaging Core for their support & assistance in this work, and the Oncogenomics core facility at Sylvester Comprehensive Cancer Center for performing high-throughput sequencing. We are grateful to Eric J. Wagner for sharing with us the U7-GFP reporter construct. We wish to acknowledge Greet Bervoets and Odessa Van Goethem for technical support.

### Funding

JB received a PhD research fellowships from the International VIB PhD program (4 years) and by the Emmanuel van der Schueren fellowship nr. ZKD5324 (1 year). EL was supported by the King Baudouin Foundation Fund Emile Carpentier—André Vander Stricht—Van Damme 2017 J1810830-207301 and by the Belgian Federation for Cancer. TKK is supported by the Developmental Origins of Chronic Diseases in Children Network (DEVOTION) of Manitoba. THJ was supported by the Novo Nordisk Foundation.

Research reported in this publication was supported by funding from VIB, grants R01 GM078455 and GM105754, and DP1 CA228041 from the National Institute of Health to R.S. and the National Cancer Institute of the National Institutes of Health under Award Number P30CA240139. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author contributions

JB designed and conducted experiments, acquired, analyzed and interpreted the data. GG analyzed STORM images. EB performed the RNA-seq experiments and generated the STORM images. FB performed the 3'mRNA-seq experiments, and total RNA-seq, small

RNA-seq and 3'mRNA-seq analyses. MMT performed eCLIP experiments, NK processed eCLIP data. TK re-analyzed publicly available transcriptomics data and performed PCA analysis. EL, RS and J-CM conceptualized, designed research studies. J–CM and EL wrote the manuscript. All authors read and edited the manuscript.

**Competing interests**

JB and JCM are inventors on a patent related to this work filed by VIB (no. PCT/EP2015/052663, Nov 4th 2019). JCM is scientific founder and scientific adviser at Flamingo Therapeutics. The authors declare no competing interests.

**Data and materials availability**

High-throughput data are deposited at the Gene Expression Omnibus (GEO) under accession number GSE125534. 3'mRNA-seq is deposited at the Gene Expression Omnibus (GEO) under accession number GSE125535.

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

## Figures and Tables

Fig. 1

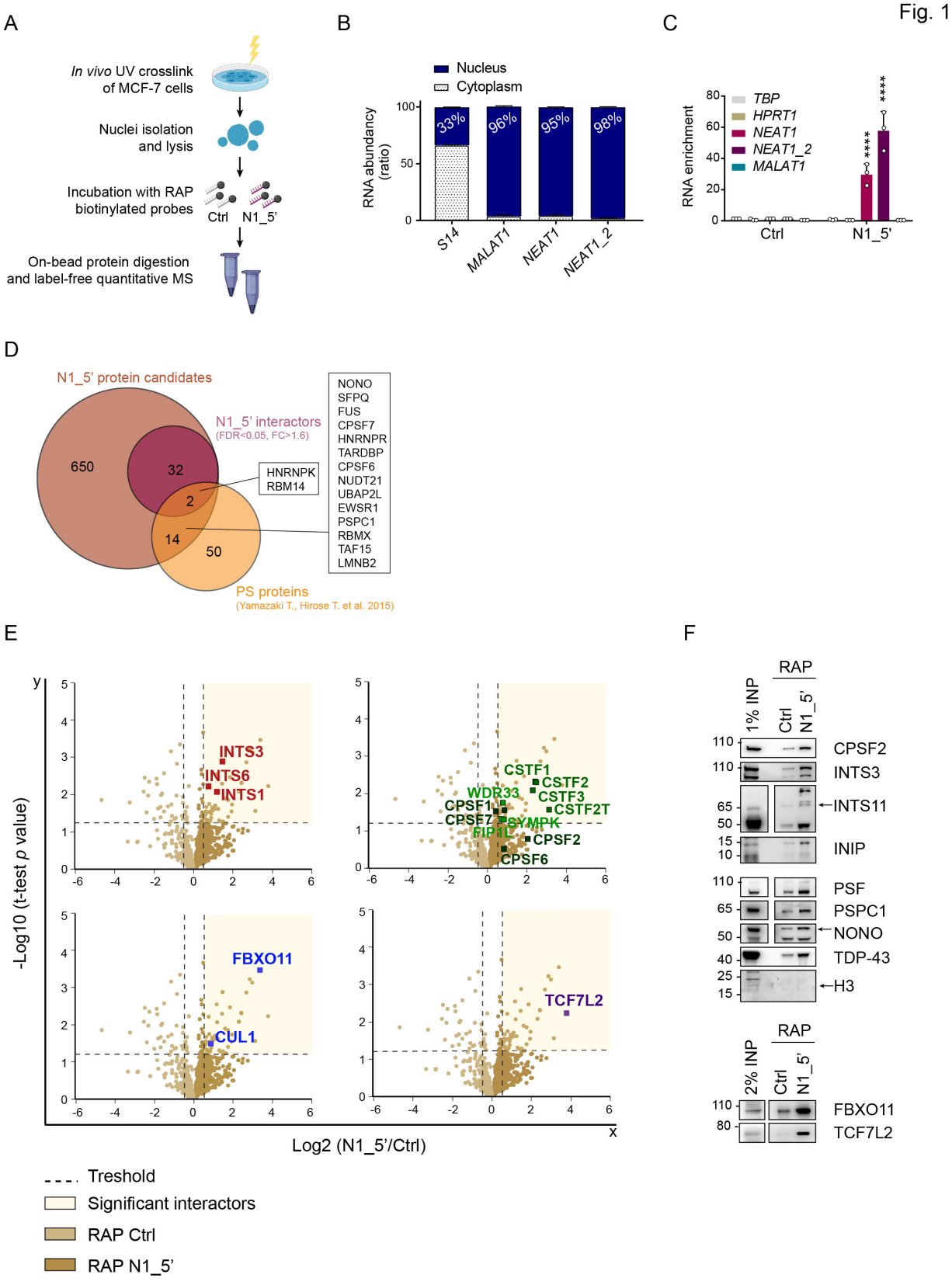


**Fig. 1. *NEAT1* interactome is enriched in 3'-end processing factors**

(**A**) Scheme of the RAP-MS approach to study the protein interactome of the lncRNA *NEAT1*. MCF-7 cells were *in vivo* UV crosslinked and fractionated to isolate nuclei. Nuclear lysates were incubated with RAP biotinylated probes targeting either the melanoma-specific *LINC00698* (Ctrl probes) or the 5' region of *NEAT1* capturing both short and long isoforms (N1_5' probes). Protein interactors were eluted by on-bead digestion, and quantified by label-free mass spectrometry (MS). (**B**) RT-qPCR of the cytoplasmic and nuclear fractions of MCF-7 cells. Localization is shown as relative abundancy of each target in the cytoplasmic and in the nuclear fraction. We measured the cytoplasmic RNA encoding the 40S ribosomal protein *S14*, and the exclusively nuclear lncRNAs *NEAT1* and *MALAT1*. (**C**) RT-qPCR to evaluate the efficiency of the RAP for *NEAT1* transcript. RT-qPCR primers that detect both forms (*NEAT1*) and the long form specifically (*NEAT1_2*) were used. Three abundant transcripts, the lncRNA *MALAT1* and the mRNAs *TBP* and *HPRT1*, are shown as negative controls. Error bars represent mean $\pm$ SD, *p* values were calculated by two-way ANOVA, using 3 biological replicates. ****$p<0.0001$. (**D**) Venn diagram represents the overlap between *NEAT1* protein partners previously identified (*43*) (in orange), and candidates identified in the present study *as bona fide* novel *NEAT1* interactors (brown). In purple are shown the 34 significantly enriched protein interactors of *NEAT1* considered for further analysis, including 2 known paraspeckle proteins. (**E**) The volcano plots indicate the interactors significantly enriched (t-test *p* value<0.05 and FC>1.6) by N1_5' RAP probes. We highlight protein candidates in separate volcano plots grouping them by protein complexes. The *x* axis indicates the difference between N1_5' and Ctrl RAP probes as a ratio N1_5'/Ctrl, in Log2 scale. The *y* axis is the – Log10 (t-test *p* value). (**F**) RAP-western blot experiments for the validation of novel *NEAT1* protein interactors. Western blot was probed with the PS proteins PSF, PSPC1, NONO and TDP-43 (as positive controls) and the putative candidates of *NEAT1*: CPSF2, INTS3, INTS11, INIP, TCF7L2 and FBXO11. H3 detection is used as negative control. Input loaded on the gel was either 1% or 2% of the total nuclear lysate.

Fig. 2

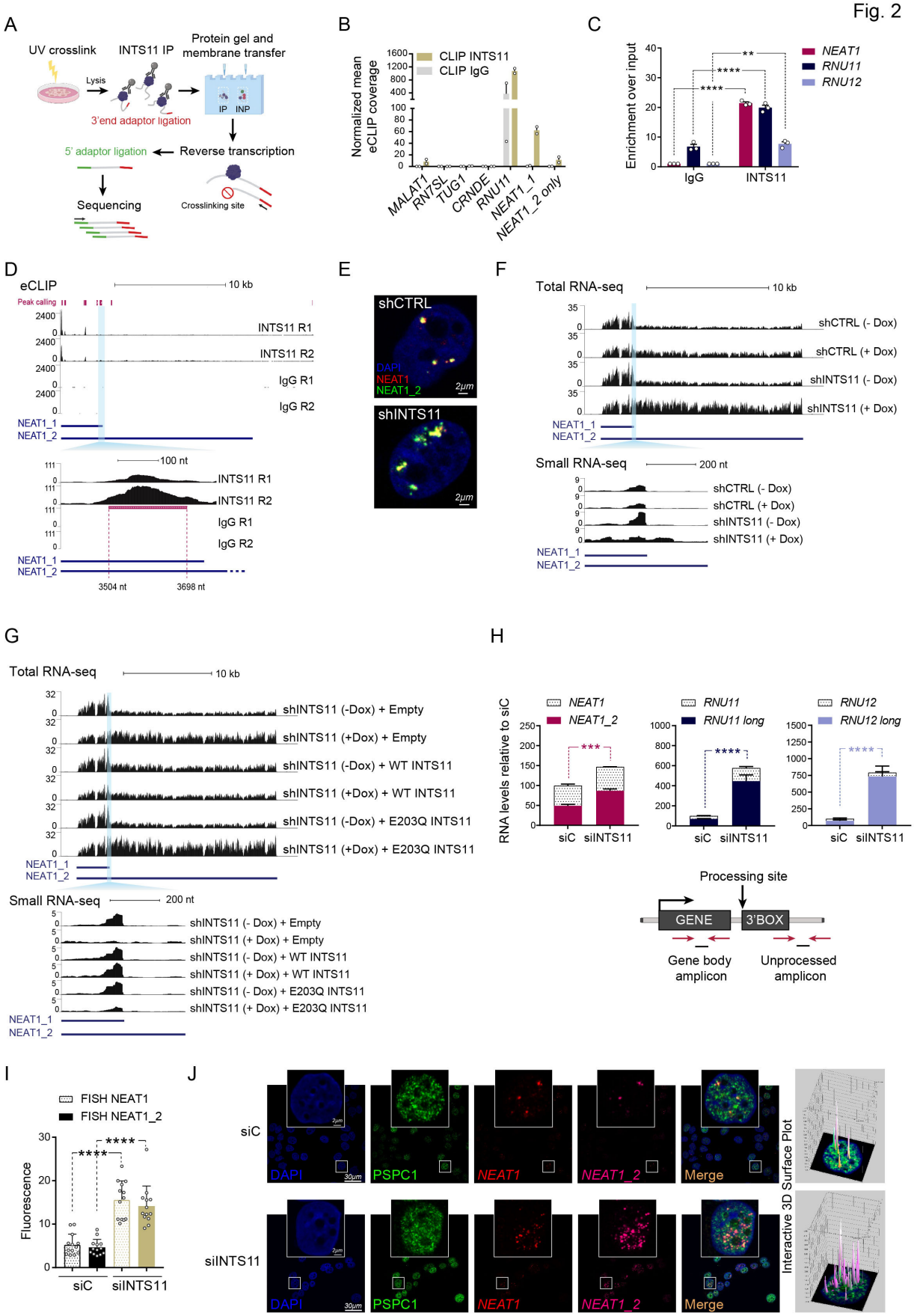
A
UV crosslink
INTS11 IP
Protein gel and membrane transfer
Lysis
IP INP
3'end adaptor ligation
5' adaptor ligation
Reverse transcription
Sequencing
Crosslinking site
B
CLIP INTS11
CLIP IgG
Normalized mean eCLIP coverage
MALAT1
RNTSL
TUG1
CRNDE
RNU11
NEAT1_1
NEAT1_2 only
C
NEAT1
RNU11
RNU12
Enrichment over input
IgG
INTS11
D
eCLIP
10 kb
Peak calling
INTS11 R1
INTS11 R2
IgG R1
IgG R2
NEAT1_1
NEAT1_2
100 nt
3504 nt
3698 nt
E
shCTRL
shINTS11
DAPI
NEAT1
NEAT1_2
2μm
F
Total RNA-seq
10 kb
shCTRL (- Dox)
shCTRL (+ Dox)
shINTS11 (- Dox)
shINTS11 (+ Dox)
Small RNA-seq
200 nt
G
Total RNA-seq
10 kb
shINTS11 (-Dox) + Empty
shINTS11 (+Dox) + Empty
shINTS11 (-Dox) + WT INTS11
shINTS11 (+Dox) + WT INTS11
shINTS11 (-Dox) + E203Q INTS11
shINTS11 (+Dox) + E203Q INTS11
Small RNA-seq
200 nt
shINTS11 (- Dox) + Empty
shINTS11 (+ Dox) + Empty
shINTS11 (- Dox) + WT INTS11
shINTS11 (+ Dox) + WT INTS11
shINTS11 (- Dox) + E203Q INTS11
shINTS11 (+ Dox) + E203Q INTS11
H
RNA levels relative to siC
NEAT1
NEAT1_2
RNU11
RNU11 long
RNU12
RNU12 long
siC
siINTS11
Processing site
GENE
3'BOX
Gene body amplicon
Unprocessed amplicon
I
FISH NEAT1
FISH NEAT1_2
Fluorescence
siC
siINTS11
J
siC
siINTS11
DAPI
PSPC1
NEAT1
NEAT1_2
Merge
Interactive 3D Surface Plot

**Fig. 2. Integrator is a novel *NEAT1* interactor that restrains paraspeckle biogenesis**

(**A**) Schematic representation of the eCLIP approach. After UV crosslink, INTS11-RNA complexes were immunoprecipitated, separated by PAGE and transferred on nitrocellulose membrane. INTS11-RNA complexes (sizes 65-145 kDa) were extracted from the membrane, the RNA recovered and subjected to library preparation followed by sequencing. (**B**) Quantification of the eCLIP signal coverage of INTS11 and size-matched IgG (two replicates each) at the following loci: *NEAT1_1* (5' region) and at the region exclusive of *NEAT1_2* (*NEAT1_2 only*), *RNU11* a known target of INTS11 (here used as positive control), highly expressed ncRNAs such as *RN7SL* (3000-fold higher expressed than *NEAT1_2*), the moderately expressed lncRNA *MALAT1* (35-fold higher expressed than *NEAT1_2*), and low expressed lncRNAs *TUG1* (0.5-fold) and *CRNDE* (0.1-fold). Normalization was performed relative to lncRNA expression levels based on RNA-seq. (**C**) RT-qPCR of the RIP for INTS11 (and IgG as negative control) in freshly isolated nuclei of MCF-7 cells. A significant enrichment in *NEAT1* transcripts as well as *RNU11* and *RNU12* (positive controls) was detected. Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA, using 3 biological replicates. ****$p<0.0001$, **$p<0.01$. (**D**) eCLIP for INTS11 performed in biological duplicates in HeLa cells, with paired size IgG isotype at the *NEAT1* locus was visualized in the UCSC genome browser (GRCh37/hg19). The binding site of INTS11 on *NEAT1* transcript was determined by CLIPper peak calling (using a cut off of an enrichment Log2>3.7 and a *p* value$<10^{-26}$) and is indicated in pink. In the insert panel below we highlight the 3'end peak with the corresponding nt position annotated on the *NEAT1* transcript. (**E**) FISH staining for *NEAT1* (both transcripts - red) and *NEAT1_2* (long transcript – green) in HeLa stable cell lines expressing a doxycycline-inducible knock down of INTS11 (shINTS11) or Control (shCtrl). (**F**) UCSC genome browser (GRCh37/hg19) tracks derived from total RNA-seq (above) and small RNA-seq (below) data obtained in HeLa cells at the *NEAT1* locus upon doxycycline induction of INTS11 knock down (shINTS11) or Control (shCtrl). The blue shade indicates *NEAT1* termination site. (**G**) Here the experiment described in (F), was performed in HeLa cells stably expressing a catalytic dead mutant of INTS11 (E203Q). HeLa cells stably expressing an empty vector (empty), or wild-type INTS11 (WT) are shown as a control. (**H**) RT-qPCR of MCF-7 transiently transfected for 72 hours with siRNA against INTS11. Relative abundancy of *NEAT1* and *NEAT1_2* transcripts, normalized on siC. Two well established Integrator targets (*RNU11*, *RNU12*) with their read-through transcripts are also shown as positive controls. Error bars represent mean ± SD, *p* values were calculated by two-tailed Student's *t*-test, using 3 biological replicates. ****$p<0.0001$, ***$p<0.001$. (**I**) Quantification of the signal intensity of the FISH staining for both forms of *NEAT1*

(FISH NEAT1) and specific for the long form *NEAT1_2* (FISH NEAT1_2) in MCF-7 cells. A total of 25 cells were selected from 3 biological replicates. Error bars represent mean ± SD, *p* values were calculated by paired two-tailed Student's *t*-test. ****$p<0.0001$. (**J**) Immunofluorescence for the PS protein PSPC1 (green) combined to RNA FISH for *NEAT1* isoforms (total *NEAT1*: red, *NEAT1_2*: magenta; nuclei are counterstained with DAPI, blue) in MCF-7 cells upon INTS11 depletion. Scale bars are represented for both the full field image and for the insert panel. The last right panel is a 3D representation of the cell shown in the insert panel, peaks represent signal intensities at the site of colocalization.

Fig. 3

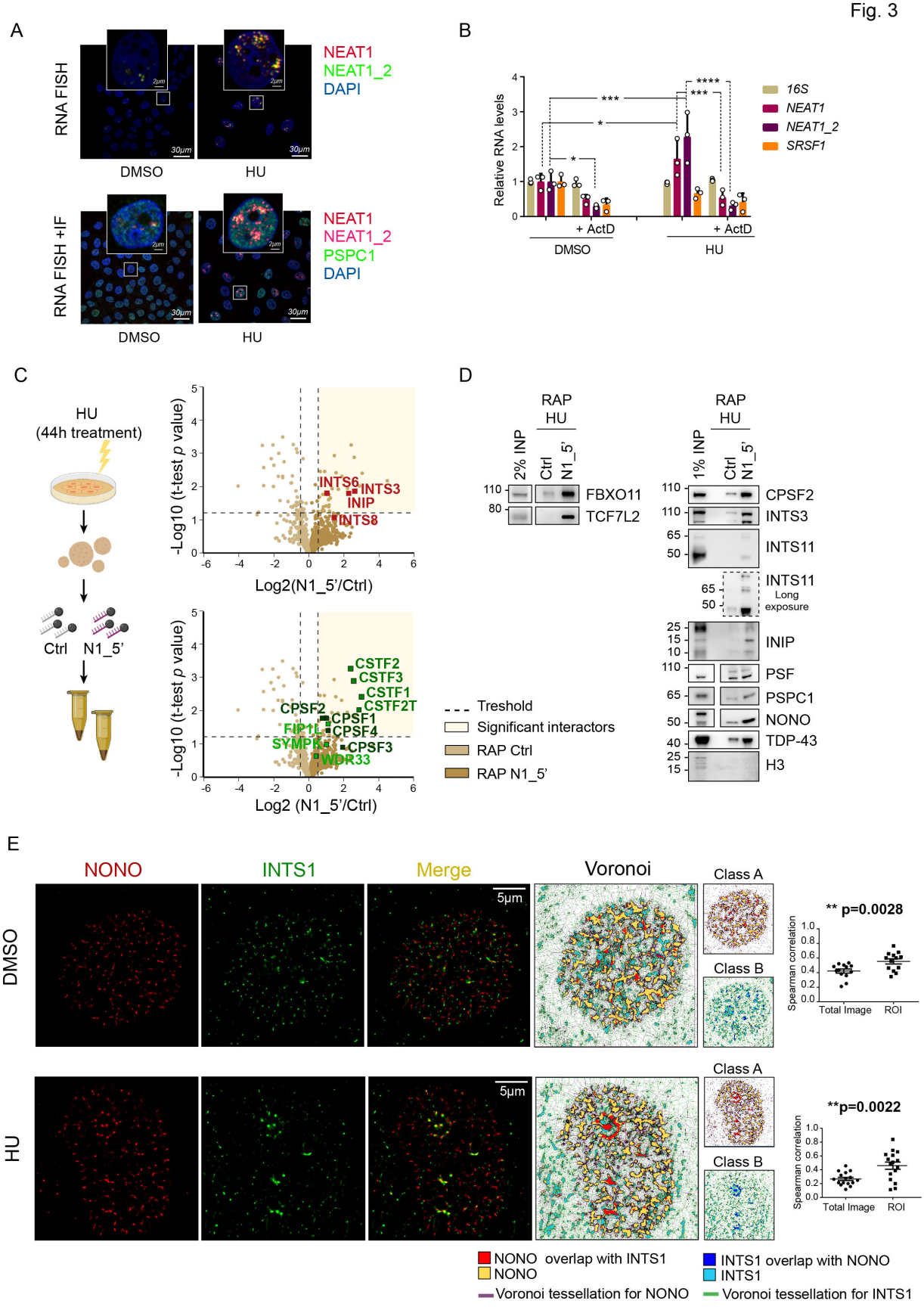
A
RNA FISH
NEAT1
NEAT1_2
DAPI
DMSO
HU
RNA FISH +IF
NEAT1
NEAT1_2
PSPC1
DAPI
DMSO
HU
B
Relative RNA levels
16S
NEAT1
NEAT1_2
SRSF1
+ ActD
+ ActD
DMSO
HU
C
HU
(44h treatment)
Ctrl
N1_5'
-Log10 (t-test p value)
Log2(N1_5'/Ctrl)
INTS6
INTS3
INIP
INTS8
CSTF2
CSTF3
CSTF1
CSTF2T
CPSF2
CPSF1
CPSF4
FIP1L
SYMPK
CPSF3
WDR33
Log2 (N1_5'/Ctrl)
Treshold
Significant interactors
RAP Ctrl
RAP N1_5'
D
RAP
HU
2% INP
Ctrl
N1_5'
FBXO11
TCF7L2
RAP
HU
1% INP
Ctrl
N1_5'
CPSF2
INTS3
INTS11
INTS11
Long exposure
INIP
PSF
PSPC1
NONO
TDP-43
H3
E
NONO
INTS1
Merge
Voronoi
Class A
Class B
5μm
** p=0.0028
Spearman correlation
Total Image
ROI
DMSO
HU
**p=0.0022
NONO overlap with INTS1
NONO
Voronoi tessellation for NONO
INTS1 overlap with NONO
INTS1
Voronoi tessellation for INTS1

**Fig. 3. *NEAT1* interactors do not dissociate upon DNA damage-induced paraspeckle formation**

(**A**) RNA FISH for *NEAT1* isoforms (total *NEAT1*: red, *NEAT1_2*: green, nucleus: blue DAPI – panel above) and RNA-FISH combined to immunofluorescence (total *NEAT1*: red; *NEAT1_2*: magenta; PSPC1: green; nucleus: blue DAPI – panel below) in MCF-7 cells exposed to the DNA damaging agent HU (1mM for 44 hours). Scale bars are indicated. (**B**) RT-qPCR measures the levels of total *NEAT1* transcript (*NEAT1*) and of the long form (*NEAT1_2*), in MCF-7 exposed to HU after a pulse treatment with 3µM of Actinomycin D (Pol II inhibitor). *SRSF1* is used as positive control and the ribosomal RNA *16S* as negative control (transcribed by Pol I). Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA, using 3 biological replicates. **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001. (**C**) Scheme of the RAP-MS experimental approach in MCF-7 cells treated with 1mM of HU for 44 hours. In the volcano plots we separately highlight the significantly enriched protein candidates (t-test *p* value<0.05 and FC>1.6) that are part of different protein complexes. (**D**) RAP – western blot performed in MCF-7 cells exposed to 1mM of HU for 44 hours. RAP capture was performed with Ctrl probes, or N1_5' probes for both forms of *NEAT1*. Western blot was probed with the PS proteins PSF, PSPC1, NONO, TDP-43 (positive controls) and the newly identified *NEAT1* interactors: CPSF2, INTS3, INTS11, INIP, TCF7L2 and FBXO11. H3 is used as negative control. Input loaded is either 1% or 2% of the total nuclear lysate. (**E**) STORM images of the PS protein NONO (red) and INTS1 (green) in MCF-7 cells treated with 1mM of HU for 44 hours. Scale bars are indicated. Voronoi analysis of the colocalization between NONO (class A) and INTS1 (class B). For quantification the whole image was compared to the selected Region of interest (ROI, 1.5µm x 1.5µm) defining the area of the paraspeckle (high NONO signal), 13 images were acquired for DMSO (vehicle) and 16 images for HU.

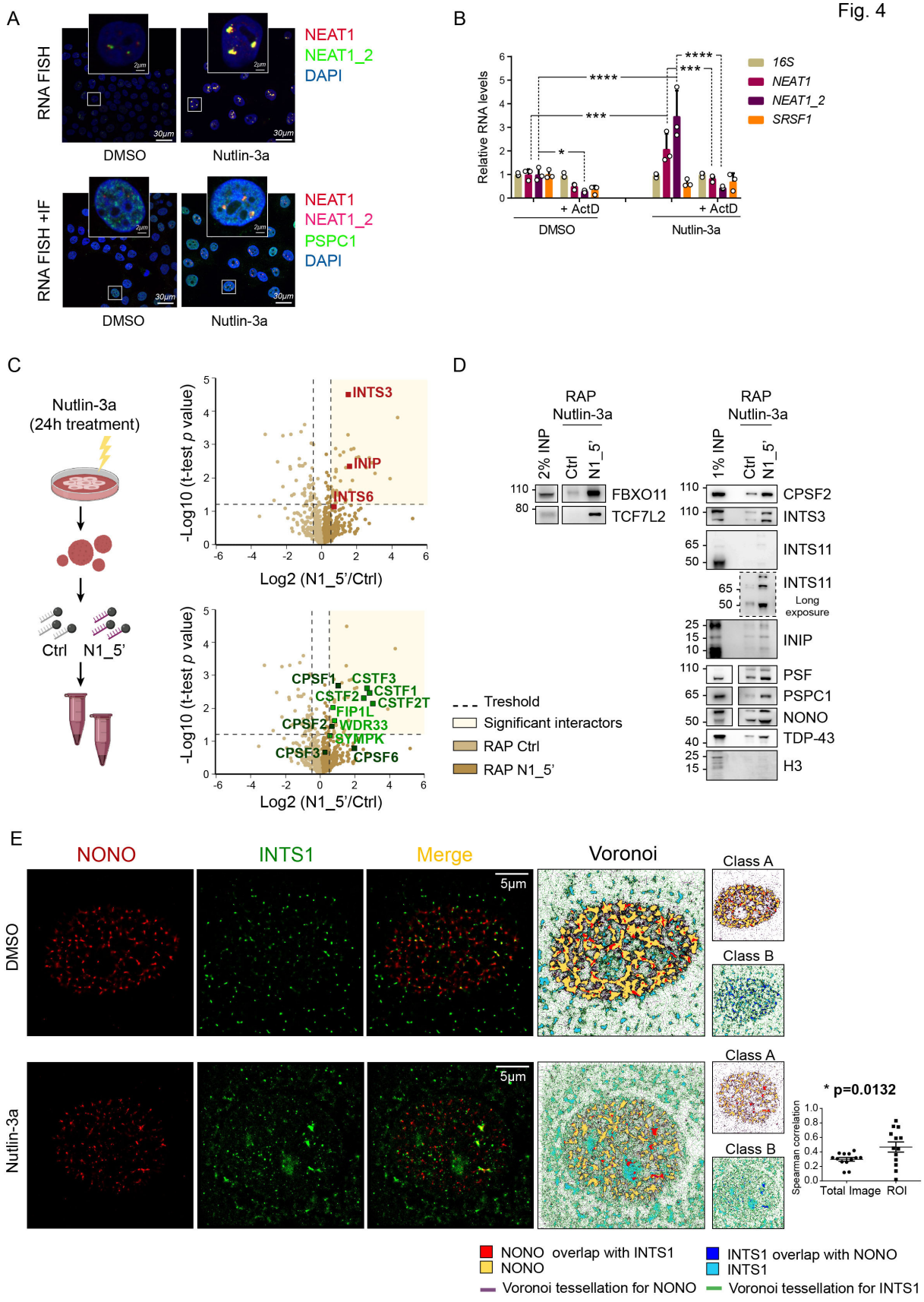
A
RNA FISH
NEAT1
NEAT1_2
DAPI
DMSO
Nutlin-3a
RNA FISH +IF
NEAT1
NEAT1_2
PSPC1
DAPI
DMSO
Nutlin-3a
B
Relative RNA levels
16S
NEAT1
NEAT1_2
SRSF1
+ ActD
+ ActD
DMSO
Nutlin-3a
C
Nutlin-3a
(24h treatment)
Ctrl
N1_5'
-Log10 (t-test p value)
Log2 (N1_5'/Ctrl)
INTS3
INIP
INTS6
CPSF1
CSTF3
CSTF2
CSTF1
CSTF2T
FIP1L
CPSF2
WDR33
SYMPK
CPSF3
CPSF6
D
RAP
Nutlin-3a
2% INP
Ctrl
N1_5'
FBXO11
TCF7L2
1% INP
CPSF2
INTS3
INTS11
INTS11
Long exposure
INIP
PSF
PSPC1
NONO
TDP-43
H3
Treshold
Significant interactors
RAP Ctrl
RAP N1_5'
E
NONO
INTS1
Merge
Voronoi
Class A
Class B
DMSO
Nutlin-3a
5µm
* p=0.0132
Spearman correlation
Total Image
ROI
NONO overlap with INTS1
NONO
INTS1 overlap with NONO
INTS1
Voronoi tessellation for NONO
Voronoi tessellation for INTS1

**Fig. 4. *NEAT1* interactome does not change upon paraspeckle induction driven by oncogenic stress**

(**A**) RNA FISH for *NEAT1* isoforms (total *NEAT1*: red; *NEAT1_2*: green; nucleus: blue DAPI) and RNA-FISH combined to immunofluorescence (total *NEAT1*: red; *NEAT1_2*: magenta; PSPC1: green; nucleus: blue DAPI) in MCF-7 cells exposed to Nutlin-3a (5μM for 24 hours). Scale bars are represented. (**B**) RT-qPCR to measure the levels of total *NEAT1* transcript (*NEAT1*) and of the long form (*NEAT1_2*), in MCF-7 exposed to Nutlin-3a after a pulse treatment with 3μM of Actinomycin D (RNA Pol II inhibitor). *SRSF1* is used as positive control and the ribosomal RNA *16S* as negative control (RNA Pol I transcript). Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA using 3 biological replicates. **p*<0.05, ****p*<0.001, *****p*<0.0001. (**C**) Scheme of the RAP-MS approach in MCF-7 cells treated with 5μM of Nutlin-3a for 24 hours. In volcano plots we separately highlight the significantly enriched protein candidates (t-test *p* value<0.05 and FC>1.6) that are part of different protein complexes. (**D**) RAP – western blot performed in MCF-7 cells exposed to 5μM of Nutlin-3a for 24 hours. RAP capture was performed with Ctrl probes, or N1_5' probes for both forms of *NEAT1*. Western blot was probed with the PS proteins PSF, PSPC1, NONO, TDP-43 (positive controls) and the newly identified *NEAT1* interactors: CPSF2, INTS3, INTS11, INIP, TCF7L2 and FBXO11. H3 is used as negative control. Input loaded is either 1% or 2% of the total nuclear lysate. (**E**) STORM images of the PS protein NONO (red) and INTS1 (green) in MCF-7 cells treated with 5μM of Nutlin-3a for 24 hours. Scale bars are indicated. Voronoi analysis of the colocalization between NONO (class A) and INTS1 (class B). For quantification the whole image was compared versus the selected Region of interest (ROI, 1.5μm x 1.5μm) defining the area of the paraspeckle (high NONO signal). 13 images were acquired for DMSO (vehicle) and 13 images for Nutlin-3a.

Fig. 5

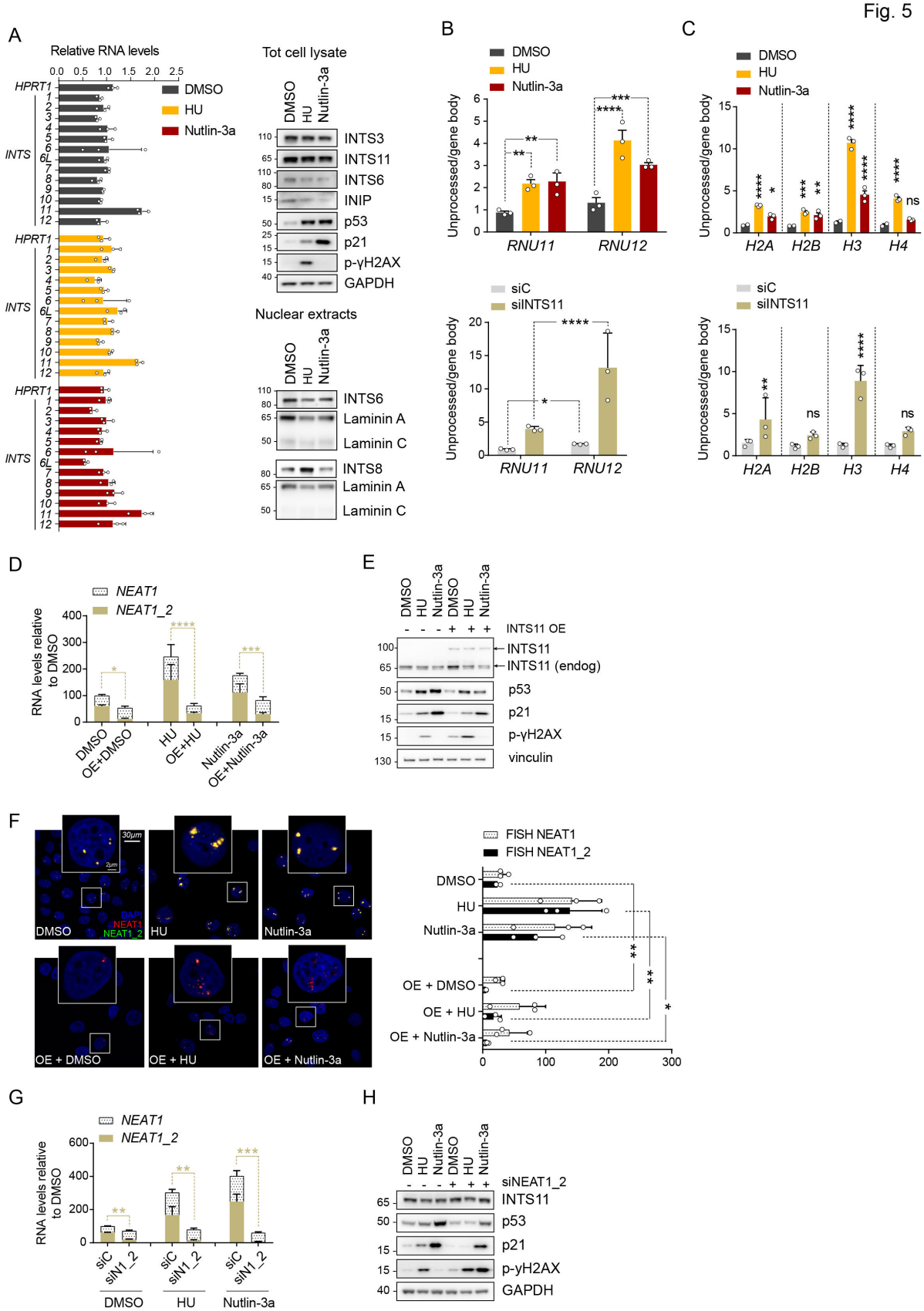
A
Relative RNA levels
DMSO
HU
Nutlin-3a
HPRT1
INTS
Tot cell lysate
INTS3
INTS11
INTS6
INIP
p53
p21
p-γH2AX
GAPDH
Nuclear extracts
INTS6
Laminin A
Laminin C
INTS8
B
Unprocessed/gene body
RNU11
RNU12
siC
siINTS11
C
H2A
H2B
H3
H4
ns
D
NEAT1
NEAT1_2
RNA levels relative to DMSO
OE+DMSO
OE+HU
OE+Nutlin-3a
E
INTS11 OE
INTS11
INTS11 (endog)
vinculin
F
DAPI
FISH NEAT1
FISH NEAT1_2
OE + DMSO
OE + HU
OE + Nutlin-3a
G
siC
siN1_2
H
siNEAT1_2

**Fig. 5. Stress compromises Integrator activity and determines read-through at target genes**

(**A**) (Left panel) RT-qPCR of MCF-7 cells exposed to either DMSO (vehicle), 5μM of Nutlin-3a for 24h, or 1mM of HU for 44h, to measure the expression levels of all the integrator subunits. Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA, using 3 biological replicates (ns for all samples). (Right panel) Western blot analysis of total cell lysate or nuclear extracts of MCF-7 cells, and detection of some Integrator subunits: INTS3, INTS11, INTS6, INIP and INTS8. GAPDH and Laminin A+C are used as loading controls, p53, p21 and phospho-γH2AX (p-γH2AX) confirm the efficacy of the treatment. (**B-C**) RT-qPCR of MCF-7 cells exposed to a prolonged stress (108h - DMSO, HU or Nutlin-3a) or to INTS11 KD (siINTS11) to measure the read-through effect on known targets of INTS11: *RNU11* and *RNU12* (B), and on the Histone transcripts *H2A*, *H2B*, *H3* and *H4* (C). Primers recognize the total gene body, or a region downstream of the 3'end, corresponding to the unprocessed long transcript. After normalization of the expression levels on three housekeeping genes (*TBP*, *HPRT1*, *GAPDH*), ratios between read-through and gene body have been calculated and used to build the graph (unprocessed/gene body). Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA using 3 biological replicates. **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001, *****p*<0.0001, ns (not significant). (**D**) RT-qPCR of MCF-7 cells overexpressing WT INTS11 (or empty vector as control) and constantly exposed to prolonged stress (for 80 hours). Total *NEAT1* and *NEAT1_2* expression levels are normalized on three housekeeping genes (*TBP*, *HPRT1* and *GAPDH*) and represented relative to the control condition (DMSO). Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA using 3 biological replicates. **p*<0.05, ****p*<0.001, *****p*<0.0001. (**E**) Western blot analysis shows the overexpression (OE) of WT INTS11 and the endogenous levels of INTS11. Moreover p53, p21 and phospho-γH2AX (p-γH2AX) indicate the effect of the rescue of INTS11 expression on p53 activity and DNA damage; vinculin is used as loading control. (**F**) FISH staining and relative quantification (biological triplicate) of *NEAT1* (both transcripts – red) and *NEAT1_2* (long transcript – green) in MCF-7 cells with WT INTS11 OE and exposure to prolonged stress. Scale bar is indicated. Quantification of the FISH signal intensity for total *NEAT1* and the long *NEAT1_2* is also shown. Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA using 3 biological replicates. **p*<0.05, ***p*<0.01. (**G**) RT-qPCR to measure *NEAT1* and *NEAT1_2* in MCF-7 cells transiently transfected with siN1_2 and exposed to stress for 24 hours. Expression levels are normalized on three housekeeping genes (*TBP*, *HPRT1* and *GAPDH*) and represented relative to the control condition (siC). Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA using 3 biological replicates. **p*<0.05, ***p*<0.01. (**H**) Western

blot analysis shows p53, p21 and phospho-γH2AX (p-γH2AX) to confirm the efficacy of the treatments, GAPDH is used as loading control.

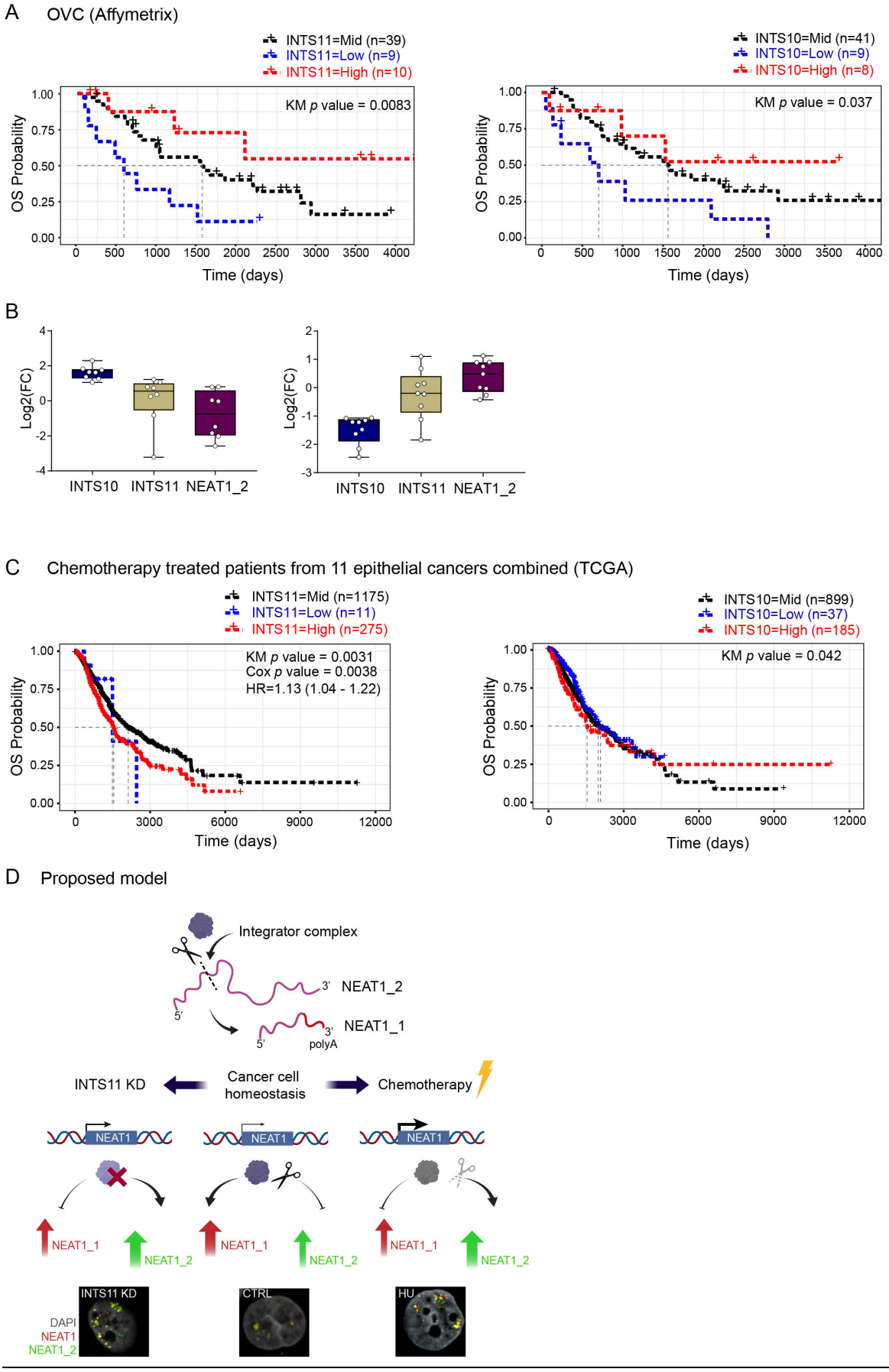
A
OVC (Affymetrix)
INTS11=Mid (n=39)
INTS11=Low (n=9)
INTS11=High (n=10)
KM p value = 0.0083
OS Probability
Time (days)
INTS10=Mid (n=41)
INTS10=Low (n=9)
INTS10=High (n=8)
KM p value = 0.037
B
Log2(FC)
INTS10
INTS11
NEAT1_2
C
Chemotherapy treated patients from 11 epithelial cancers combined (TCGA)
INTS11=Mid (n=1175)
INTS11=Low (n=11)
INTS11=High (n=275)
KM p value = 0.0031
Cox p value = 0.0038
HR=1.13 (1.04 - 1.22)
INTS10=Mid (n=899)
INTS10=Low (n=37)
INTS10=High (n=185)
KM p value = 0.042
D
Proposed model
Integrator complex
3' NEAT1_2
5'
3' NEAT1_1
5'
polyA
INTS11 KD
Cancer cell homeostasis
Chemotherapy
NEAT1
NEAT1_1
NEAT1_2
INTS11 KD
CTRL
HU
DAPI
NEAT1
NEAT1_2

**Fig. 6. Low levels of Integrator components correlate with poorer survival and response to chemotherapy**

(**A**) Kaplan Meier (KM) survival curves of patients from the ovarian cancer Affymetrix cohort (OVC, GSE30161). Patients treated with chemotherapy were selected and stratified based on the expression levels of INTS11 and INTS10.The KM log-rank test *p* values are shown to indicate the difference in overall survival between the patients with significantly low (and high) expression levels of INT subunit. (**B**) Bar plots represent the log fold changes of INTS10 and INTS11, and the corresponding fold changes for *NEAT1_2* among respective samples. The Integrator subunit (10 or 11) is considered "high/low" in samples for which the Log2 fold change is greater/less than ±1.96 respectively. Fold changes were calculated using z-scores to represent the classic null hypothesis of no overall fold change in the mean of all samples (**C**) Kaplan Meier (KM) survival curve of patients treated with chemotherapy from 11 epithelial cancer cohorts combined (TCGA database). Patients were stratified in 3 groups based on the expression levels of INT subunits (blue=low, black=intermediate, red=high). Graphs refer to the subunits INTS11 and INTS10, whose lower expression levels significantly correlate with worse prognosis of chemo-treated patients. We also indicate the *p* value of the Cox Analysis and the Hazard Ratio (HR) calculated for INTS11. (**D**) Proposed model: Integrator restrains paraspeckles formation by promoting the processing of *NEAT1*.

# Supplementary Materials

S1

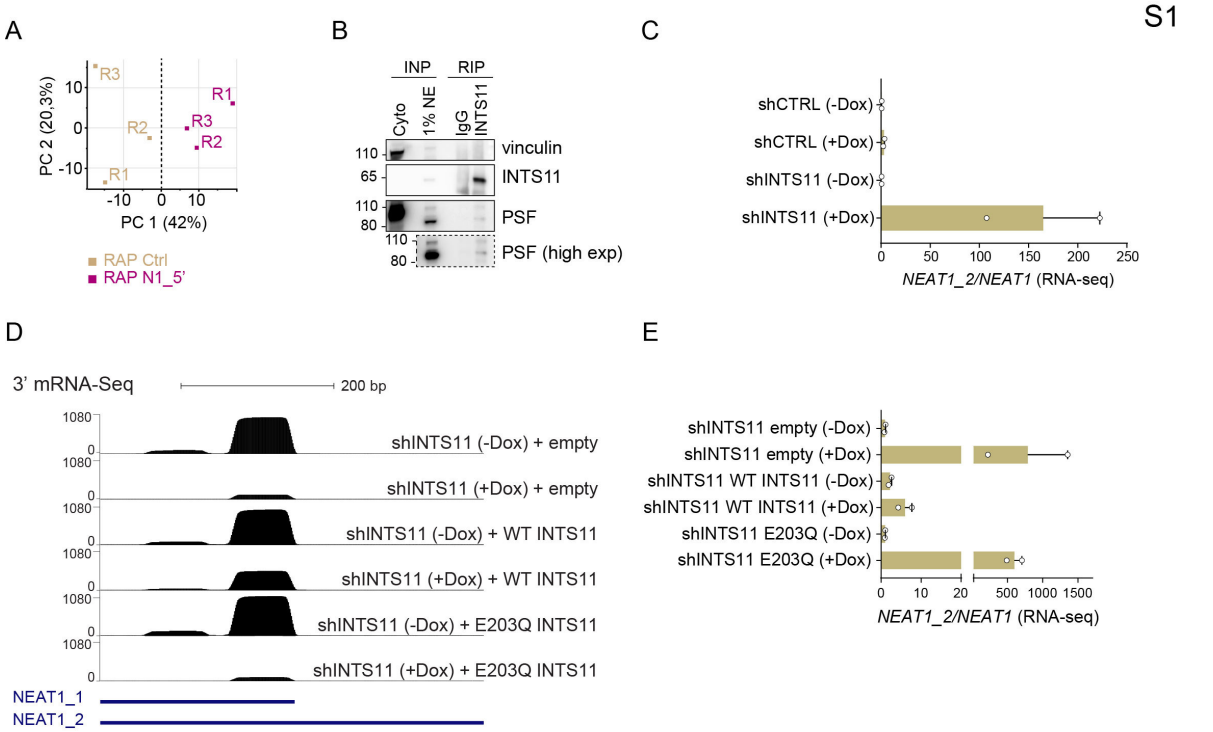

**Fig. S1. Integrator restrains paraspeckle assembly**

(**A**) PCA analysis of MS replicates in untreated MCF-7 cells: ellipses represent 95% confidence intervals, around each cluster's centroid, calculated using Hotelling's $T^2$ statistics. The axes are the respective first and second principal components (PCs) with the percent variance captured by each PC in the parentheses. (**B**) Western blot analysis of the RIP for INTS11 and IgG. 1% of nuclear extract (NE) was loaded on the gel. Membrane probed with INTS11, vinculin (used as negative control) and the known paraspeckle protein PSF. (**C-E**) Quantification of *NEAT1* read-through (*NEAT1_2/NEAT1*) calculated using the FPKM values from 2 biological replicates for each condition. (**D**) 3'mRNA-seq shows specifically the region of *NEAT1_1*. A decrease in 3' processing is detectable upon INTS11 KD, and is rescued by ectopic expression of wild-type INTS11 (WT) and not by expressing the catalytic dead mutant of INTS11 (E203Q). (**F**) UCSC genome browser (GRCh37/hg19) tracks derived from total RNA-seq data obtained in HeLa cells at the *RNU11* locus (used as positive control). (**G**) Total RNA-seq of HeLa cells stably expressing a catalytic dead mutant of INTS11 (E203Q), an empty vector (empty), or wild-type INTS11 (WT) at the locus of *RNU11*. (**H**) Western blot analysis to detect MCF-7 total cell lysate, 72 hours post transfection with siRNA against INTS11 or control (siC). GAPDH is used as loading control; 20µg of protein lysate were loaded on the gel. (Panel below) RT-qPCR to evaluate *INTS11* KD efficiency. (**I**) Signal quantification of FISH for *NEAT1* isoforms combine to IF for PSPC1. For each condition (siC and siINTS11) a total of 25 cells were selected from biological triplicates and the colocalization between channels was quantified (coefficient of colocalization r). In the right panel a representative image of a cell, the dashed circle indicates the nucleus, individual channels are shown for PSPC1, NEAT1 and NEAT1_2.

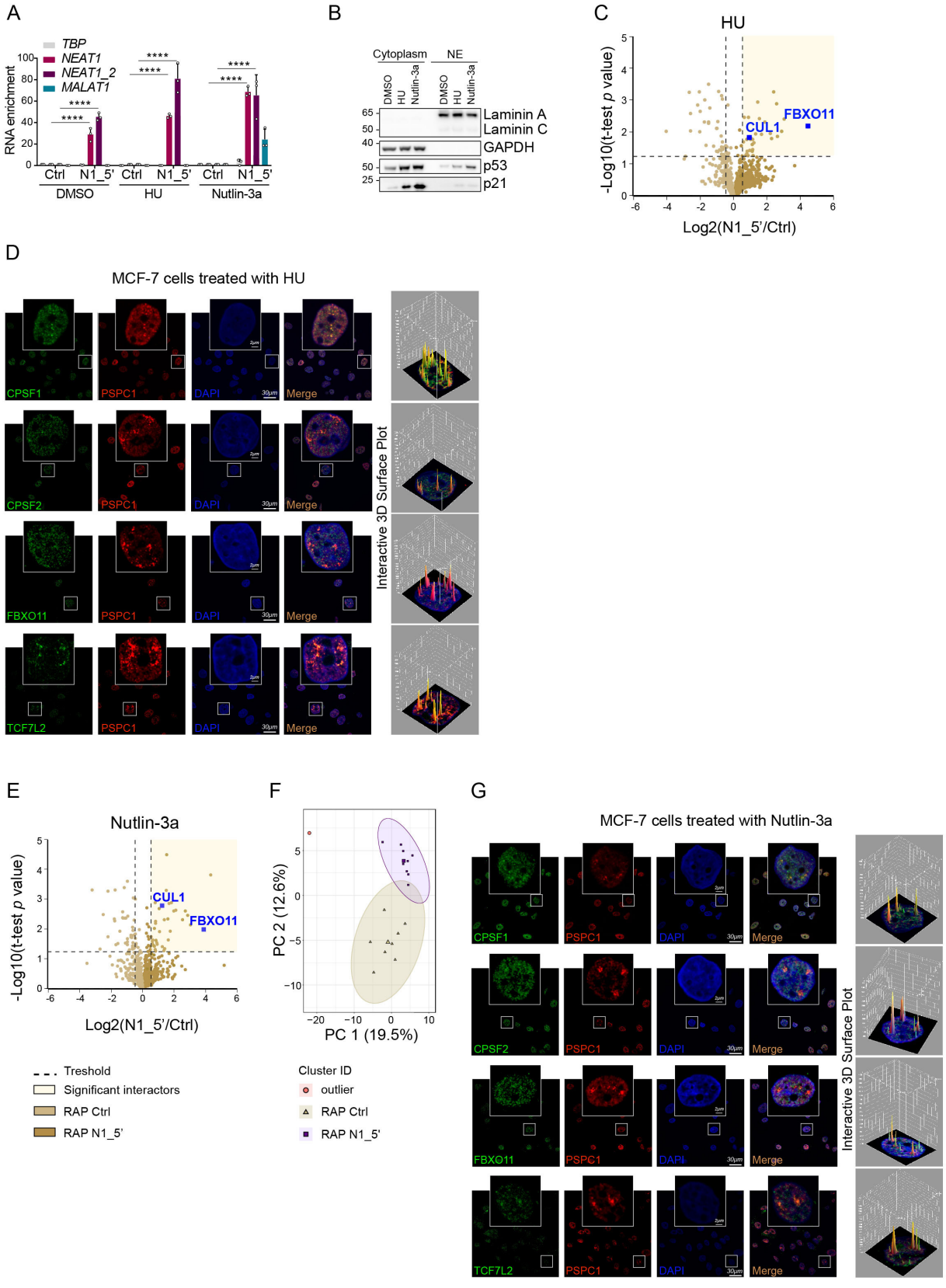
A
TBP
NEAT1
NEAT1_2
MALAT1
RNA enrichment
Ctrl
N1_5'
DMSO
HU
Nutlin-3a
B
Cytoplasm
NE
Laminin A
Laminin C
GAPDH
p53
p21
C
HU
-Log10(t-test p value)
Log2(N1_5'/Ctrl)
CUL1
FBXO11
D
MCF-7 cells treated with HU
CPSF1
CPSF2
FBXO11
TCF7L2
PSPC1
DAPI
Merge
Interactive 3D Surface Plot
E
Nutlin-3a
-Log10(t-test p value)
Log2(N1_5'/Ctrl)
CUL1
FBXO11
Treshold
Significant interactors
RAP Ctrl
RAP N1_5'
F
PC 2 (12.6%)
PC 1 (19.5%)
Cluster ID
outlier
RAP Ctrl
RAP N1_5'
G
MCF-7 cells treated with Nutlin-3a
CPSF1
CPSF2
FBXO11
TCF7L2
PSPC1
DAPI
Merge
Interactive 3D Surface Plot

**Fig. S2. *NEAT1* interactors do not dissociate from PS under stress conditions**

(**A**) RT-qPCR to evaluate the efficiency of purification of *NEAT1* transcripts by N1_5' RAP probes as opposed to Ctrl probes, in MCF-7 cells exposed to 1mM of HU for 44 hours (or to 5µM of Nutlin-3a for 24 hours). Primers recognizing total *NEAT1* (*NEAT1*) and primers specific for the long form of *NEAT1* (*NEAT1_2*) are shown. Two abundant transcripts, the lncRNA *MALAT1* and *TBP* mRNA, are shown as negative controls. Error bars represent mean ± SD, *p* values were calculated by two-way ANOVA, using 3 biological replicates. ****$p<0.0001$. (**B**) Western blot analysis of cytoplasmic and nuclear extract (NE) of MCF-7 cells exposed to the abovementioned stresses. Laminin A+C and GAPDH are used as loading controls and show the purity of the fractionation, p53 and p21 confirm the efficacy of the treatments. (**C**) Volcano plot showing significant enrichment (t-test $p$ value<0.05 and FC>1.6) of FBXO11 and CUL1 by N1_5' RAP probes upon 1mM of HU treatment. (**D**) Immunofluorescence staining of four newly identified candidates (CPSF1, CPSF2, FBXO11 and TCF7L2 - green) and colocalization with paraspeckle protein PSPC1 (red) in MCF-7 cells exposed to HU. The nucleus is counterstained with DAPI (blue). The interactive 3D surface plot indicates as peaks the areas of signal colocalization. Scale bars are represented. (**E**) Volcano plot showing significant (t-test $p$ value<0.05 and FC>1.6) enrichment of FBXO11 and CUL1 by N1_5' RAP probes upon 5µM of Nutlin-3a treatment. (**F**) PCA of all RAP samples to show the consistency of the approach (**G**) Immunofluorescence staining of four newly identified candidates (CPSF1, CPSF2, FBXO11 and TCF7L2 - green) and colocalization with paraspeckle protein PSPC1 (red) in MCF-7 cells exposed to Nutlin-3a. The nucleus is counterstained with DAPI (blue). The interactive 3D surface plot indicates as peaks the areas of signal colocalization.

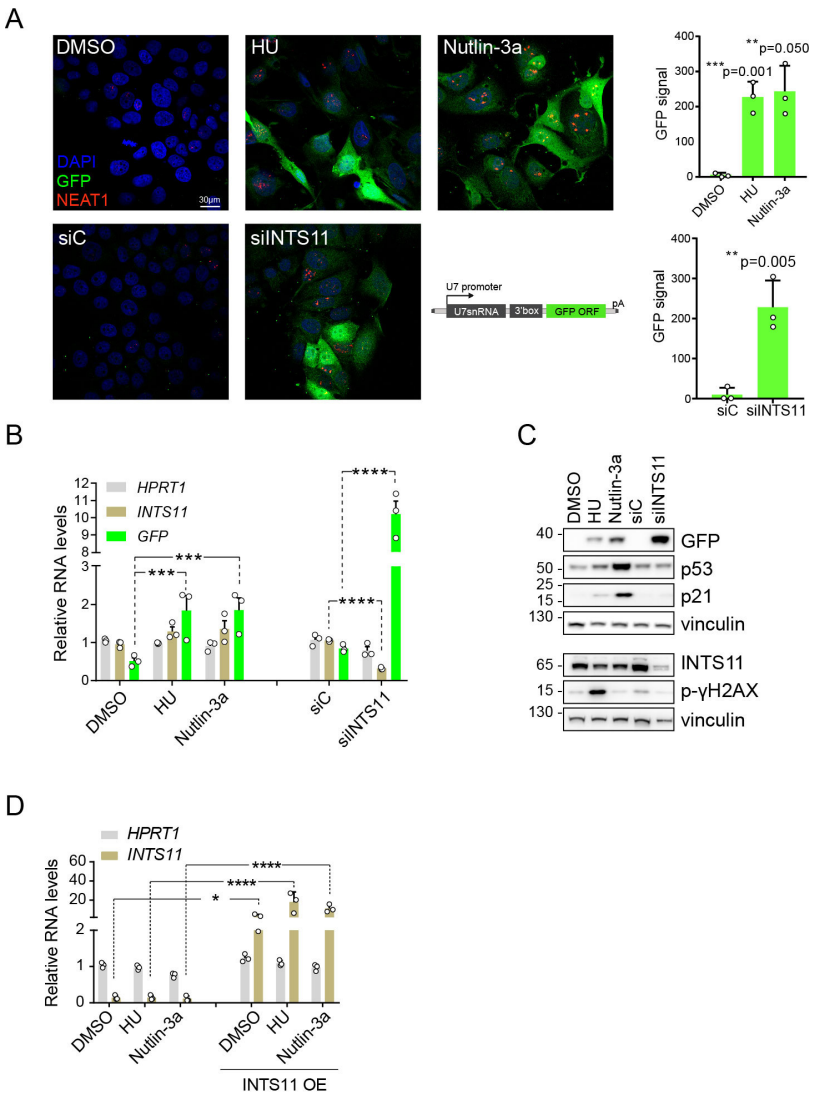


**Fig. S3. Stress causes read-through at INTS11 target genes**

(**A**) FISH staining for *NEAT1* transcripts (red) in MCF-7 cells transiently transfected with the reporter construct U7-GFP to detect read-through (scheme of the construct on the right). 24h post transfection cells were exposed to prolonged treatment (108 hours) with DMSO, Nutlin-3a 5µM, HU 1mM, or to INTS11 KD. GFP intensity was visualized at the microscope and quantified. (**B**) RT-qPCR where INTS11 KD efficiency and GFP reporter expression is measured. (**C**) Western blot analysis to detect GFP and INTS11 protein levels, as well as p53, p21 and phospho-γH2AX (p-γH2AX) to confirm the efficacy of the treatments; vinculin is used as loading control. (**D**) RT-qPCR of MCF-7 cells OE WT INTS11 and exposed to 80 hours of stress. Error bars represent mean ± SD, *p* values were calculated by paired two-tailed Student's *t*-test. * $p<0.05$; **** $p<0.0001$.

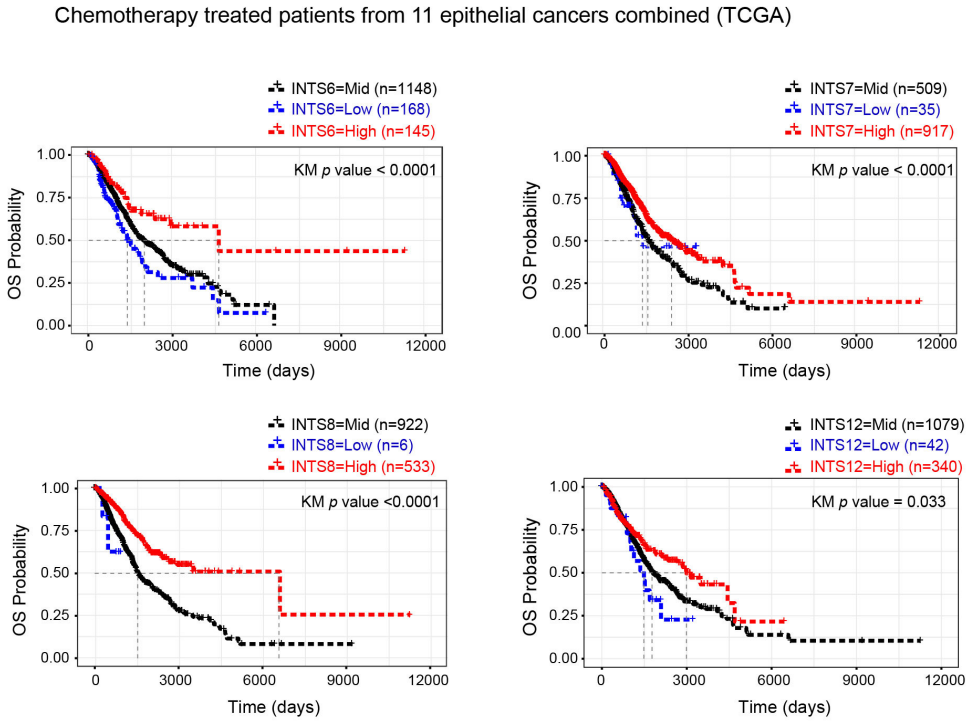


**Fig. S4. Integrator levels correlate with survival and response to chemotherapy**

Kaplan Meier (KM) survival curve of patients treated with chemotherapy from 11 epithelial cancer cohorts combined (TCGA database). Patients are stratified in 3 groups based on the expression levels of INT subunits (blue=low, black=intermediate, red=high). Graphs refer to the subunits INTS6, INTS7, INTS8, INTS12 whose lower expression levels significantly correlate with worse prognosis of chemo-treated patients.

**Additional files** that cannot be embedded into this Word files are the following data files (excel tables)

**Table S1.** List of the 34 significant candidates (t-test *p* value<0.05 and FC>1.6) quantified by RAP-MS in unstimulated MCF-7 cells.

**Table S2.** Gene Ontology (GO) Analysis for "Biological Processes" using the Search Tool for Recurring Instances of Neighboring Genes (STRING; https://string-db.org). Performed using the 34 significantly enriched *NEAT1* protein interactors (t-test *p* value<0.05 and FC>1.6, see Supplementary Table 1) quantified by RAP-MS approach, this analysis indicates that the majority of *NEAT1* interactors are mainly involved in key aspects of RNA biogenesis and processing.

**Table S3.** List of the protein candidates quantified by RAP-MS in MCF-7 cells treated with 1mM HU for 44 hours.

**Table S4.** List of the candidates quantified by RAP-MS in MCF-7 cells treated with 5µM of Nutlin-3a for 24 hours.

**Table S5.** *In silico* analysis.

- "TCGA Cox Analysis" sheet: multivariate analysis using the Cox proportional hazards model on 11 combined TCGA epithelial cancer types. In addition to gene expression levels, we adjusted for the effect of other risk factors (covariates) such as age, race, stage and gender. Highlighted in bold the subunits with both a *p* value<0.05 and HR>1 (increase in hazard).
- "Affymetrix" sheet: Kaplan Meier (KM) survival analysis performed individually on the following Affymetrix cohorts of patients treated with chemotherapy: 2 colorectal cancers (CRC), 3 breast cancers (BRCA) and 1 ovarian cancer (OVC). Highlighted in bold the INT subunits with KM p value<0.05.

**Table S6.** Sequence of the RT-qPCR primers.