Aus dem Biomedizinischen Centrum (BMC)

der Ludwig-Maximilians-Universität München

Lehrstuhl Anatomie III – Zellbiologie

Vorstand: Prof. Dr. Michael Kiebler

The function of paraspeckle components in pluripotency maintenance and differentiation

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

Vorgelegt von

Markus Grosch

aus

Rosenheim

2020

Gedruckt mit Genehmigung der Medizinischen Fakultät

Der Ludwig-Maximilians Universität München

Betreuer: Prof. Dr. Michael Kiebler

Zweitgutachter: Prof. Dr. Olivier Gires

Dekan: Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 15.5.2020

Table of Contents

Table of Contents

Table of Contents

Abstract

The differentiation of pluripotent stem cells (PSCs) to progenitors of the three germ layers mimics one of the earliest events in mammalian development and is regulated by an intricate network of transcription factors, RNA binding proteins (RBPs) and chromatin-remodeling complexes. Moreover, a handful of long non-coding RNAs (lncRNAs) were implicated in this process, however the vast majority of lncRNAs have not been analyzed, yet. Here, I demonstrated that nuclear lncRNAs, including *NEAT1*, which scaffolds membraneless condensates named paraspeckles, generally exhibited dynamic regulation during multi-lineage differentiation of human PSCs. By analyzing paraspeckle formation in 24 human cell types, I revealed general traits such as nucleus size and differentiation characteristics that can explain the variability in paraspeckle numbers between cells within and across different cell populations. On a molecular level, paraspeckle formation is regulated by the RBP TDP-43 via post-transcriptional processing of *NEAT1*. Furthermore, by treatment with DNA-binding chemotherapeutic reagents, which I showed for the first time to dissolve chromatin-bound lncRNA condensates, I determined that assembly of paraspeckles across the germ layers relies on DNA accessibility.

To interrogate the connection of paraspeckles and human embryonic development, I generated genetically-modified PSCs that exhibited altered expression of either one or both *NEAT1* isoforms. The differentiation of these lines revealed that *NEAT1_2*, but not *NEAT1_1*, is fine-tuning the early differentiation process by maintaining the expression of pluripotency and differentiation genes, amongst others of the transcription factors *NANOG* and *PAX6*, which respectively sustain pluripotency or drive neural differentiation.

Finally, I dissected the function of paraspeckle core proteins SFPQ, NONO and PSPC1 in human PSCs. Whereas the knock-out of *NONO* and *PSPC1* induced stem cell differentiation, the depletion of SFPQ is lethal for PSCs. Transcriptome analysis revealed that NONO regulates the expression of cholesterol-producing enzymes, whereas PSPC1 is mainly involved in adipogenesis. Moreover, I identified SFPQ as part of the polyadenylation complex that regulates the expression of genes involved in cell cycle and homeostasis.

Collectively, this study provides a comprehensive analysis of paraspeckle components during mammalian stem cell differentiation. Paraspeckles are paradigmatic for phase-separated, chromatin-embedded condensates and principles of their formation, dynamics and perturbations by small molecules, which have been demonstrated here, might be exploited in medicine as many diseases are accompanied by accumulation of lncRNA condensates whose functions have not been overtly addressed, yet.

Zusammenfassung

Die Differenzierung von pluripotenten Stammzellen (PSCs) zu Vorläufern der drei Keimblätter bildet eines der frühesten Ereignisse in der Säugetierentwicklung nach und wird durch ein kompliziertes Netzwerk an Transkriptionsfaktoren, RNA Bindeproteinen (RBPs) und Chromatinkomplexen reguliert. Außerdem wurden wenige lange, nicht-kodierende RNAs (lncRNAs) gefunden, welche diesen Prozess beeinflussen, allerdings sind die meisten lncRNAs noch nicht diesbezüglich analysiert worden. Ich demonstrierte hier, dass lncRNAs aus dem Zellkern, einschließlich *NEAT1*, welches das Gerüst für membranlose Kondensate namens Paraspeckles ist, dynamisch während der Zelldifferenzierung von humanen PSCs reguliert ist. Durch die Analyse von Paraspeckles in 24 humane Zellarten habe ich allgemeine Merkmale wie Kerngröße und Differenzierungstatus entdeckt, welche die Variabilität bei der Anzahl der Paraspeckles zwischen verschiedenen Zelltypen erklären können. Auf molekularer Ebene werden Paraspeckles durch das RBP TDP-43 via post-transkriptioneller Prozessierung von *NEAT1* reguliert. Desweiteren habe ich zum ersten Mal gezeigt, dass DNA-bindende Moleküle, welche zur Chemotherapie eingesetzt werden, Chromatin-gebundene lncRNA Kondensate auslösen können, woraus ich schloss, dass der Aufbau von Paraspeckles von der DNA Zugänglichkeit abhängt.

Um die Beteiligung von Paraspeckles an der humanen embryonalen Entwicklung herauszufinden, habe ich genetisch-modifizierte PSCs hergestellt, welche veränderte Expression von einer, oder beiden *NEAT1* Isoformen aufweisen. Die Differenzierung dieser Zelllinien zeigte, dass *NEAT1_2*, aber nicht *NEAT1_1*, den frühen Differenzierungsprozess feintunen, in dem es die Expression von Pluripotenz- und Differenzierungsgenen aufrechterhält darunter die der Transkriptionsfaktoren *NANOG* und *PAX6*, welche entweder die Pluripotenz erhalten oder neurale Differenzierung vorantreiben.

Abschließend habe ich noch die Funktion der Paraspecklekernproteine SFPQ, NONO und PSPC1 in humanen PSCs untersucht. Während der Knockout von *NONO* und *PSPC1* die Differenzierung von Stammzellen induziert, ist die Ausschaltung von SFPQ tödlich für PSCs. Die Transkriptomanalyse zeigte, dass NONO die Expression von Cholesterin-produzierenden Enzymen reguliert, während PSPC1 hauptsächlich die Adipogenese beeinflusst. Zusätzlich identifizierte ich SFPQ als Teil des Polyadenylierungskomplexes, welcher die Expression von Zellzyklus- und Homöostasegenen reguliert.

Zusammengefasst stellt diese Studie eine vollständige Analyse von Paraspecklekomponenten während der Differenzierung von Säugetierstammzellen dar. Paraspeckles stehen paradigmatisch für Phasen-getrennte, Chromatin-gebundene Kondensate und die hier demonstrierten Prinzipien ihrer Bildung, Dynamik und Veränderungen durch kleine Moleküle könnten für medizinische Zwecke genutzt

werden, da viele Krankheiten die Akkumulation von lncRNA Kondensaten aufweisen, deren Funktionen

bis jetzt noch nicht genau adressiert wurden.

1. Introduction

1.1 Exordium

The central question of developmental biology is as follows: how can a multicellular organism with trillions of highly specialized cells arise from a single cell that is created after the fusion of male and female gametes? The importance of addressing this question is highlighted by the fact that 6% of infants born worldwide exhibit serious birth defects caused by genetic or environmental abnormalities during gestation [1]. The desire to control and correct potential threats for the embryo even beyond ethical boundaries was recently demonstrated by genetic germline manipulation using the CRISPR/Cas9 system to mutate HIV-targeted receptors in the human embryo [2]. Moreover, many parents need to turn to assisted reproductive technology, such as *in vitro* fertilization (IVF), to conceive a child. Even though seminal progress was made during the past decades, the success rate of IVF is still relatively low and a better understanding of human embryonic development is required to increase chances of a successful pregnancy [3].

Developmental studies in humans have been historically difficult due to limited access to material, however, a milestone was reached with the successful isolation and cultivation of human embryonic stem cells (hESCs) [4] that recapitulate the early pluripotent state during development when the cells are still capable to differentiate to all cell types of the human embryo. In the following sections, I will introduce principles of human development with focus on pluripotent stem cells (PSCs) that are widely used, also in this study, as paradigm for human early embryogenesis.

1.2 The early stages of human embryonic development

Embryonic development begins with the fusion of sperm cell and oocyte, followed by reprogramming of both gametes into the totipotent zygote that can give rise to all embryonic cell types and extraembryonic tissue. While the transcriptional program controlling this process relies initially only on maternally provided mRNAs, this changes during maternal-to-zygotic transition, a process that is conserved in all animal clades, during which the zygotic genome is activated [5]. After a series of cellular divisions, the blastocyst is formed 5 days post fertilization (**Fig. 1**), a process with a success rate of only 50% [6]. The blastocyst contains two types of cells, one surrounding cell layer, the trophectoderm, which will form the placenta, and the inner cell mass (ICM) that contains the founder cells for the embryo, namely pluripotent stem cells.

The next milestone during embryonic development is the implantation of the blastocyst into the uterine wall, which is an intricate process that relies on successful execution of the following steps: a) hormone-controlled formation of a receptive uterus, b) escape of the blastocyst from their outer shell (zona

pellucida), c) apposition and adhesion of the blastocyst to the uterine wall and finally, d) the invasion of fetal trophoblast cells with the reconstruction of maternal spiral arteries to ensure blood flow between mother and fetus [7]. The timing of implantation is not conserved and occurs in *H.Sapiens* 6 to 12 days after fertilization [8].

After implantation, gastrulation, the process of germ layer development, is initiated by formation of the primitive streak, which arises from a two-dimensional layer of ICM-originating cells, the epiblast cells, that undergo epithelial-to-mesenchymal transition (EMT) to form the mesoderm and endoderm [9]. Conversely, neural cells arise from the ectodermal cell layer that is positioned directly above the notochord, a cylindrical accumulation of mesoderm cells that extends from the posterior to the anterior axis [10]. These processes lay the groundwork for subsequent patterning and development of functional organs, which begins between gestational weeks 3 and 8 and lasts until and beyond birth [11].
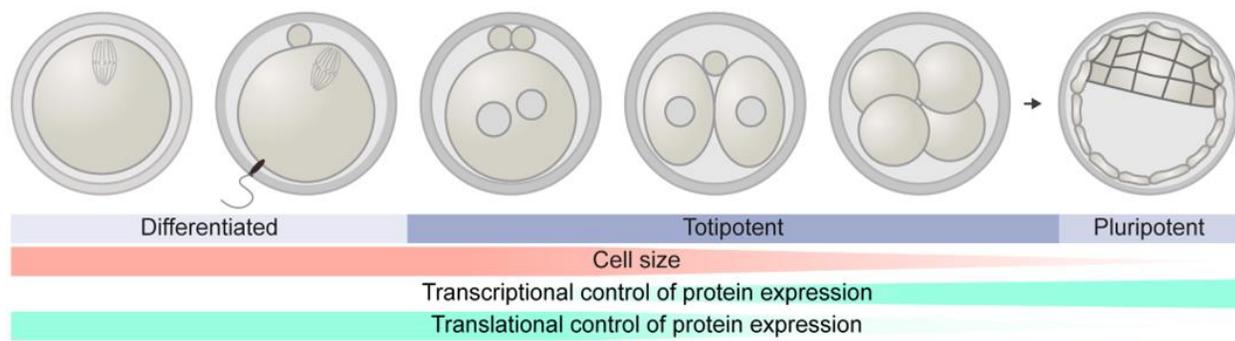
Figure 1: **Scheme of the first stages of embryonic development.** Features that are concomitant with blastocyst formation. Adapted from [12].

1.3 Modeling embryonic development with pluripotent stem cells

The study of human embryonic development is hindered by ethical concerns regarding the usage of human embryos as a research object [13], however, these concerns were to some extent circumvented by the successful isolation and cultivation of human embryonic stem cells from the ICM of IVF embryos that would have been discarded otherwise [4]. With hESCs, it is now possible to mimic germ layer formation *in vitro*, which has fueled a plethora of studies to understand the molecular mechanisms of human pre- and post-implantation development. Most notably, studies have identified the core transcriptional network underpinning pluripotency maintenance, which is composed of the genes *OCT4* (*POU5F1*), *SOX2* and *NANOG*. All three factors co-localize at the chromatin to collectively stabilize the pluripotent state [14]. Both SOX2 and OCT4 protein production is tightly regulated and expression changes of either one induces stem cell differentiation in mouse ESCs (mESCs) [15], [16]. The pluripotency factors work not in isolation but are associated with many other factors, including ESRRB, REX1 and SALL4 that have been identified

and extensively characterized by proteomic pull-down and chromatin immunoprecipitation studies [17]. Collectively, the OCT4/SOX2/NANOG complex binds to roughly 600 genes that are part of the extended network of pluripotency [18]. The acquired knowledge of the core transcriptional program in ESCs was exploited in a seminal study by Yamanaka and colleagues, who overexpressed *Oct4*, *Sox2*, *Klf4* and *Myc* (now known as Yamanaka factors) to reprogram somatic mouse fibroblasts to induced pluripotent stem cells (iPSCs) [19]. This was recapitulated shortly after in human cells by replacing *KLF4* and *MYC* with *NANOG* and the RNA binding protein *LIN28A* [20]. Nowadays, iPSCs are widely used as a research paradigm to model embryonic development and as a therapeutic tool to understand, recapitulate and correct genetic diseases [21].

1.3.1 Extrinsic factors regulating PSC maintenance

OCT4, *SOX2* and *NANOG* expression relies on an intricate network of signaling pathways fueled by extrinsic factors that are added to the culture medium. Here, profound differences in the maintenance of human and mouse ESCs are observed. Activation of the JAK/STAT signaling pathway by the addition of leukemia inhibitory factor (Lif) is instrumental to propagate mouse ESCs. Furthermore, two inhibitors of the mitogen-activated protein kinase (MAPK) and the glycogen synthase kinase 3β (GSK3β) are required to keep mESCs in a stable pluripotent state [22].

In contrast, the ligands ACTIVIN and NODAL were shown to maintain pluripotency characteristics of human ESCs by activation of the transcription factors SMAD2/3 that bind to promoters of the master pluripotency genes *OCT4* and *NANOG* [23]. Moreover, the use of FGF2 is crucial for the maintenance of hESCs by activating the RAS-MAPK, PI3K-AKT, PLCγ and STAT intracellular signaling pathways [24]. The differences in stem cell maintenance have fueled the discussion whether mouse and human ESCs are molecularly related, or in fact represent two different stages of development [25].

1.3.2 Naïve and primed ESCs

Although ESCs of human or mouse origin were both isolated from the ICM of a developing blastocyst, they depict clear molecular differences. Studies have shown that mESCs are in a "ground", or naïve state of pluripotency, which is defined by various naïve marker genes, and most importantly the state prior to X-chromosome inactivation (XCI) that takes place after blastocyst implantation. Contrarily, hESCs are considered to be in a primed state of pluripotency, analogous to mouse epiblast stem cells that can be induced from mESCs by removal of LIF and addition of ACTIVIN and FGF [26]. Besides culture requirements, hESCs differ from mESCs in their morphology (flat vs. dome-shaped), transcriptional profile

(similar to mouse epiblast stem cells) and X-chromosome activation status (X-chromosome is already silenced). Furthermore, naïve cells primarily employ mitochondrial respiration for energy generation, whereas primed cells rely on glycolysis [27], [28]. During the last 5 years, numerous studies were published that reported the conversion of human primed to naïve ESCs by using a cocktail of small molecules. All protocols are derived from the 2i/Lif culture condition for mESCs, however mostly with the supplement of additional inhibitors or growth factors such as, but not limited to, BMP4 and JNK inhibitors in combination with FGF2 or ACTIVIN [29]. Nevertheless, all studies report somewhat different gene expression profiles and morphology of naïve hESCs, hence the molecular profile of naïve hESCs is still controversial, as well as to what extent they are similar to mESCs [30].

1.3.3 Germr layer differentiation of pluripotent stem cells

The power of PSCs is undoubtedly their ability to differentiate into virtually any cell type, which opens the door for potential therapeutic applications. A myriad of studies has been published in the last 3 decades that describe differentiation protocols for the generation of various cell types from PSCs. Generally, three main signaling pathways are targeted to induce germ layer differentiation. Endoderm differentiation relies on hyperactivation of the ACTIVIN/NODAL signaling pathway [31], whereas mesoderm commitment can be achieved by activation of the WNT pathway through inhibition of GSK3β which, as consequence, increases the concentration of free β-CATENIN, the signal transducer of the WNT pathway [32]. In contrast, neuroectoderm development is commonly induced by inhibition of the TGFβ pathway including besides the ACTIVIN clade, the BMP4 signaling cascade [33]. Combinatorial treatment with distinct cytokines, often for many months, leads to differentiation of more sophisticated cell types, such as motor neurons [34], astrocytes [35], nephrons [36], hepatocytes [37] and others, which were also generated for this study.

1.4 RNA binding proteins regulate pluripotency-differentiation transition

Transcriptional and epigenetic mechanisms that govern stem cell maintenance and differentiation have been extensively researched, however, more recently, many RNA binding proteins (RBPs) were identified to be instrumental for pluripotency maintenance, mainly by regulating co- or post-transcriptional processes (**Fig. 2**). The fate of an mRNA molecule is tightly controlled by various means, including transcription initiation, capping, polyadenylation, splicing, export, translation and degradation [38], all processes that are regulated by RBPs. Transcriptome comparison across 31 different tissues showed that 6% of RBPs exhibit tissue specificity [39], and thus are able to regulate cell type-specific post-transcriptional events, which makes RBPs a versatile tool for the control of cell fate decisions.

1.4.1 Alternative splicing is crucial for pluripotency maintenance

Historically, alternative splicing has been linked first to the maintenance of ESCs and many pluripotency factors, including *OCT4*, *SALL4*, *TCF3*, *NANOG* and *DNMT3B* contain various splice isoforms that exhibit differences in their ability to maintain pluripotency [40]. This was further demonstrated by the identification of an ESC-specific isoform of the transcription factor FOXP1, which is regulated by alternative splicing of exon 18b, that confers changes in DNA binding capability of FOXP1 between pluripotent and differentiated cells [41]. Recently, a handful of RBPs were found to regulate alternative splicing in ESCs, thereby functioning either as positive or negative regulators of pluripotency. Genome-wide RNAi screens for pluripotency factors resulted in the identification of the spliceosome-associated RBP SON, which regulates splicing of pluripotency genes in hESCs [42]. Conversely, MBNL1 represses the pluripotency state in ESCs, mainly by inhibiting inclusion of exon 18b of FOXP1, thus changing its transcriptional circuit towards a differentiation program [43].

1.4.2 The alternative polyadenylation profile changes during stem cell differentiation

Besides alternative splicing, changes in the mRNA sequence can be mediated by alternative polyadenylation (APA), a process which amongst others, leads to changes in the length of the 3`untranslated region (UTR) [44]. Technologies for mapping and identification of APA sites significantly improved over the last years and demonstrated that approximately 70% of all transcripts undergo APA [45]. Differentiation of PSCs and conversely, the reprogramming of fibroblasts, is accompanied by respective lengthening or shortening of 3`UTRs, providing an intriguing connection between APA and the pluripotency-differentiation transition. A mechanistic link between APA and stem cell differentiation has been lacking, until Lackford and colleagues showed that FIP1, a subunit of the canonical cleavage and polyadenylation specificity factor (CPSF), is important for self-renewal of mESCs. The authors demonstrated that *Fip1* is developmentally regulated and changes the polyadenylation pattern of several hundred genes, many of which are important for embryonic development [46]. Along these lines, we have recently identified TDP-43 as an important regulator of stem cell differentiation and somatic reprogramming. TDP-43 is a multifunctional RBP that is highly expressed in ESCs but down-regulated upon differentiation. We showed that many pluripotency factors change their polyadenylation profile upon TDP-43 depletion, in a manner that is similar to the changes that occur during differentiation. This was confirmed by the identification of TDP-43 binding sites at UG-repeats surrounding the polyadenylation site of deregulated transcripts, which include amongst others the mRNA of the master

pluripotency factor SOX2. Together, we found that TDP-43 is important for pluripotency-differentiation transition in m/hESCs by regulating APA of many pluripotency-associated transcripts [47].

To summarize, a growing body of literature underscores the importance of RBP-mediated post-transcriptional gene regulation for stem cell maintenance and differentiation (**Fig. 2**).
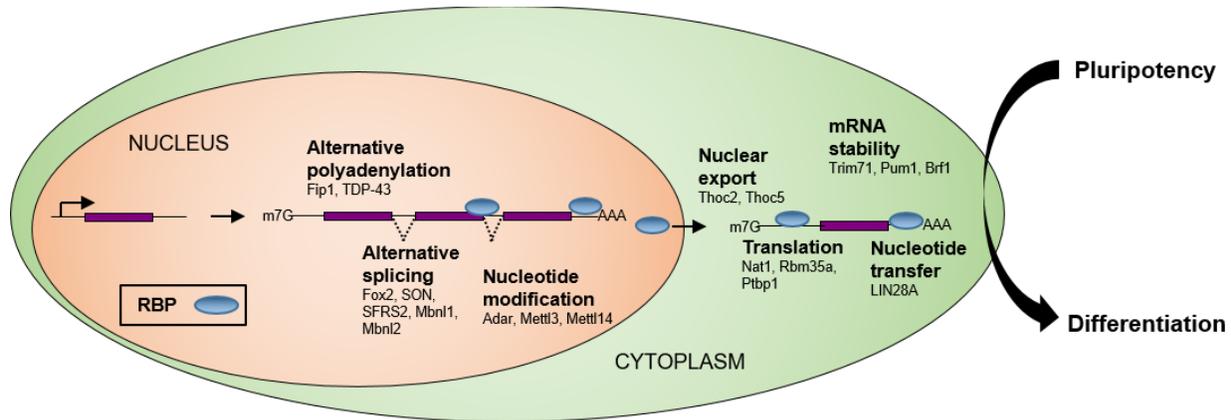
Figure 2: **RNA binding proteins mediate the pluripotency-differentiation transition.** Adapted from [48].

1.5 Long non-coding RNAs are new players in embryonic development

Long non-coding (lnc) RNAs represent a class of non-coding RNAs that have a profound impact on cell fate decisions [49]. More than 50000 lncRNAs have been detected in *H.Sapiens*, defined by having a length of >200 nucleotides, however only ~1000 exhibit a moderate-to-high expression level, out of which, 300 are conserved across mammals and other vertebrate species [50]. LncRNAs are typically shorter, have fewer exons and are one order of magnitude lower expressed than mRNAs. Moreover, they tend to be more temporarily and spatially regulated and evolve much faster compared to mRNAs, which is manifested by the lack of homologs for many lncRNAs [51]. LncRNAs can adapt a complex secondary or higher-ordered structure to orchestrate binding of RBPs [49]. There are nuclear lncRNAs that have been implicated in guiding chromatin modifiers to mediate transcription and cytoplasmic lncRNAs that control stabilization and mRNA translation [49].

Nuclear lncRNAs can be broadly distinguished in *cis*- and *trans*-acting lncRNAs depending on their mode of action which is either confined to the lncRNA gene locus (*cis*) or to a distal gene locus (*trans*) [52] (**Fig. 3**). Examples of *cis*-acting lncRNAs include *Xist* and *Kcnq1ot1*, which have been implicated respectively in X-chromosome inactivation and genomic imprinting. *Xist* is transcribed from the X-chromosome to which it binds to initiate the formation of transcriptionally inactive heterochromatin [53]. Interestingly, *Xist* itself is regulated by lncRNAs such as its antisense non-coding transcript *Tsix*, which represses *Xist* expression by inducing epigenetic modifications at its promoter, and *Jpx*, which acts as an

activator of *Xist* by sequestration of the transcriptional repressor CTCF [54]. A similar mechanism was shown for *Kcnq1ot1*, which is required for genomic imprinting, the gene silencing mechanism of only one parental chromosome during gametogenesis [55].

Moreover, lncRNAs are involved in the regulation of *HOX* genes, which are instrumental for spatiotemporal control of body axis formation [56]. A lncRNA that regulates *HOX* gene expression in *cis* is *HOTTIP*, which recruits the histone complex MLL1 and activates distal *HOX* gene promoters [57]. *HOX* genes are also regulated in *trans* by the lncRNA *HOTAIR*, which interacts with repressive histone-modifying complexes and recruits them to specific target genes [58].

Furthermore, many lncRNAs have been implicated in organogenesis, including *braveheart* and *Fendrr*, which are required for coordinated heart development [59], [60], *H19*, which sequesters miRNA *let-7* to regulate muscle differentiation [61] and *TINCR*, which is required for keratinocyte differentiation, likely by binding to STAU1 and stabilizing differentiation-associated transcripts [62]. Similarly, brain development is affected by lncRNAs such as *Malat1*, a conserved, abundant lncRNA that regulates synapse formation in cultured mouse hippocampal neurons [63] and *Dlx6os1*, which mediates expression of neighboring *Dlx* genes that are important for forebrain development [64].

Lastly, genome-wide mapping of chromatin marks of actively transcribed genes outside of known protein-coding genes revealed that PSCs express thousands of lncRNAs [65]. A comprehensive loss-of-function screen of 147 lncRNAs by Guttman and colleagues showed that many of them affect the pluripotency characteristics of mESCs [66], however, lacking mechanistic insights and how this relates to human differentiation. Some of these lncRNAs are controlled by OCT4 or NANOG, for instance *MIAT* and *AK141205*, which in turn regulate pluripotency maintenance [67]. Moreover, lncRNA *TUNA* maintains mESCs by interaction with RBPs that bind to the promoters of the pluripotency factors *Nanog*, *Sox2* and *Fgf4* [68]. In hESCs, the lncRNA *lncPRESS1* was shown to interact with the histone H3 deacetylase SIRT6 to prevent its access to chromatin, thus maintaining histone acetylation at promoters of pluripotency genes [69]. Moreover, the lncRNAs *lncRNA-ES1* and *lncRNA-ES2* are abundantly expressed in hESCs, where they interact with the repressive Polycomb protein SUZ12 and the pluripotency factor SOX2 to block neural differentiation [70]. Another example includes *linc-RoR* which is enriched in iPSCs and sequesters *miRNA-145* to impair differentiation [71]. There are also developmentally regulated lncRNAs such as *DIGIT* and *TERRA* that are respectively up-regulated during endoderm differentiation [72] or down-regulated upon exit from pluripotency [73], however it is not clear whether they have a function during this process.

To summarize, there are several lncRNAs that regulate gene expression during mammalian development, nevertheless, the function of many lncRNAs remains uncharacterized, especially in human cell types, which is one objective of this thesis.
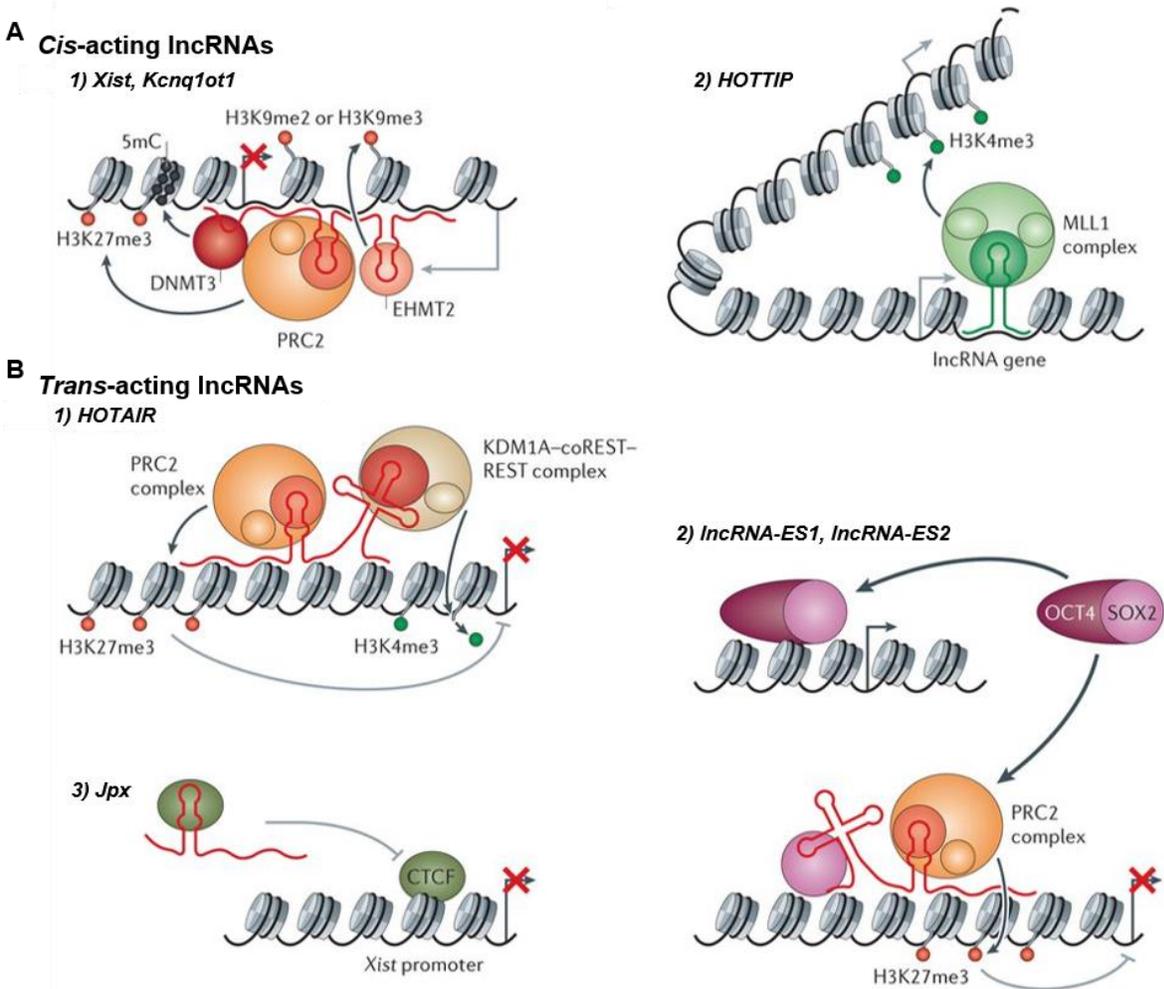
Figure 3: **The mode-of-action of nuclear lncRNAs**. Nuclear lncRNAs are classified into *cis-* (**A**) and *trans-*acting (**B**), depending on whether they mediate gene expression of neighboring or distal genes, respectively. Modified from [49].

1.6 Membraneless organelles are phase-separated entities

Many RBPs and lncRNAs exhibit spatial organization by aggregation into higher-ordered, macromolecular structures, often referred to as granules, membraneless organelles or condensates, which are prevalent in the nucleus (for example: nucleoli, Cajal bodies, gems, speckles, paraspeckles, histone-locus bodies, promyelocytic leukemia (PML) bodies) and the cytoplasm (for example: P-bodies, stress granules, germ granules, RNA transport particle) [74]. Due to the lack of a confining membrane, these granules are highly dynamic and able to react rapidly to environmental changes in order to spatiotemporally control

biochemical reactions. In contrast to membrane-surrounded organelles, it is still unsettled how formation and maintenance of membraneless organelles is achieved [75]. Hyman and colleagues demonstrated in a seminal publication that P-bodies in *C.elegans* have fluid-like properties, as they are spherical, can be deformed under stress and recover quickly after laser-induced photobleaching of a fluorescently-labeled P-body protein [76]. By now, many other membraneless organelles were identified to behave like "liquid droplets", amongst others nucleoli [77], stress granules [78] and the centrosome [79]. The liquid-like property was postulated for other granules as well, however, a rigorous experimental assessment is yet awaiting.

Based on their properties, it was proposed that membraneless aggregates form by liquid-liquid phase separation, a process which is analogous to an oil-water vinaigrette, where after vigorous mixing, oil and water remain separated in two phases. Whereas entropy would favor a mixing of the two substances, homophobic interactions between the molecules lead to a system with lower free energy, thereby inhibiting the mixing process [75]. A similar process was proposed for the aggregation of RBPs, which often possess intrinsically-disordered domains that mediate intermolecular interactions [74]. Recent developments indicate that also the formation of heterochromatin is driven by phase separation [80].

To summarize, the process of phase separation is an emerging concept in cell biology by which the formation of membraneless granules is explained, which happens mainly through interactions of molecules with similar biophysical properties.

1.7 Composition and function of paraspeckles

One type of granule that is thought to assembly by liquid-liquid phase separation are nuclear paraspeckles [81]. In 2002, paraspeckles were identified by Archa Fox and colleagues who performed a proteomic screen for nucleoli proteins and found the RBPs SFPQ and PSPC1 that form distinct nuclear foci in close proximity to splicing speckles [82]. Almost a decade later, a comprehensive screen for paraspeckle components was performed, extending the repertoire of known paraspeckle proteins to 40 members [83] (**Fig. 4A**). Intriguingly, paraspeckles contain mainly RBPs with annotated function in mRNA processing, which makes them a potential hub for co- and post-transcriptional processes [84].

Besides proteins, various types of RNAs are enriched in paraspeckles (summarized in **Fig. 4B**), most notably, the lncRNA *NEAT1* [85]–[88]. The *NEAT1* gene produces a short and a long isoform, from here on referred to as *NEAT1_1* (3.7 kb) and *NEAT1_2* (23 kb), previously known as *MENε/β*. Both are single exon RNAs, however with different modes of terminal end processing. While *NEAT1_1* is polyadenylated, *NEAT1_2* contains a triple helix t-RNA-like structure at its 3`end, a feature that is shared

with 129 other lncRNAs in vertebrates [89]. Electron microscopy and super-resolution studies have uncovered that *NEAT1_1* and the 3`and 5` arms of *NEAT1_2* are located at the outer shell of paraspeckles while the middle segment of *NEAT1_2* is positioned in the paraspeckle center [90], [91] (**Fig. 4C**). Paraspeckle proteins exhibit a similar spatial relationship [91], which provides evidence that paraspeckles, similar to other phase-separated entities like germ granules [92], stress granules [93] and nucleoli [94] contain sub-compartments that might assume different functions within the granule. By differential over-expression of short and long *NEAT1* isoform, it was shown that only *NEAT1_2* is crucial for paraspeckle formation [85], [86]. While *NEAT1_2* is solely localized in paraspeckles, *NEAT1_1* can also occupy space outside of paraspeckles, where it mostly exists as single molecules, however, it is unknown, whether *NEAT1_1* has paraspeckle-independent functions [95].

In mammalian cells, transcription of *NEAT1_2* is regulated by an intricate network of RBPs that bind to the *NEAT1* polyadenylation site. By mutating binding sites of the polyadenylation complex CFIm and the RBP hnRNPK, Naganuma and colleagues showed that the former is inducing polyadenylation, thereby producing *NEAT1_1*, while binding of the latter prevents polyadenylation by capturing and inactivating a subunit of the CFIm complex, consequently leading to transcription of *NEAT1_2* [83]. Following those events, the essential paraspeckle proteins SFPQ and NONO bind to the nascent *NEAT1_2* transcript and stabilize it. More paraspeckle proteins are recruited that are likely responsible for compaction and keeping all proteins in place before transcription of *NEAT1_2* is terminated and a mature paraspeckle diffuses away from the transcription start site [81].
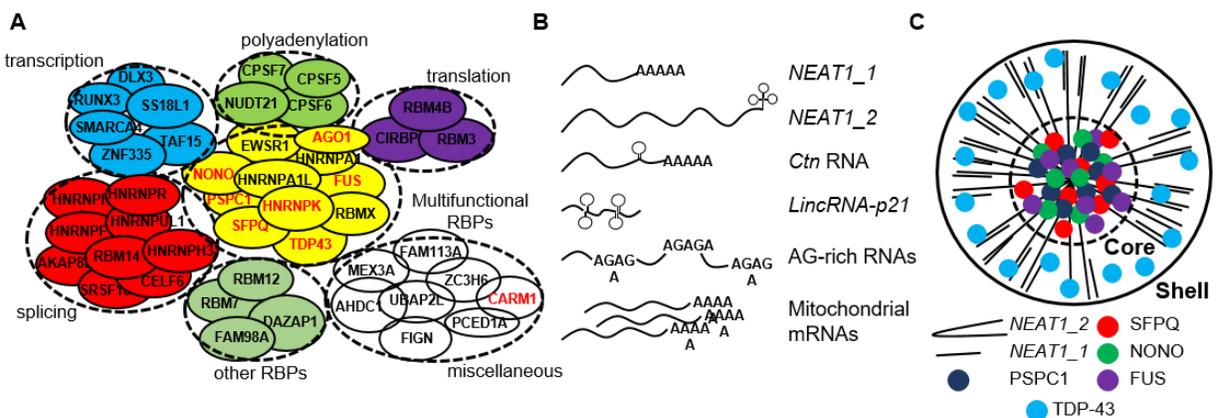
Figure 4: **A scheme of paraspeckle components**. **A)** A summary of paraspeckle proteins as listed before [81] including AGO1 and CARM1 that have been identified, recently [91], [96]. Proteins in red are mentioned more in detail in this study. **B)** A summary of paraspeckle RNA components. Adapted from Fox *et al*. [81] and including *lincRNA-p21* and mitochondrial mRNAs, which were recently found to be localized in paraspeckles [97], [98]. **C)** A schematic depiction of the paraspeckle substructure as shown before [91].

1.7.1 Molecular mechanism of paraspeckles

The last decade of research uncovered three main mechanisms of how paraspeckles exert their function, which is either by retention of RNA, sequestration of proteins or binding to chromatin, or likely a combination of all three mechanisms (**Fig. 5**).

1.7.1.1 RNA retention

RNA retention was shown first for the RNA CTN, which is the 3`UTR-extended isoform of the mouse cationic amino acid transporter 2 (mCat2) mRNA. While the long CTN-RNA is retained in the nucleus, it is endonucleolytically cleaved upon stimulation of interferon gamma or polysaccharide receptors, for instance during viral infection. This results in the production of the shorter mCat2 mRNA, which is exported to the nucleus and translated [99]. There is evidence that nuclear RNA retention is triggered by binding and sequestration of the core paraspeckle protein NONO to hyper-edited stretches within an RNA. RNA editing is defined as the change of the RNA sequence, for instance by an adenosine-to-inosine switch that is mediated by the enzyme ADAR via adenosine deamination [100]. ADAR-mediated editing occurs primarily on double-stranded RNA sequences, likely as part of an immune response to target viral double-stranded RNA for degradation [101]. Recent computational analysis showed that humans contain 333 mRNAs with putative double-stranded regions, which are mostly originating from primate-specific short interspersed elements (SINEs), so-called *Alu* elements that comprise 11% of the human genome [102]. Inverted repeat *Alu* elements in an mRNA can base-pair and form double-stranded regions that are targeted by ADAR and then potentially bound by NONO and retained in paraspeckles. Evidence for this hypothesis was provided by Chen and colleagues who fused GFP reporter mRNAs with inverted *Alu* repeat elements and observed their translocation to paraspeckles [103].

Besides *Alu*-element containing mRNAs, pull-down studies of *Neat1*-associated RNAs revealed that paraspeckles in murine cells contain AG-rich, intronic RNA sequences [91]. A similar study in human U2OS and HEK293 cells uncovered that mRNAs encoding for mitochondrial genes were found in paraspeckles, which was shown to be the mechanism for cross-talk between paraspeckles and mitochondria [97]. Given the little overlap in the identified RNA species between those studies, it is likely that the RNA repertoire of paraspeckles is dynamic and dependent on the cellular context.

1.7.1.2 Protein sequestration

Paraspeckles, similar to many other membraneless organelles [104], can sequester proteins and by this alter or inactivate their function in the nucleus. A particularly well-studied example is the translocation of the

core paraspeckle protein SFPQ from the nucleoplasm to paraspeckles in HeLa cells that were transfected with poly I:C double-stranded RNA, which mimics viral infection and increased the formation of paraspeckles. ChIP-qPCR analysis indicated that SFPQ binds and represses IL-8, a major immunological response gene [105]. The authors propose a model where virally-induced paraspeckle formation triggers SFPQ relocation from chromatin to paraspeckles, which in turn leads to expression of IL-8 for appropriate viral response. A question that is rarely addressed is how much protein is sequestered to paraspeckles and whether this is meaningful, considering that SFPQ and other paraspeckle proteins are also found in the nucleoplasm. Hirose *et al*. estimated that inhibition of the proteasome activity, which induces *NEAT1* expression, resulted in relocation of 50% of nucleoplasmic SFPQ and NONO to paraspeckles. They showed that SFPQ is required for expression of RNA-specific adenosine deaminase B2 (ADARB2), which consequently, is down-regulated upon proteasome inhibition and SFPQ translocation [106]. Furthermore, it was shown that SFPQ and NONO bind to many primary miRNA transcripts in the nucleus, which enhanced their processing by the Drosha-DGCR8 microprocessor complex. The authors demonstrated that *NEAT1_2* sequesters SFPQ, NONO and the microprocessor, which is required for effective pri-miRNA processing, thereby providing a mechanistic link for many previous studies that described aberrant miRNA expression upon *NEAT1* misregulation in various cancer cells [107].

1.7.1.3 Chromatin binding

Many lncRNAs exert their function by direct binding to DNA and attracting chromatin-remodeling complexes [108]. Using complementary capture oligonucleotides, West *et al*. identified binding sites of *NEAT1* at hundreds of genomic loci, the majority of which are positive for H3K4me3, a marker of actively transcribed genes and interestingly also co-occupied by *MALAT1*, a lncRNA that is located next to *NEAT1* in the genome [109]. This indicates that *NEAT1* acts as sponge for chromatin regulators and might even be involved in the spatial organization of the chromatin. Indeed, the catalytic subunit of the SWI/SNF complex BRG1 was found to be associated to paraspeckles [110] and it was shown that nuclear AGO1 is sequestered in paraspeckles where it is involved in maintaining chromatin organization [96]. Recently, it was reported that *NEAT1* undergoes triple helix formation with the double-stranded DNA via Hoogsteen base-pairing, suggesting that the mode of binding is sequence-mediated [111].

To summarize, paraspeckles have vast possibilities to alter gene expression and it remains to be analyzed for each cell type individually, whether regulation of RNA, protein or chromatin is the prevalent mode-of-action of paraspeckles.
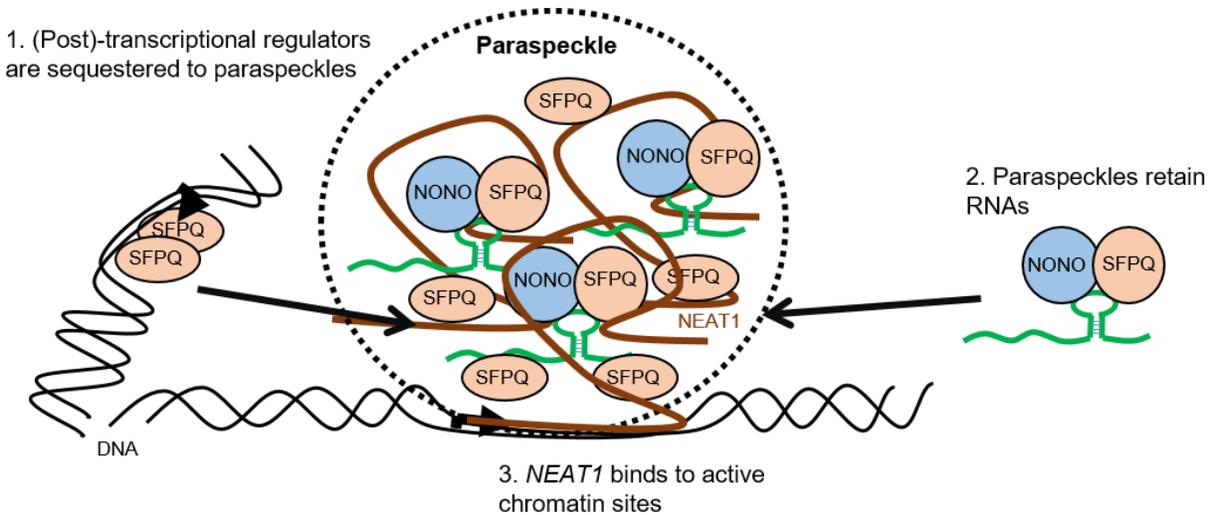
Figure 5: **A schematic overview of the molecular functions of paraspeckles**. References are listed in the section above.

1.7.2 Paraspeckles in development and disease

Countless studies have demonstrated the up-regulation of paraspeckles in various disease models and in certain stages of development (**Fig. 6**), however, addressing mostly correlation and failing to establish causative relationships. The following sections outline the physiological functions of paraspeckles in mammals.

1.7.2.1 Paraspeckles in development

The first link of paraspeckles to developmental processes was established after demonstrating their up-regulation during differentiation of myoblasts to myotubes, a study that identified *NEAT1* as the core component of paraspeckle [87]. Following the generation of *Neat1-/-* mice, Nakagawa and colleagues observed that the number of parturitions from female *Neat1-/-* were reduced by 50%, concomitant with reduced litter size and serum progesterone levels [112]. The latter is a hormone that is produced in the ovarian corpus luteum, a tissue that contains many paraspeckles and which was disrupted in *Neat1-/-* mice. A thorough dissection of *Neat1* dynamics in pre-blastocyst development revealed that in mice, paraspeckles are highly up-regulated during the 4-cell developmental stage before being down-regulated upon blastocyst formation [113]. The authors showed that paraspeckles contain the arginine methyltransferase CARM1, which is instrumental for the arginine methylation of histone H3, an activating mark for the development of embryonic rather than extra-embryonic tissue [114]. Over-expression or down-regulation of *Carm1* inhibited aggregation of the core paraspeckle constituent NONO and conversely knock-down of

Neat1 reduced CARM1 speckles in the 4-cell embryo. Intriguingly, down-regulation of *Neat1* or *Nono* resulted in a developmental arrest at the 16- or 32-cell stage, which the authors explained by up-regulation of *Cdx2*, a transcription factor that is crucial for the development of extra-embryonic tissue. While this study showed convincingly the contribution of *Neat1* to pre-blastocyst development, the function and dynamics of paraspeckles in gastrulation remains enigmatic and is one objective of this study.

1.7.2.2 Paraspeckles in disease

1.7.2.2.1 Paraspeckles in cancer

Whereas the developmental aspect of paraspeckles is understudied, many publications established a link between paraspeckle formation and disease progression (**Fig. 6**). Undeniably, the bulk body of literature focuses on paraspeckle contribution in cancer. Up to now, studies showed elevated *NEAT1* expression in the following tumors: lung cancer, esophageal squamous cell carcinoma, laryngeal squamous cell carcinoma, oral squamous cell carcinoma, nasopharyngeal carcinoma, endometrial cancer, cholangiocarcinoma, colorectal cancer, hepatocellular cancer, breast cancer, ovarian cancer, prostate cancer, pancreatic cancer, thyroid carcinoma, osteosarcoma, renal cell carcinoma and glioma [115], [116]. These studies suggested that for many cancer types, *NEAT1* could serve as a prognostic biomarker whose up-regulation indicates poor patient outcome. The causes of *NEAT1* up-regulation are diverse and range from an increase in copy numbers [115], mutations of the *NEAT1* promoter [117] and transcription factors that target the *NEAT1* locus such as hypoxia-inducible factor 2 (HIF-2) [118] and RUNX1 [119]. Moreover, it was shown that the EGFR signaling pathway with its downstream effectors STAT3 and NF-ƙB activates the *NEAT1* promoter in glioblastoma [120]. Adriaens *et al.* recently demonstrated that the tumor suppressor gene p53 induces paraspeckle formation in skin fibroblasts and that silencing of *NEAT1* impaired skin tumorigenesis [121]. Mechanistically, it was shown that *NEAT1* sponges and inhibits many miRNAs, which leads to increased cell proliferation, migration, invasion, EMT and chemoresistance [115]. Moreover, *NEAT1* binds EZH2, a subunit of the polycomb repressive complex and change expression of its downstream targets [122]. While these studies clearly emphasize the role of *NEAT1* as oncogene, there is also evidence that *NEAT1* acts as tumor suppressor in acute promyelocytic leukemia where *NEAT1* is significantly down-regulated compared to healthy subjects. This discrepancy might be due to the different environment in solid tumors relative to cancer arising from the blood [116].

1.7.2.2.2 Paraspeckles during viral infection

The infection with RNA viruses including Japanese encephalitis, HIV, rabies, influenza, Hantaan and herpes simplex virus (HSV) is concomitant with up-regulation of *NEAT1* [123]. Hantaan virus induces *NEAT1* by the RIG-I signaling pathway and establishes a feed-forward loop by sequestration of paraspeckle core protein SFPQ, which in turn cannot exert its inhibitory effect on *RIG-I* expression, thus leading to more paraspeckles [124]. A different mechanism was demonstrated for HIV-1 infection where *NEAT1* retains HIV-1 mRNAs in the nucleus and a down-regulation of *NEAT1* is accompanied by enhanced virus replication [125]. *NEAT1* was also attributed with pro-viral activities as shown in HSV whose DNA is bound by the paraspeckle proteins PSPC1 and NONO to facilitate the interaction between STAT3 and viral gene promoters, which increased viral infection [126]. To summarize, it depends on downstream mechanisms whether *NEAT1* is pro- or antiviral, which is reminiscent to the function of paraspeckles in cancer [123].

1.7.2.2.3 Paraspeckles in neurodegenerative diseases

Finally, paraspeckle formation was linked to the onset or progression of diverse neurodegenerative diseases. Whereas healthy neurons are devoid of paraspeckles, Nishimoto and colleagues found paraspeckles in spinal motor neurons from patients with the sporadic form of amyotrophic lateral sclerosis (ALS) [127]. This was supported by similar findings in familial ALS caused by mutations in the *C9ORF72* and *TARDBP* genes. Likely, the increase in paraspeckles is due to nuclear depletion and inactivation of TDP-43, which we and others have shown to regulate *NEAT1* expression [128] and which happens in 95% of all sporadic ALS patients. Of note is that 8 out of 25 proteins, genetically associated with ALS, are paraspeckle proteins. These proteins can be mutated and sequester other paraspeckle proteins, thereby disrupting proper paraspeckle-mediated signaling in ALS [129]. Furthermore, *NEAT1* up-regulation was observed in Parkinson`s [130], Huntington`s [131] and Alzheimer`s disease [132], however the mechanistic impact of paraspeckles in these pathologies remains to be studied.
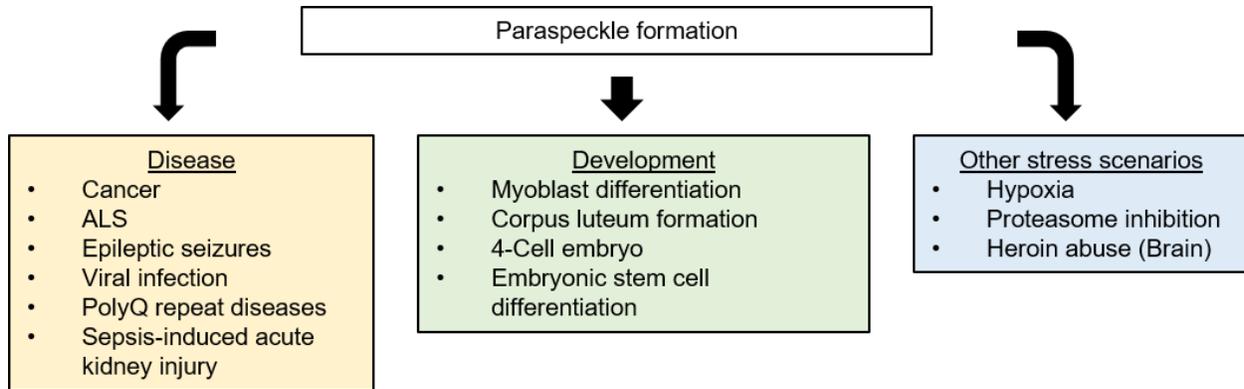
Figure 6: A summary of conditions and diseases that are concomitant with up-regulation of *NEAT1*. References are listed in the text above.

1.8 DBHS proteins are involved in transcriptional and post-transcriptional gene regulation

DBHS proteins were the first components identified in paraspeckles [133]. Humans express three members of the DBHS protein family, namely SFPQ (PSF), NONO (p54nrb) and PSPC1 (PSP1), which all contain conserved RNA recognition domains (RRMs), a NonA/paraspeckle domain (NOPS) and a C-terminal coiled-coil domain (**Fig. 7**) [134]. Importantly, DBHS proteins rarely act alone but are able to homo- and heterodimerize with each other via reciprocal interaction of RRM2, NOPS and coiled-coil domains of both proteins, thereby forming a globular structure with extended coiled-coil domains [135]. Recently, Lee *et al.* showed that the coiled-coil extension is used as oligomerization site of many DBHS protein dimers, which is crucial for paraspeckle formation [136]. On some occasions, the loss of one DBHS protein can be compensated, for instance by up-regulation of *Pspc1* in *Nono⁻ᐟ⁻* mouse fibroblasts to form a functionally intact heterodimer with SFPQ [135]. On the other hand, cognitive disabilities in mice are a result of NONO depletion and cannot be compensated by SFPQ [138] and similarly, knock-down of PSPC1 in HeLa cells is not compensated by SFPQ and NONO and results in reduced cell proliferation and increased DNA damage [139]. It should be emphasized that many studies did not discuss the possibility of heterodimerization and focused only on the function of one DBHS protein. Nevertheless, important insights into molecular functions of DBHS proteins were generated in those studies, which are summarized in the next paragraphs.

Figure 7: Protein domains of DBHS family members SFPQ, NONO and PSPC1. Adapted from Knott *et al*. [135]. Low complexity domains are indicated in dashed boxes, as well as the uncharacterized DNA binding domain in SFPQ. Numbers indicate amino acid boundaries in *H. sapiens*.

1.8.1 Molecular functions of DBHS proteins

Members of the DBHS protein family bind to single- and double-stranded DNA and RNA and hence have multiple functions depending on the cellular context **(Fig. 8)** [135]. They have been described in all stages of the mRNA live cycle, starting with transcription that is mainly mediated by NONO, which binds to many gene promoters, as shown by chromatin immunoprecipitation and sequencing (ChIP-seq) experiments, including photoreceptor genes to regulate their expression in mouse retina cells [140]. Another study demonstrated that NONO acts as a bridge between RNA polymerase II (RNA-PolII) and a coactivator complex of the cAMP response pathway, thereby mediating the expression of cAMP-response genes [141]. In contrast to NONO, SFPQ acts mainly as transcription repressor, exemplified by studies that identified SFPQ binding to promoters of hormone receptors where it recruits the epigenetic silencer proteins HDAC or Sin3A [142], [143]. Moreover, SFPQ and NONO are required for transcription elongation by binding to RNA-PolII and mediating co-transcriptional processing and termination, the latter by recruiting the exonuclease XRN2 [142]. Furthermore, they are known to stabilize transcripts, most prominently the lncRNA *NEAT1*, but also histone-encoded mRNAs [145]. Splicing of pre-mRNAs was the first activity that was described for SFPQ [146]. Many studies have identified SFPQ and NONO as associated non-essential factors of the human spliceosome that regulate alternative splicing of the tyrosine phosphatase CD45 in T cells [145], the microtubule-binding protein Tau [148], neural-specific genes [149] and the spinal muscular atrophy genes *SMN1*/*SMN2* [148]. DBHS proteins also mediate RNA export and transport, which was shown for U snRNAs whose export is facilitated by SFPQ and NONO [151]. Furthermore, neurons have a cytoplasmic pool of SFPQ and NONO, which are part of the RNA transport granule [152]. Recently, Cosker and colleagues demonstrated that SFPQ binds and localizes multiple, functionally related mRNAs essential for axon survival [153]. The molecular functions of DBHS proteins in human PSCs have not been analyzed, yet, and are one aim of this study.

1.8.2 Physiological roles of DBHS proteins

DBHS proteins are best characterized by their function in a) circadian rhythm, b) DNA damage repair and c) paraspeckle assembly. The circadian rhythm is a mechanism where organisms adapt to a 24 hour, day-and-night cycle [154]. In mammals, this is mainly controlled by the PER proteins (*Period1 and Period2*), which regulate the transcription of further components of the circadian rhythm in an oscillating manner [155]. Mass spectrometry analysis identified NONO as a PER protein interactor, which antagonizes its function and consequently, depletion of the *Drosophila* homolog NonA results in arrhythmic flies [156]. Moreover, it was shown that the SFPQ protein amount oscillates with the day-and-night cycle and that it recruits the HDAC-Sin3A histone deacetylation complex to repress the expression of *Per* genes [157]. DBHS proteins are also crucial for DNA double-strand repair by non-homologous end joining or homologous recombination. SFPQ and NONO bind directly to the DNA ends and interact with other proteins of the double-strand break pre-ligation complex [158], [159]. Furthermore, SFPQ is involved in homologous recombination by assisting in strand invasion, D-loop formation and topoisomerase activity [135].
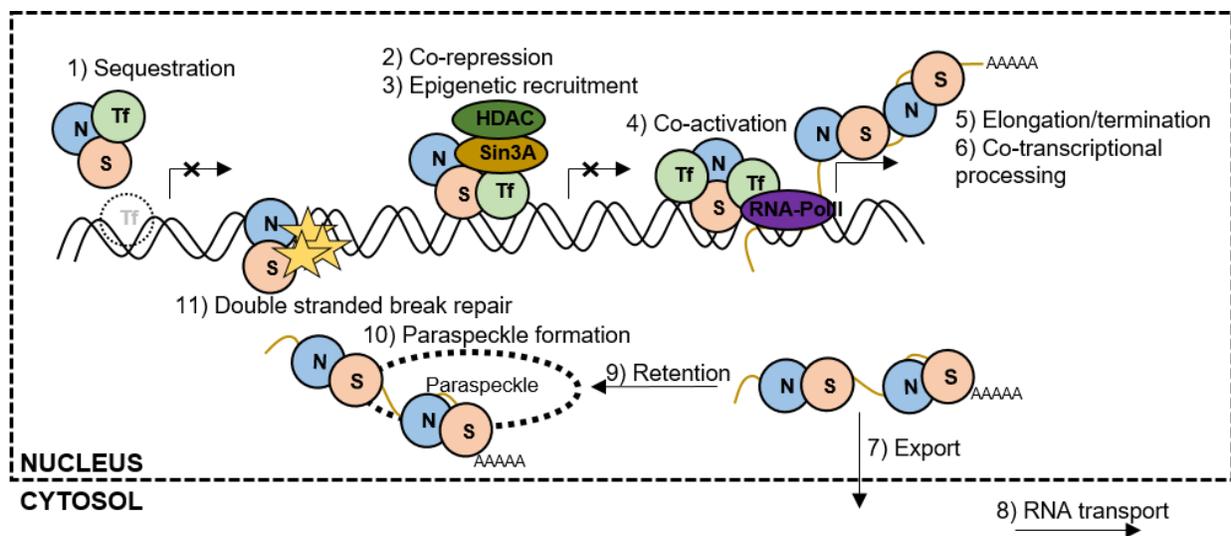
Figure 8: A scheme depicting the molecular functions of SFPQ and NONO. Adapted from Knott *et al.* [135].

1.8.3 DBHS proteins in disease

Large-scale analysis of genetic variations in humans has shown that DBHS proteins belong to the class of genes with the lowest tolerance of missense and loss-of-function mutations [160], stressing their importance in disease and development.

Similar to paraspeckles, DBHS proteins are deregulated in many cancer types, however often exhibiting trends that are not correlated with paraspeckles. *NONO* is up-regulated in malignant breast

cancer where it binds and stabilizes SREBP-1A, a master regulator of lipid biogenesis [161]. Both *SFPQ* and *NONO* are also up-regulated in colorectal cancers and a potential oncogenic function of both proteins is hypothesized [162], [163]. Another feature of DBHS proteins is their ability to modulate the immune response after viral infection, exemplified by SFPQ, which influences the transcription and processing of virus RNAs from hepatitis delta [164], influenza A [165] and HIV [166], the latter also being regulated by NONO [167]. Lastly, DBHS proteins affect neural development and mutations in *NONO* were connected to intellectual disability in mice and humans [138]. Moreover, all three DBHS proteins were shown to be important for neuronal development and axonal growth [135].

Taken together, DBHS proteins exert a multitude of cellular functions due to their ability to oligomerize, and to interact with DNA, RNA and a plethora of other proteins. DBHS proteins are sequestered to paraspeckles and are in fact important for their structural integrity, hence their function is linked with the appearance of paraspeckles, although many studies fail to analyze this connection.

1.9 Aims and impact of this work

LncRNAs can be important regulators of mammalian embryogenesis and disease, however, the function of many lncRNAs during human embryonic development remains enigmatic. Given that there is a plethora of lncRNAs that form condensates in the nucleus which likely have an impact on gene expression, I sought to first identify nuclear lncRNAs that exhibit dynamic regulation during germ layer differentiation. To this end, I curated a panel of 27 lncRNAs and analyzed their expression in PSC-derived multipotent progenitor cells. I then focused on the paraspeckle lncRNA *NEAT1*, which is up-regulated at the onset of differentiation in a lineage-independent manner. Developmental studies of *NEAT1* were primarily performed in the murine system and data for paraspeckle formation in human cell types was lacking. To overcome this knowledge gap, I employed PSCs to construct an atlas of paraspeckle trajectories in 24 human cell types including multipotent progenitors and terminally differentiated cells and dissected the expression of *NEAT1* isoforms. I identified many novel cell types, especially from the mesenchymal and glial lineage tree that contain a high amount of paraspeckles. To explain the heterogeneity in the number of paraspeckles between cell types, I analyzed cellular and molecular features that correlate with paraspeckle appearance and found that nuclear size is one factor that can predict paraspeckle formation. Moreover, I identified the RBP TDP-43, which promotes the polyadenylation of *NEAT1_1* in PSCs and its down-regulation upon pluripotency exit induced paraspeckle formation. Next, I sought to address molecular features of paraspeckle formation and found that small DNA-binding molecules, which are

regularly used in chemotherapy, can disintegrate paraspeckles and other chromatin-embedded lncRNA condensates, thereby demonstrating that DNA accessibility is important for the formation of nuclear DNA-associated granules. Given that paraspeckles are up-regulated in neurodegenerative diseases, but also in many tumors, this finding might be relevant to develop strategies to dissociate paraspeckles and thereby having impact on disease progression.

Moreover, by genome editing approaches, I sought to interrogate the functional connection of *NEAT1* isoforms and stem cell differentiation, which revealed that the architectural isoform of paraspeckles, *NEAT1_2*, is required for coordinated differentiation, whereas *NEAT1_1* is dispensable for germ layer commitment. These findings were supplemented by the analysis of DBHS proteins SFPQ, NONO and PSPC1, which are regulated by translocation to paraspeckles upon differentiation. Selective knock-out of each member revealed that depletion of *SPPQ* could not be tolerated, whereas knock-out of *NONO* and *PSPC1* primed hESCs for differentiation. Functional analysis showed that NONO mainly regulates genes of the cholesterol pathway, which is important for stem cell differentiation, whereas SFPQ generally regulates expression of genes involved in cell homeostasis, probably by acting as a polyadenylation factor.

To summarize, my work represents the first attempt to dissect the function of individual members of paraspeckles in maintenance and differentiation of pluripotent stem cells. I uncovered novel principles for the formation of lncRNA-containing, chromatin-embedded granules, which adds a layer of complexity to understanding mechanisms of formation of phase-separated granules. Furthermore, this study provides a comprehensive analysis of paraspeckle trajectories in human cell types and is important as a resource to tackle cell type-specific functions of paraspeckles.

2. Materials and Methods

2.1 Chemicals and kits

All chemicals were of research-grade; routinely used reagents are denoted in **Supplementary Table 1**. Kits that were used in this study are listed in **Supplementary Table 2**.

2.2 PSC culture

Human ESCs of the H9 line (WiCELL Research Institute) and iPSCs were cultured in StemMACS iPS-Brew XF (Miltenyi Biotec) and passaged by StemMACS Passaging Solution (Miltenyi Biotec) on tissue culture-treated plates (Sigma) coated with Matrigel (ThermoFisher Scientific) diluted 1:100 in DMEM/F-12 (ThermoFisher Scientific). All differentiation experiments were carried out with H9 cells, except lung progenitor and cortical neuron differentiation, which were performed with iPSC lines, namely NKX2.1-P2A-eGFP [168] and foreskin fibroblast-derived iPSCs [169], respectively. For paraspeckle measurements in trophoblast progenitors and neural crest cells, I used differentiation protocols, as previously described [170], [171].

2.3 Fibroblast reprogramming

The reprogramming of human neonatal dermal fibroblasts was performed using StemRNA 3rd Gen Reprogramming Kit (Reprocell) according to the manufacturer`s protocol. The RNA transfection cocktail included synthetic, non-modified RNA of reprogramming factors OCT4, SOX2, KLF4, cMYC, NANOG and LIN28A, immune evasion mRNAs of E3, K3, B18 and reprogramming-enhancing mature, double-stranded microRNAs from the 302/367 cluster. $1.0x10^4$ fibroblasts were plated per 60 mm organ culture dish (Corning) and reprogramming was started the following day by lipofection of the mRNA cocktail and incubation overnight. Transfections were repeated daily for three days and on day 9, distinct iPS colonies were forming.

2.4 Spontaneous differentiation

One day prior to the beginning of spontaneous differentiation, $5.0x10^5$ cells, which were dissociated using Accutase (Sigma), were transferred to one Matrigel-coated well of a 12-well plate with StemMACS iPS-Brew XF and 10 μM Y-27632 (R&D Systems). After 24 h, medium was replaced with medium containing 20% KnockOut Serum Replacement (KSR), 1% GlutaMAX, 1% non-essential amino acids (NEAA) and 0.1 mM beta-Mercaptoethanol (all ThermoFisher Scientific). Fresh medium was applied daily for up to 3 days.

2.5 Mesenchymal stem cell (MSCs), adipocyte and osteocyte differentiation

MSC differentiation was induced by exchanging StemMACS iPS-Brew XF medium with differentiation medium containing 20% KSR, 1% GlutaMAX, 1% NEAA and 0.1 mM beta-Mercaptoethanol supplemented with 10 μM SB431542 (Miltenyi Biotec). Fresh medium was applied every other day and after 7 days, cells were transferred in a 1:3 ratio to a non-coated tissue culture treated plate with MSC expansion medium (Miltenyi Biotec). Fresh medium was applied daily before splitting the cells at differentiation day 14. Process control of MSC differentiation was performed by flow cytometry and RT-qPCR on day 21. On day 21, MSCs were differentiated to adipocytes or osteocytes using StemMACS AdipoDiff Media or StemMACS OsteoDiff Media (both Miltenyi Biotec), respectively. Fresh medium was applied every 3 days for 20 days before process control by OilRed O or Alizarin Red staining, respectively.

2.6 Cardiomyocyte differentiation

Cardiomyocytes were generated according to a published protocol [172]. Briefly, 1.0×10^6 cells were dissociated as single cells using Accutase and plated in a well of a 12-well plate with StemMACS iPS-Brew and differentiation was induced the following day by changing the medium to RPMI-1640 (Sigma) with 2% B-27 supplement without Insulin (ThermoFisher Scientific) and 10 μM CHIR99021(R&D Systems). Same medium was used the following day and on day 3, half of the medium was replaced with RPMI/B-27 without insulin supplemented with 10 μM IWP-2 (Santa Cruz Biotechnology). On day 5 and 7, RPMI/B-27, first without insulin and then with full B-27 (ThermoFisher Scientific), were used. Fresh medium was applied after 3 days and cultures beginning to contract around day 12 were used for experiments. Process control of lateral mesoderm markers was performed on day 3.

2.7 Nephron differentiation

The protocol for differentiation of nephrons was optimized based on a published protocol [36]. Starting with undifferentiated cell cultures of ~70% confluency, a medium containing RPMI-1640, 1% GlutaMAX and 2% B-27 supplement (basal medium), 10 μM CHIR99021 and 500 nM dorsomorphin (Tocris) was used. Fresh medium was applied every other day and from day 4 onwards, the basal medium was supplemented with 10 ng/ml of ACTIVIN A (R&D Systems). On day 7, basal medium was supplemented with 10 ng/ml FGF9 (R&D Systems) and at day 9, with 3 μM CHIR99021 in addition for 48 h. Afterwards, basal medium supplemented with FGF9 was applied daily until day 21. Process controls were performed on day 7 for intermediate mesoderm markers, on day 14 for nephron progenitor markers and on day 21 for nephron markers by RT-qPCR and immunostaining.

2.8 Definitive endoderm, lung progenitor and hepatocyte differentiation

The protocol for differentiation of definitive endoderm was based on a published protocol [173]. Briefly, hPSCs were dissociated using Accutase and 4×10^5 single cells were seeded in a Matrigel-coated 24-well in RPMI-1640 medium, supplemented with 2% B-27, 50 U/ml of penicillin/streptomycin (Pen/Strep; ThermoFisher Scientific), 100 ng/ml ACTIVIN A, 1 μM CHIR99021 and 10 μM Y-27632. Fresh medium was applied daily until day 6 without Y-27632, but with 0.25 mM sodium butyrate (Sigma) on the first day and 0.125 mM afterwards. Process controls were performed on day 6 by flow cytometry and RT-qPCR.

Subsequent differentiation towards lung progenitor cells was based on a published protocol [174]. Briefly, foregut endoderm was induced using day 6 definitive endoderm cells by DMEM/F-12 medium, supplemented with 1% GlutaMAX, 2% B-27, 1% N-2 (ThermoFisher Scientific), 50 U/ml Pen/Strep, 0.05 mg/ml of L-ascorbic acid (Sigma), 0.4 mM of monothioglycerol (Sigma) (basal medium), 2 μM dorsomorphin and 10 μM SB431542. Fresh medium was applied daily and on day 10, lung progenitor differentiation was induced by applying basal medium supplemented with 20 ng/ml recombinant human BMP4 (R&D Systems), 50 nM retinoic acid (Sigma) and 3 μM CHIR99021. Fresh medium was applied daily until differentiation day 15 when expression of *NKX2.1* was observed.

Hepatocyte differentiation was based on a published protocol [37]. Briefly, 1.5×10^5 definitive endoderm cells were dissociated with Accutase, transferred to a Matrigel-coated 24-well and treated by DMEM/F-12 with 10% KSR, 1% NEAA, 1% GlutaMAX and DMSO (Sigma) together with 10 μM Y-27632 and 100 ng/ml recombinant human hepatocyte growth factor (R&D Systems). Medium was changed daily without Y-27632 for 10 days and process controls were conducted by RT-qPCR and immunofluorescence.

2.9 Neuronal stem cell differentiation

The protocol for differentiation of neural stem cells (NSCs) was based on the generation of neurospheres [33]. Briefly, hESCs were harvested using a 2 mg/ml Collagenase IV solution (ThermoFisher Scientific) and resuspended in DMEM/F-12 medium supplemented with 20% KSR, 1% NEAA, 1% GlutaMAX, 10 μM SB431542, 5 μM dorsomorphin, 20 μM CHIR99021, 10 μM purmorphamine (Miltenyi Biotec) and 10 μM Y-27632, and plated on an ultra-low attachment 6-well plate (Corning). Fresh medium was applied without Y-27632. 48 h later, the basal medium was exchanged with N2B27-based medium containing a 1:1 mixture of DMEM-F-12 and Neurobasal A (ThermoFisher Scientific) with 0.5% N-2, 1% B-27 minus Vitamin A, 1% NEAA and 1% GlutaMAX, and the small molecules described above. At day 5, N2B27-based medium supplemented with 50 μg/ml L-ascorbic acid, SB431542 and dorsomorphin was applied. On day 7, the neurospheres were mechanically dissociated and plated on Matrigel-coated plates. 24 h before the

replating, the medium was supplemented additionally with 5 ng/ml bFGF (Peprotech). Plated neurospheres were maintained for 7 days using the same medium and on day 14, confluent neuroepithelial outgrowths were passaged in a 1:10 dilution using Collagenase IV. The NSC cultures were passaged every 7 days and maintained in N2B27 medium with SB431542, dorsomorphin and bFGF at same concentrations as above with medium change every other day. Process control of NSC differentiation was performed on day 21.

2.10 Astrocyte differentiation

The protocol of astrocyte differentiation was based on a published protocol [35]. Briefly, tissue culture-treated plates were coated for 2 h with 10 ng/ml laminin/poly-L-ornithine (Sigma) and day 21 NSCs were dissociated using Accutase and plated at a ratio of 2.8×10^5 cells per well of a 12-well plate with N2B27 medium supplemented with 20 ng/ml bFGF, 10 ng/ml BMP4 and 5 ng/ml CNTF (R&D Systems). On day 15, medium was supplemented with 10 ng/ml bFGF, 10 ng/ml EGF (Sigma) and 10 ng/ml Neuregulin (R&D Systems) and the cells were differentiated for additional 15 days and then analyzed.

2.11 Motor neuron differentiation

The protocol of motor neuron differentiation was based on a published protocol [34]. Briefly, plates were coated, first with 10 ng/ml laminin, poly-L-ornithine, collagen I and collagen IV (Sigma) for 1 h each and then with 10 ng/ml vitronectin (Peprotech) for 1 h. 10 ng/ml fibronectin (Sigma) instead of vitronectin was used for later passaging. 1.5×10^5 day 21 NSCs were seeded per well of a 12-well plate with N2B27 medium supplemented with 100 ng/ml SHH, 10 ng/ml BDNF, 10 ng/ml GDNF, 10 ng/ml IGF (all from R&D System) and 100 nM retinoic acid. After 15 days, the medium was supplemented with 0.1 µM y-secretase inhibitor XXI (Merck) and 0.1 µM cAMP (Sigma Aldrich). Cells were analyzed at day 75.

2.12 Cortical neuron differentiation

The protocol of cortical neuron differentiation was based on a previously published protocol [175], with minor modifications. Briefly, iPSCs were plated in a 1:1 mixture of DMEM/F-12 and Neurobasal A, 1% N-2, 2% B-27, 1% GlutaMAX, 1% NEAA, 1000 U/mL Pen/Strep, 5 µg/ml human insulin (ThermoFisher Scientific) and 0.1 mM β-mercaptoethanol with 10 µM SB431542 and 1 µM dorsomorphin, and fresh media was applied daily. At day 10, cells were dissociated with Accutase and plated on poly-L-ornithine (1:1000) and laminin (1:200) coated plates at 1:4 dilution with the same medium supplemented with 10 µM Y-27632. From the next day onwards, the cells were treated by medium without SB431542 and dorsomorphin. Cells

were passaged every six days. Process control for neural induction and cortical neuron progenitor differentiation was performed after 15 and 35 days.

2.13 Somatic cell lines

Somatic cell lines used in this study were GIBCO® Human Skeletal myoblasts that were cultured for two days in DMEM (ThermoFisher Scientific) together with 2% horse serum (ThermoFisher Scientific), which induced differentiation to myotubes. Additionally, primary human epidermal keratinocytes (ATCC® PCS-200011™), primary adult human dermal fibroblasts (ATCC® PCS201012™), primary human neonatal foreskin fibroblasts (ATCC® CRL-2522™) and primary human astrocytes (ScienCell™ Research Laboratories, #1800) were cultured according to provider`s instructions.

2.14 Derivation of primary murine mesenchymal stem cells

Cultures of murine MSCs were established from the femoral bone marrow of female FVB/N mice (Charles River Laboratories, Sulzbach, Germany) by aspiration from the marrow cavity with 1 ml ice-cold PBS and a 0.4 mm injection needle. A solution of single cells was produced by pipetting, filtering through a 70 μm cell strainer (BD) and 5 min centrifugation at 300 g. Cells were plated in 12 ml of DMEM/F-12 with 1g/l glucose, 10% MSC-qualified FBS (ThermoFisher Scientific), 1% GlutaMAX and 10 μM Y-27632 in T75 cell culture flasks. Cells were kept under hypoxic conditions (2% O_2, 5% CO_2) at 37°C in a humidified atmosphere. Non-adherent cells were depleted by exchanging the medium 2 and 4 h after initial plating, whereas later on, fresh medium was applied every 3.5 days. When reached approximately 80% confluency, cells were passaged in a 1:3 ratio using Accutase.

2.15 Derivation of primary murine astrocytes

Primary mouse astrocytes of the C56BL/6 P3 strain were derived from whole cortex preparations. The brain was washed with HBSS (Sigma) supplemented with 50 U/ml Pen/Strep and meninges and blood vessels were removed. The cortex was isolated and cut into smaller pieces, and further resuspended in 10 ml HBSS/Pen/Strep. The minced tissue was plated on poly-D-lysine-coated plates (40 μg/ml, 1 h incubation) in DMEM/F-12 supplemented with 10% FBS, 50 U/ml Pen/Strep, 10 ng/ml FGF2 and10 ng/ml EGF. Fresh medium was applied every other day until the culture became confluent.

2.16 Derivation of primary murine cardiomyocytes

Primary mouse cardiomyocytes cultures were prepared using the Primary Cardiomyocyte Isolation Kit (ThermoFisher Scientific) according to the manufacturer`s instructions.

2.17 Derivation of primary murine hepatocytes

The protocol of primary hepatocyte derivation was based on a published protocol [176]. Liver was obtained from 14-week old C56BL/6 mice and digested using 2 mg/ml collagenase IV solution (ThermoFisher Scientific) at 37ºC for 45 min. The digested tissue was plated in a 10 cm dish with Williams E medium (Sigma) supplemented with 5% FBS and mechanically dissociated. Then, cells were filtered using a 70 μm cell strainer and 6 ml cell suspension was layered on top of a Percoll (Sigma) gradient of 1.12 g/ml, 1.08 g/ml and 1.06 g/ml in PBS. Cells were centrifuged for 20 min at 800 g and washed with Williams E medium with 5% FBS. After another centrifugation at 300 g for 10 min, the cells were resuspended in Williams E medium with 5% FBS, 1% GlutaMAX, 50 U/ml Pen/Strep, 50 ng/ml EGF, 1 μg/ml Insulin, 10 μg/ml transferrin (Sigma), and 1.3 μg/ml of hydrocortisone (Sigma) and plated on 10 μg/ml rat tail collagen I (Sigma) coated plates with daily medium change.

2.18 Animal data

Mouse keeping was done at the central facilities at the Helmholtz Center Munich in accordance with the German animal welfare legislation and guidelines of the Society of Laboratory Animals (GV-SOLAS) and of the Federation of Laboratory Animal Science Associations (FELASA).

2.19 Oil Red O staining

Following adipocyte differentiation, cells were washed twice with PBS, fixed with 10% neutral buffered formalin (Sigma) for 45 min, then washed twice with tap water and fixed again with 2-propanol (Sigma) for 5 min. Filtered Oil Red O solution (1.8 mg/ml in 2-propanol; Sigma) was added to the cells and incubated for 10 min. After two washes with PBS, cells were counterstained with Mayer`s hematoxylin solution (Sigma) for 3 min, before two washes with tap water, addition of PBS and imaging with a phase-contrast microscope. All steps were performed at RT.

2.20 Alizarin Red staining

Following osteocyte differentiation, cells were washed twice with PBS and fixed with 10% neutral buffered formalin (Sigma) for 45 min. Next, cells were washed twice with tap water and incubated with filtered

alizarin red staining solution (20 mg/ml; Sigma) for 45 min. After 4 washes with de-ionized water, PBS was added to the cells and images were obtained with a phase-contrast microscope. All steps were performed at RT.

2.21 Immunofluorescence staining

Cells were grown on imaging slides (Ibidi), washed 3 times with PBS, fixed with 4% paraformaldehyde (Sigma) in PBS for 10 min, followed by 3 washes using PBS. After permeabilization using 0.5% Triton-X-100 (Sigma) in PBS at 4°C overnight and 3 washes with PBS, slides were blocked with 0.1% Triton-X-100 and 1% FBS in PBS for 1 h at room temperature. Incubation with primary antibodies was performed at 4°C overnight. After 3 washes with PBS, slides were incubated with the species-corresponding secondary antibodies (ThermoFisher Scientific) for 2 h at room temperature in the dark and washed 3 times with PBS afterwards. The samples were mounted with ProLong® Gold Antifade Reagent with DAPI (ThermoFisher Scientific) on a coverslip and imaged with an Axio Observer.Z1 inverted epifluorescence microscope (Zeiss) equipped with a 10x/0.3 Plan-NEOFLUAR objective (Zeiss). Primary antibodies were diluted 1:100 unless stated otherwise and secondary antibodies 1:1000 in blocking buffer. Primary antibodies that were used in this study are listed in **Supplementary Table 3**.

2.22 Single-molecule fluorescence *in situ* hybridization (smFISH)

Cells were plated on imaging slides (Ibidi), fixed with 4% paraformaldehyde, washed twice with PBS and permeabilized with 70% ethanol overnight at 4°C. After 2 washes with PBS and pre-hybridization solution (10% deionized formamide, 2x SSC), slides were incubated with 50 µl hybridization solution containing 2x SSC, 10% formamide, 50 µg competitor E.coli tRNA (Roche Diagnostics), 10% Dextran Sulfate (VWR), 2 mg/ml BSA (UltraPure; Life Technologies), 10 mM vanadyl-ribonucleoside complex (NEB) and 1 ng/µl smFISH probes) for 6 h at 37°C. Afterwards, slides were washed twice with pre-hybridization solution at 37°C, then twice with PBS with subsequent mounting with ProLong® Gold Antifade Reagent with DAPI. Slides were imaged after 12 hours when the mounting medium was fully cured on an Axio Observer.Z1 inverted epifluorescence microscope equipped with a 63x/1.4 Plan-APOCHROMAT objective (Zeiss).

Probe Designer software by Biosearch Technologies was used to design probes for hNEAT1 5` segment and mNEAT1 middle segment, both conjugated to Quasar®670 fluorescent dye. Sequences are listed in **Supplementary Table 4**. Probes for hNEAT1 middle segment, mNEAT1 5` segment and MALAT1 (all conjugated to Quasar®570) were pre-designed by Biosearch Technologies.

2.23 Chemicals used for DNA binding

Cells were treated either by 2 μM Actinomycin D (ThermoFisher Scientific), 100 μg/ml Hoechst 33342 (ThermoFisher Scientific), 50 μM α-Amanitin (Cayman Chemical) and 5 μM Mithramycin A (Abcam). Vincristine (Selleckchem), Etoposide (Selleckchem) and Flavopiridol (Biomol) were used at concentrations specified in **Fig. 20C**.

2.24 Image analysis for paraspeckle counting

The spot detection program *Airlocalize* [177] was used for paraspeckle quantification based on 3D image stacks with 6 μm depth as described previously [47]. The averaged number of paraspeckles was calculated from images containing 20-150 cells. 7 images were analyzed per condition and replicate.

2.25 Image analysis for *NEAT1_2* single-molecule counting

Quantification of single *NEAT1_2* molecules based on smFISH was done with the *Fiji* software. For every image, a maximum intensity z-projection was generated and subjected background subtraction using the rolling ball method with 5-pixel thickness. Afterwards, the remaining intracellular background was measured by five consecutive intensity measurements and subtracted from the image. A threshold was applied to mask remaining spots and the integrated intensity of those spots was analyzed. A typical image yielded 500-2500 *NEAT1_2* foci whose intensities were binned to generate a distribution histogram. The peak with the lowest intensity was defined as the intensity of a single *NEAT1_2* molecule, and further peaks were defined accordingly with 2, 3, or more copies of *NEAT1_2*. Next, the intensity of all detected spots was divided by the intensity value for a single *NEAT1_2* molecule to determine the total amount of *NEAT1_2* molecules, which was then normalized by the number of cells. Importantly, this analysis was performed only in cells that were treated with Actinomycin for 0.5 – 1h, dependent on the cell type and

before any RNA degradation was observed. The method was adapted from a previously published paper [92] and is depicted in **Fig. 9**.
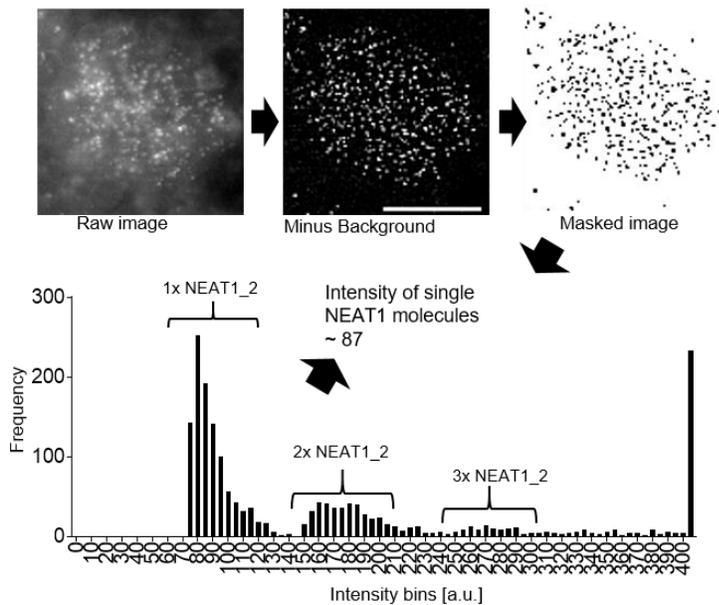
Figure 9: Image-based *NEAT1_2* single-molecule counting. A scheme and representative image and histogram displaying background subtraction and binning of *NEAT1_2* signal intensities (arbitrary units) representing the analysis of one image containing ~30 cells and the mean intensity of a single *NEAT1_2* molecule.

2.26 Quantification of nucleus size

Quantification of nucleus size based on DAPI staining was done using the *Fiji* software. Per image, an intensity threshold was determined to mask the DAPI staining in a maximum projection of a 3D image stack with 6 μm depth. The total DAPI area was divided by the number of cells per image to determine the average nucleus size per cell per image. The determination of nuclear size in single cells (**Fig. 17B**) was done by manually masking DAPI labeled nuclei and analyzing the nuclear area by the "Analyze Particles" function in *Fiji*.

2.27 Flow cytometry analysis

Surface marker staining was performed by washing dissociated cells with FACS buffer (1% FBS in PBS), centrifugation, removal of supernatant and incubation with primary antibodies in FACS buffer for 30 min on ice. Next, after centrifugation and removal of supernatant, cells were incubated with species-corresponding secondary antibody for 30 min on ice, before washing and final resuspension in FACS

buffer. A similar protocol was carried out with primary antibodies that were already conjugated to fluorophores.

Intracellular staining was performed according to instructions of the Inside Stain Kit (Miltenyi Biotec). Primary antibodies were incubated for 1 h at room temperature with 2.0×10^5 cells. Secondary antibodies were incubated for 30 minutes on ice. Cells were washed once with Inside Perm solution before resuspending them in FACS buffer for analysis.

Unconjugated primary antibodies were diluted 1:100 unless specified otherwise and secondary antibodies 1:1000 in FACS buffer. Samples were analyzed using the BD FACSAria III cell sorter (BD Biosciences) and data was processed using FlowJo software. Primary antibodies are listed in **Supplementary Table 3**.

2.28 SmFISH combined with flow cytometry

Based on smFISH signal, cells were sorted by flow cytometry according to a previously published protocol [178]. Briefly, 1×10^6 cells were harvested and resuspended in 500 µl of 4% paraformaldehyde and incubated at RT for 5 min. Fixed cells were centrifuged at 1000 g for 5 min, washed with 70% EtOH and resuspended in 70% EtOH with incubation overnight at +4°C. Next, the sample was split into half and resuspended in 100 µl RNA protection and hybridization buffer containing 300 mM NaCl, 30 mM sodium citrate, 2.1 M ammonium sulfate, 10 mM EDTA, 1 mg/ml *E. coli* tRNA, 500 µg/ml BSA, 25% formamide with or without 2 ng/µl *NEAT1* probe. After incubation at 30°C for 12 h, cells were washed twice with wash buffer containing 25% formamide and 2x SSC before resuspension in 2x SSC and cell sorting using the BD FACSAria III cell sorter into ice-cold buffer consisting of 200 mM NaCl, 20 mM sodium citrate, 1.5 M ammonium sulfate, 5 mM EDTA, pH 5.2 and 2x SSC. The top and lowest 5% of cells were sorted, centrifuged at 1000 g for 5 min and resuspended in 100 µl reverse crosslink buffer that contained 100 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA, 0.5 % SDS (v/v) supplemented with 500 µg/ml proteinase K (ThermoFisher Scientific), followed by incubation at 50°C for 1 h. RNA was isolated using the QIAzol lysis reagent (Qiagen) according to the manufacturer`s instructions.

2.29 RNA extraction and quantitative RT-PCR (RT-qPCR)

RNA extraction was performed using the RNeasy Mini Kit (Qiagen) according to the manufacturer`s instructions. Reverse transcription was performed using the Verso cDNA Synthesis Kit (ThermoFisher Scientific) with 200 ng RNA per reaction. RT-qPCR was performed in 384-well plates using 5 µl of SYBR Green PCR Master Mix (ThermoFisher Scientific), 1 µl cDNA and 1 µl of 5 µM primer forward and reverse

mix in a 10 µl reaction. PCR conditions were 2 min at 50°C and 10 min at 95°C followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. Relative expression levels were calculated using the Delta-Delta Ct method normalized with *GAPDH*. Statistical analysis was performed with the GraphPad Prism 7 software. RT-qPCR primers are listed in **Supplementary Table 5**.

2.30 Western blot

Cells were lysed in 50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1% Igepal CA-630 (Sigma I8896), 0.1% SDS, 0.5% sodium deoxycholate and 1x protease inhibitor (Roche). After 30 min incubation on ice, the lysate was centrifuged at 21000 g for 30 min. The supernatant was transferred to a new tube and 2x Laemmli buffer (Biorad) was added in a 1:1 ratio together with beta-mercaptoethanol and incubated at 95°C for 5 min. The protein sample was loaded on a Mini PROTEAN® TGX stain-free gel (Biorad) together with SDS running buffer containing 3% (wt/vol) Tris base, 14.4% (wt/vol) glycine and 1% (wt/vol) SDS. Subsequent blotting was performed for 1 h with 100 V on a nitrocellulose membrane in blotting buffer containing 25 mM TRIS, 192 mM glycine and 20% methanol. Next, the membrane was blocked with TBST buffer consisting of 20 mM TRIS, 150 mM NaCl, 0.1% Tween 20 and 5% non-fat dried milk powder (Sigma) for 45 min at RT. The primary antibody was added in a 1:1000 dilution in blocking buffer and incubated at 4°C ON. After three washes with 1x TBST buffer, the membrane was incubated with a 1:10000 dilution of secondary antibody conjugated to horse radish peroxidase (HRP) in blocking buffer for 2 h at RT. After three washes in TBST buffer for 15 minutes each, the membrane was stained with Clarity Western ECL Substrate (Biorad) and imaged after 2 min incubation using a ChemiDoc MP Imaging System (Biorad).

2.31 Generation of CRISPR-Cas9 modified hESCs

Genomic manipulations of hESCs were carried out according to a published protocol [179]. Briefly, the Protospacer Adjacent Motif (PAM) sequence was identified using the crispr.mit.edu website. *BbsI*-digested pSpCas9(BB)-2A-GFP vector (Addgene plasmid ID: 48138) was ligated with annealed forward/reverse guide RNA (gRNA) mix (1:250 dilution) using T4 ligase (NEB). NEB® 5-alpha competent *E.coli* bacteria (NEB) were inoculated with ligated plasmid and plated on agar plates. Bacteria colonies were propagated and plasmids were isolated using the GeneJET Plasmid MiniPrep kit (LifeTechnologies) according to manufacturer`s instructions. Sanger sequencing was used to screen for correct integrations. $1.0 \times 10^6$ hESCs were nucleofected with 5 µg of up- and downstream gRNA/Cas9 plasmid mix using the P3 Primary Cell 4D-Nucleofector® Kit (Lonza) according to the manufacturer`s instructions. Cells were plated 2 days later

and single clones were picked and analyzed for successful genomic deletion by PCR. Guide RNAs and primers for PCR-based screening are listed in **Supplementary Tables 6 and 7**.

Generation of *NEAT1^{YFP}* hESCs was done as previously described [95] by using plasmids provided by Addgene (IDs: 97088 for donor plasmid and 97082 for gRNA plasmid).

Generation of *NEAT1* hESCs with integrated stop cassette was performed in collaboration with Dong-Jiunn Jeffery Truong from the AG Westmeyer (Helmholtz Zentrum München, Institute of Developmental Genetics). Briefly, 4 µg of donor and 2 µg of CRISPR/Cas9 plasmid were transfected and selected for successful insertion with 0.5 µg/mL puromycin (ThermoFischer Scientific) for 7 days. Clones were picked and successful genomic insertion was confirmed by PCR. Plasmid sequences and exact cloning strategy are proprietary knowledge of the AG Westmeyer.

2.32 Generation of *SunTag* hESCs and transient *NEAT1* over-expression

Generation of hESCs expressing the *SunTag* complex under control of a doxycycline-inducible promoter was performed as previously described [180]. 1.0×10^6 hESCs were nucleofected with 2 µg of PB-pCAG-rtTA, SunTag PiggyBac and PBase vector provided by the authors of this publication. Cells were selected with 200 µg/ml Hygromycin B (Life Technologies) and 1 µg/ml doxycycline (Sigma) for 8 days before picking single green fluorescent clones, which were propagated and used for *NEAT1* gRNA transfections. *NEAT1* gRNA expressing vectors were generated as described in section 2.31. *SunTag* hESCs were transiently transfected with 5 µg of *NEAT1* gRNA vector and seeded with 1 µg/ml doxycycline and 10 µM Y-27632. Spontaneous differentiation was induced after 24 h by adding KSR differentiation medium and cells were analyzed after 48 h of differentiation. *NEAT1* gRNA sequences are listed in **Supplementary Table 6**.

2.33 DNA extraction and polymerase chain reaction (PCR)

Isolation of genomic DNA for screening of KO clones after transfection of CRISPR/Cas9 was performed using 30 µl QuickExtract™ (Biozym) according to the manufacturer`s instructions. PCR was performed using Q5 Polymerase master mix (NEB) with 100 ng DNA.

2.34 SiRNA and antisense oligonucleotide (ASO) transfection

H9 cells were transferred as single cells in a 24-well plate format and grown until 60% confluency was reached. Lipofectamine™ RNAiMAX Transfection Reagent (ThermoFisher Scientific) was used with minor deviation from the protocol suggested by the provider. Briefly, per reaction, 100 µl Opti-MEM®

(ThermoFisher Scientific) and 5 µl Lipofectamine® were added to a mix of 100 µl Opti-MEM® and 0.1 µM pre-designed Silencer® Select siRNA (ThermoFisher Scientific). After 15 min incubation at RT, this mix was added to the cells together with 200 µl iPS-Brew. Medium was changed after 24 h and cells were harvested after another 24 h of incubation. The following siRNAs were used in this study: siSFPQ: s12712 and sictr: Stealth RNAi™ siRNA Negative Control medium GC (ThermoFisher Scientific). Transfection of ASOs (**Supplementary Table 8**) for down-regulation of *NEAT1* expression was performed in a similar fashion. The final concentration of ASOs per reaction was 100 nM.

2.35 Subcellular fractionation

For subcellular fractionation, a modified protocol with the reagents from the Nuclear Complex Co-IP Kit (Active Motif) was employed. Cells were harvested, washed with 1 ml of 5% Phosphatase Inhibitors in PBS and centrifuged at 500 g for 5 min at 4°C. The cells were resuspended in 250 µl of 1x Hypotonic Buffer and incubated on ice for 15 min. After adding 12.5 µl detergent and mixing, the suspension was centrifuged for 30 s at 14000 g at 4°C. While the supernatant (cytoplasmic fraction) was transferred to a fresh microcentrifuge tube, the nuclear fraction was washed three times with 1x Hypotonic Buffer supplemented with 5% detergent to remove cytoplasmic contaminations. For both fractions, RNA was isolated by adding 1 ml QIAzol Lysis Reagent and heating the samples to 55°C for 10 min. 0.2 ml chloroform was added, the tubes were mixed vigorously and centrifuged at 12000 g for 15 min at 4°C. The upper, aqueous phase was transferred to a new tube, supplemented with 0.5 ml isopropanol, vortexed and incubated ON at -20°C. This was followed by centrifugation at 12000 g for 10 min at 4°C, before washing the RNA pellet with 1 ml of 70% EtOH and centrifugation at 12000 g for 5 minutes at 4°C. After removing the supernatant and briefly air-drying the pellet, the RNA was resuspended in 30 µl RNase-free water and cleaned up using the RNeasy MinElute Cleanup Kit (Qiagen).

2.36 Transcriptome analysis

Transcriptome analysis of *NONO*[-/-] hESCs was carried out by using the QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina (REV) with Custom Sequencing Primer (Lexogen) with 500 ng total RNA. Libraries were amplified and multiplexed with barcodes under the following conditions: 98°C 30 s, 14 cycles of 98°C for 10 s, 65°C for 20 s, 72°C for 30 s, and a final extension of 72°C for 1 min. Library preparation of SFPQ knock-down samples was prepared by following instructions of the TruSeq RNA Library Prep Kit v2 (Illumina). Quality control of the libraries was performed with a 2100 Bioanalyzer using the High Sensitivity DNA Kit (Agilent Technologies). Libraries were denatured with 0.1 N NaOH, diluted to a final

concentration of 6 pM and sequenced by a HiSeq2500 machine (*NONO*[-/-] hESCs) or a NextSeq500 machine (SFPQ knock-down hESCs). Data analysis was carried out on the Galaxy platform [181]. RNA Sequencing reads were mapped and aligned to the human hg19 reference genome using TopHat2 [182]. Transcript assembly and read counts were analyzed with Cufflinks and differential transcript expression was analyzed using Cuffdiff [183].

2.37 Nuclear Co-immunoprecipitation (Co-IP)

Nuclear Co-IP was performed by following the protocol of the Nuclear Co-IP kit (Active Motif). 10×10^6 cells were used per experiment, divided in a 1:1 ratio and one half was incubated with 5 μg SFPQ or 4 μg NONO antibody and the other half with the same amounts of corresponding IgG control antibody. Incubation was performed ON at +4°C on a rotating wheel. IP-High buffer condition without additional detergent and salt was used as precipitation and wash condition. After precipitation, the lysate was incubated with pre-washed 40 μl Protein A Dynabeads ™ and incubated for 1 h at RT. Subsequent washes of the beads were performed according to the instructions with additional three washes of PBS which was necessary before submitting samples to mass spectroscopy analysis. Finally, proteins were eluted by resuspending the beads in 20 μl 1x Laemmli buffer and incubation at 95°C for 5 min. Samples were either analyzed by Western Blot or subjected to mass spectrometry in collaboration with Dr. Stefanie Hauck (Helmholtz Zentrum München, Research unit Protein Science).

2.38 ChiP-SICAP

ChIP-SICAP is a protocol to pull down proteins and DNA associated with a target protein, which was established recently by Rafiee and colleagues [184]. This experiment was performed in the lab of Dr. Jernej Ule at the Francis Crick Institute with guidance by the first author of this publication. 10×10^6 cells were washed twice with PBS, harvested with Accutase, centrifuged at 1000 rpm for 5 min and fixed with 1.5% fresh formaldehyde solution for 14 min at RT. Glycine was added to reach a final concentration of 130 mM and incubated for 5 min. Cells were centrifuged at 2500 rpm for 2 min, washed once with PBS, centrifuged again and stored at -80°C until further processing. Cells were first washed with 10 mM TE buffer, pH 7.5 with 2x protease inhibitor (Roche), centrifuged at 1000 g for 2 min, resuspended in 900 μl TE buffer, incubated on ice for 10 min before addition of 100 μl of 10% Triton-X-100 and further incubation on ice for 5 min. Next, cells were centrifuged, washed twice with TE buffer and sonicated using a Covaris S220 with the following settings: time: 430s, Duty cycle: 10%, Intensity: 5, Cycle/Burst: 200. Afterwards, tubes were centrifuged at 12000 g for 10 min and 1% Triton-X-100 was added to the supernatant. Supernatant was

supplemented with 5 µg of Ser2-Phospho RNA PolII antibody or IgG control and incubated ON at +4°C/shaking. The following day, the chromatin was centrifuged at 12000 g for 10 min at 4°C and the supernatant was transferred to new tube. Magnetic Dynabeads conjugated with protein G (LifeTechnologies) were washed with IP buffer (1% Triton-X-100, 0.5% NP-40, 50 mM Tris-HCl; pH 7.5, 5 mM EDTA, 150 mM NaCl) and 30 µl Dynabeads were added per sample and incubated for 3 h at +4°C on a rotating wheel. Beads were washed with 100 mM Tris-HCl (pH 7.5), then washed with 100 µl 1x TdT buffer and finally resuspended in 93 µl TdT buffer with 5 µl dCTP-Biotin (1 mM stock, Jenabioscience) and 2 µl Terminal Deoxynucleotidyl Transferase (20 U/µl, ThermoFisher). Beads were incubated for 30 min at 37°C with 500 rpm agitation before washing them 6 times with ice-cold IP buffer at RT and resuspension and subsequent resuspension in 100 µl SDS 7.5% and 200 mM DTT and incubation for 30 min at 37°C. Supernatant was collected and diluted with 1000 µl IP buffer and 30 µl of Streptavidin magnetic beads were added to the wash tube. Beads were rotated for 1 h at RT before washing them 3 times with SDS wash buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 200 mM NaCl, 1% SDS). Then, beads were washed three times with 40% acetonitrile, transferred to PCR tubes and resuspended in 14 µl SDS 0.1% with 50 mM ammonium bicarbonate. 1 µl of 100 mM DTT was added and incubated for 20 min at 95°C. After cooling down the tubes, 1 µl Indole-3-acetic acid (IAA, 200 mM) was added and incubated 30 min in the dark. Another 1 µl DTT was added to neutralize the IAA, liquid was transferred to a new tube and proteins were digested with 200 ng Trypsin and 50 ng LysC at 37°C for 16 h. Peptides were cleaned up according to a previously published protocol [185] and analyzed by mass spectroscopy.

3. Results

3.1 Nuclear lncRNAs in the pluripotency - differentiation transition

Embryonic stem cells express hundreds of lncRNAs with crucial gene-regulatory functions in stem cell maintenance [66], [186], however, to date there is no comprehensive analysis of lncRNA dynamics during human germ layer differentiation. To interrogate connections between lncRNAs and the regulation of early human cell differentiation, I analyzed progenitors of the human germ layers (**Fig. 10A**). Each germ layer was represented by an early progenitor population and a population of tissue progenitors that were produced by differentiation of human ESCs or iPSCs. I first established differentiation protocols of lateral mesoderm and mesenchymal stem cells (MSCs), definitive endoderm and lung progenitor cells, and neural stem cells (NSCs) and cortical neuron progenitors, which represented respectively, early and late stages of differentiation of the three germ layers mesoderm, endoderm and ectoderm. I observed the up-regulation of lateral mesoderm markers *MESP1*, *T* (*Brachyury*), *FZD4* and *MIXL1*, and transcription factors *TWIST* and *SLUG*, which regulate epithelial-to-mesenchymal transition of MSCs, as well as surface markers that are characteristic for MSCs (**Fig. 10B, E**). When differentiated to definitive endoderm, I detected the up-regulation of *SOX17*, *FOXA2*, surface markers *CXCR4*, *CD117* and *EPCAM*, and later of the master lung transcription factor *NKX2.1* (**Fig. 10C, F-H**). Moreover, the up-regulation of *PAX6*, *SOX1*, *ASCL1*, *NESTIN* and *FOXG1* transcripts and proteins confirmed the differentiation to NSCs and cortical neuron progenitors, respectively (**Fig. 10D, I, J**). Finally, in all cell types, the pluripotency factors *OCT4*, *SOX2* and *NANOG* were down-regulated (**Fig. 10B-D**).
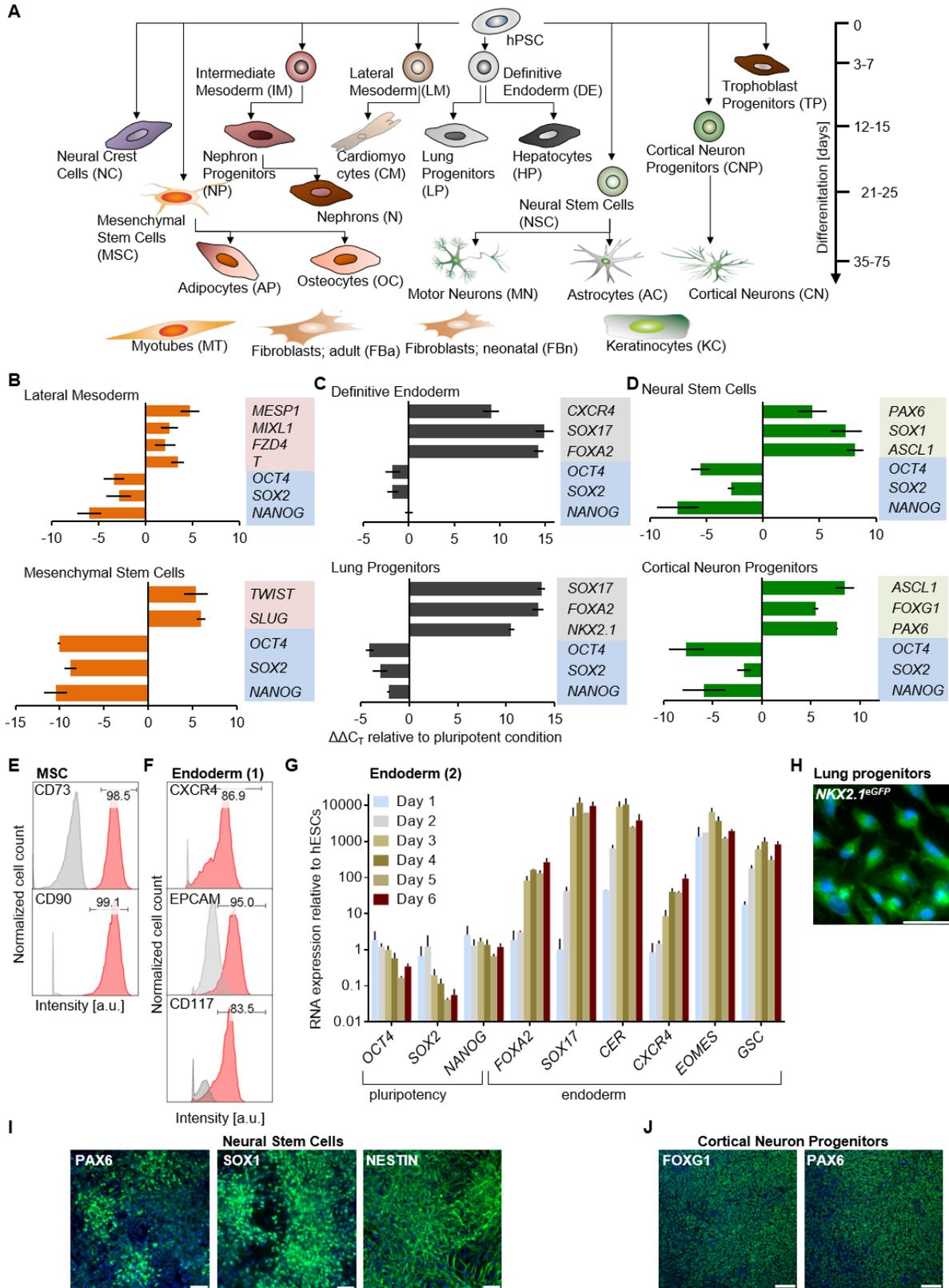
Figure 10: Characterization of germ layer progenitors. A) A scheme illustrating the cell types that were produced in this study by differentiation of hPSCs. Starting with undifferentiated cells at the top, hPSCs were differentiated to precursors of the germ layers, embryonic and extraembryonic progenitors, and terminally differentiated cells. The lineage and approximate developmental distance was estimated based on the expression of developmental markers as outlined below. In addition, primary preparations of keratinocytes, fibroblasts (adult and neonatal) and myotubes were analyzed. **B-D)** RT-qPCR analysis of lineage-selected markers corresponding to lateral mesoderm, and mesenchymal stem cells (MSCs) (**B**); definitive endoderm and lung progenitors (**C**); and neural progenitors and cortical neuron progenitors (**D**). Pluripotency genes *OCT4*, *SOX2* and *NANOG* were analyzed in all samples. **E)** Analysis of mesoderm differentiation towards MSCs showing the expression of characteristic markers *CD73* and *CD90* in accordance to [187]. **F-H)** Analysis of the differentiation towards definitive endoderm showing the up-regulation of *CXCR4*, *EPCAM* and *CD117* cell surface markers (**F**) and a cohort of characteristic markers as well as the down-regulation of pluripotency genes by RT-qPCR (**G**), and the expression of *eGFP* integrated in *NKX2.1*, which marks the formation of human lung progenitors [168] (**H**). Scale bar: 10 μm. **I, J)** Representative immunocytochemistry images of NSCs showing the expression of characteristic markers *PAX6*, *SOX1* and *NESTIN* on day 21 of NSC differentiation (**I**), and the cortical neuron progenitor markers *FOXG1* and *PAX6* (**J**) [175]. Scale bar: 50 μm.
n=2 independent experiments of cells in different passages. Error bars represent standard deviation.

Strikingly, the vast majority of lncRNAs, 24 out of 27, which were chosen for their participation in transcriptional and post-transcriptional regulation [186], were expressed in undifferentiated human ESCs and/or progenitor cells, and exhibited highly dynamic and diverse expression patterns including lineage-specific and temporal regulation (**Fig. 11**). A notable example for lineage-specific expression was the induction of *H19* (p<0.0001) in lung progenitor cells, and several lncRNAs that were up-regulated in all germ layers including *PINCR*, *LINC00472* and *NEAT1_2* (p<0.05 in ≥5 lineages). These findings indicated the involvement of nuclear lncRNAs in regulation of differentiation and diversification of human cell lineages. Because *NEAT1_2* was previously implicated in the regulation of development and differentiation [47], [113], I sought to further characterize its expression and function in the formation of the germ layers, tissue progenitors and differentiated cells.
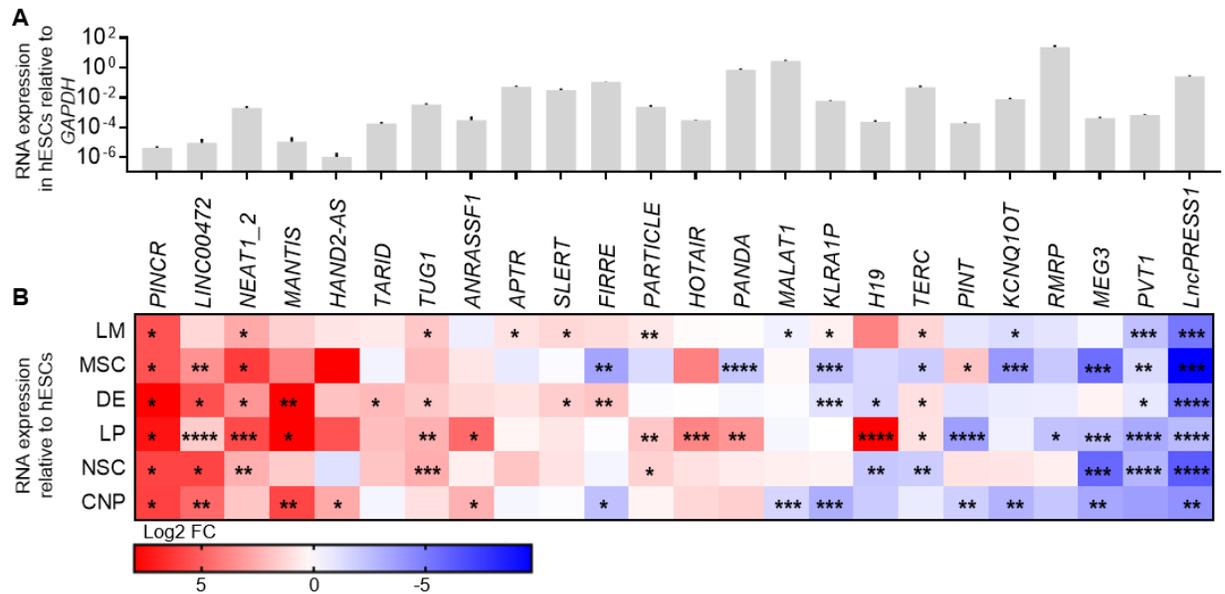
Figure 11: Analysis of developmentally regulated lncRNAs. A, B) The absolute (**A**) and relative (**B**) expression of nuclear lncRNAs in undifferentiated human ESCs and germ layer and tissue progenitors as in **Fig. 10B-D** based on RT-qPCR analysis. n=3 independent experiments, error bars represent standard deviation, Cells at different passages were used for replicates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, unpaired t-test. Abbreviations: LM: lateral mesoderm, MSC: mesenchymal stem cells, DE: definitive endoderm, LP: lung progenitors, CNP: cortical neuron progenitors, NSC: neural stem cells.

3.2 Analysis of *NEAT1* isoforms reveals dynamic regulation upon germ layer differentiation

3.2.1 Atlas of paraspeckle trajectories during cell fate conversions

To quantify paraspeckle dynamics during the differentiation of the human germ layers, I performed single-molecule FISH (smFISH) for analyzing foci of *NEAT1_2* in single cells. I expanded the repertoire of cell types by using numerous differentiation protocols as follows: mesoderm was represented by differentiating MSCs to adipocytes and osteocytes, lateral mesoderm to cardiomyocytes, and intermediate mesoderm to nephron progenitors and nephrons; definitive endoderm cells were differentiated to hepatocytes and lung progenitors; NSCs which are from ectoderm origin were differentiated into motor neurons and astrocytes, and cortical neuron progenitors were cultured to a mature state; neural crest progenitors, which represent the "fourth" germ layer, which give rise to multiple lineages that migrate throughout the body [188], were produced by a differentiation protocol that involves the formation of neurospheres [189]. Moreover, I differentiated trophoblast progenitors which represent extraembryonic tissues [170] and analyzed primary myoblasts, keratinocytes and fibroblasts from somatic sources (**Fig. 10A**).

Analysis of the expressed transcripts and proteins confirmed these lineages: differentiated MSCs exhibited lipid droplets and calcium deposits, which are observed in adipocytes and osteocytes, respectively (**Fig. 12A, B**); differentiation of lateral mesoderm progenitors led to up-regulation of cardiomyocyte progenitor markers including *NKX2.5* and *ISL1* (and cells began to spontaneously contract), while the precursor markers *T* and *MESP1* were down-regulated (**Fig. 12C**); and expression of *SIX2*, *PAX2 CDH5*, *WT1* and additional nephron progenitor markers were overtly apparent (**Fig. 12D, E, J**). In the direction of endoderm differentiation, liver markers *AFP*, *ALB*, *HNF4A* were strongly induced (**Fig. 12F, J**). Characterization of the neuronal cell populations was based on the formation of TUBB3 and NFH positive axons in the case of motor neurons, MAP2 positive axons in the case of cortical neurons, and GFAP positive star-like projections in the case of astrocytes (**Fig. 12J**). Moreover, these cell populations expressed the characteristic transcription factors *MNX1*, *ISL1*, *TBR1* and *SOX9*, respectively (**Fig. 12J**), which were confirmed by analysis of gene expression together with neuronal markers *CHAT* and *TBR2* as well as markers of astrocytes *SLC1A2* and *SLC1A3* (**Fig. 12G-I**). Finally, the identity of fibroblasts and keratinocytes was validated by expression of *VIM* / *HSP47* and *KRT14* / *IVL*, respectively (**Fig. 12J**).
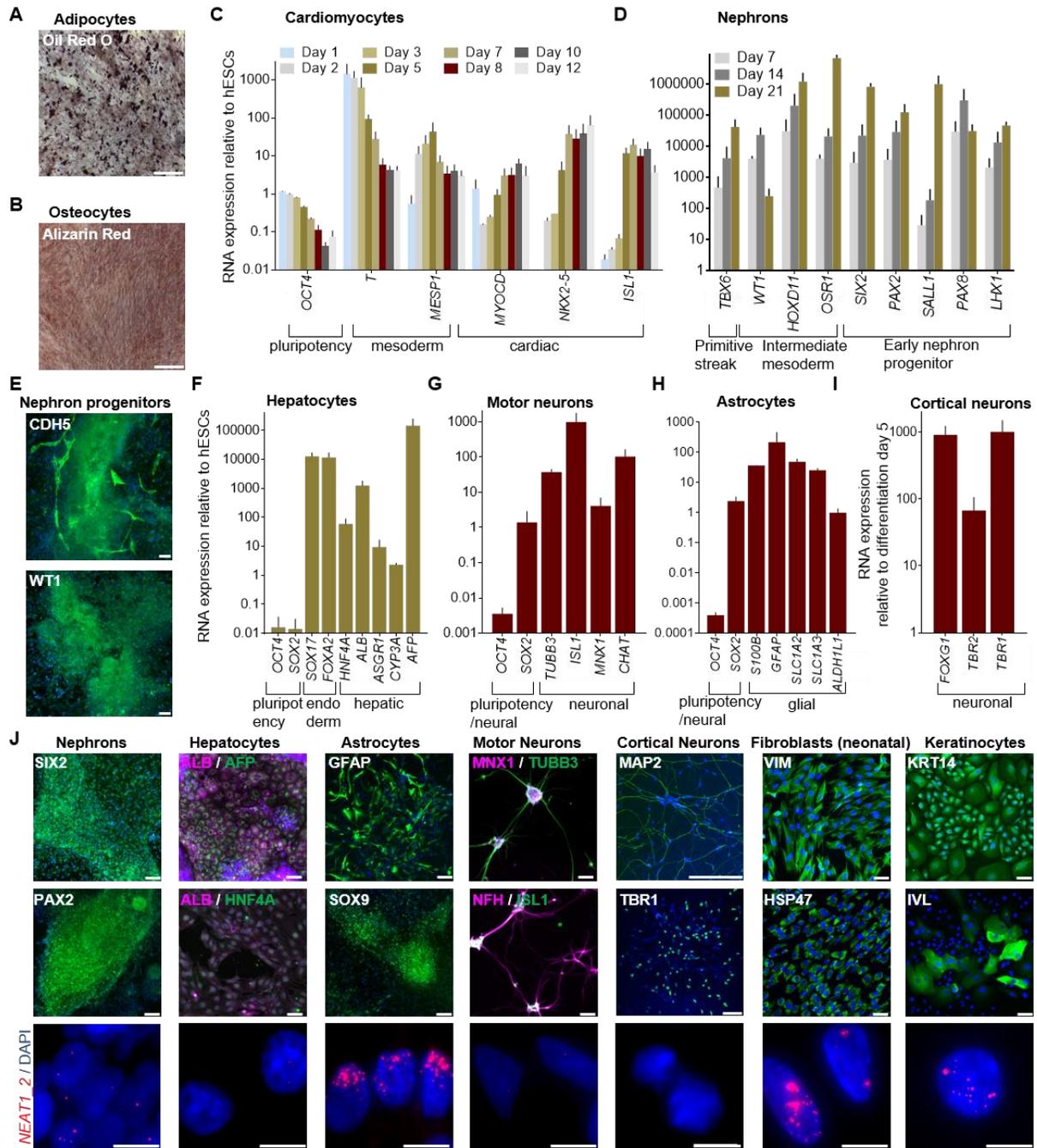
Figure 12: Characterization of differentiated cells. A, B) Oil Red O (**A**) and Alizarin Red (**B**) staining of human MSCs differentiated to adipocytes and osteocytes, respectively. Scale bar: 500 μm. **C, D)** Time course RT-qPCR analysis of representative pluripotency, mesoderm and cardiac markers during lateral mesoderm differentiation to cardiomyocytes (**C**) [172], and of representative intermediate mesoderm and nephron progenitor markers during nephron differentiation (**D**) [36]. **E)** Representative images showing the expression of characteristic nephron progenitor markers at day 14 of differentiation. **F)** RT-qPCR analysis of representative pluripotency, definitive endoderm and hepatocyte markers during differentiation to hepatocytes at day 16 [37]. **G-I)** RT-qPCR analysis of representative pluripotency, motor neuron, glial and cortical markers following differentiation to motor neurons (**G**), astrocytes (**H**) and

cortical neurons (**I**). **J)** Representative immunocytochemistry images of terminally differentiated cell preparations stained with antibodies, specific for markers of the respective cell types (scale bar upper panels: 50 µm) and analyzed by smFISH with *NEAT1_2* probe (bottom panel, probe in red, DAPI staining in blue; scale bar: 10 µm).
n=2 independent experiments, error bars represent standard deviation, cells in different passages were used for replicates.

Inspection of the atlas confirmed the previous observations that the number of paraspeckles increases when hPSCs exit the pluripotent state [47], [88] and it became apparent that this is a general phenomenon which includes precursors of the ectoderm and neural crest (**Fig. 13A-C, F**). However, the number of paraspeckles varied considerably between the different types of cells: differentiated cells of the mesoderm lineages generally exhibited a similar number of paraspeckles as their precursors, with the exception of MSCs and their adipocyte, osteocyte and fibroblast progeny, which exhibited the highest number of paraspeckles of all cell types (**Fig. 13A-C, F**). In the endoderm lineages, lung progenitors exhibited a similar number of paraspeckles as the definitive endoderm progenitor stage, and hepatocytes exhibited a smaller amount. In the neural lineages, all mature cells exhibited low number of paraspeckles, however, astrocytes, which were derived from NSCs in parallel to motor neurons, exhibited a very high amount of paraspeckles. Neural crest progenitors exhibited a similar number of paraspeckles as the other three types of germ layer precursors, but the number of paraspeckles in the extraembryonic lineage of trophoblasts was two times higher. Interestingly, adult dermal fibroblasts exhibited significantly more paraspeckles compared to fibroblasts of neonatal, foreskin origin (**Fig. 13C, F**). Finally, keratinocytes, which belong to the ectoderm, exhibited some of the highest numbers of paraspeckles (**Fig. 13F**). In general, there was only a minor correlation between paraspeckle number and differentiation time per cell type (**Fig. 13D**), whereas a stronger positive correlation with *NEAT1_2* expression was observed (**Fig. 13E**). I concluded that the quantity of paraspeckles is not overtly related to a specific germ layer or developmental stage, and that cells of mesenchymal and glial origin exhibited high amount of paraspeckles, whereas neuronal, hepatic and terminally differentiated cells of the mesoderm lineage exhibited little paraspeckle formation. Importantly, the individual cells within all populations exhibited heterogeneous amounts of paraspeckles (**Fig. 13G**), which confirms paraspeckle heterogeneity observed in tumor cell lines [190].
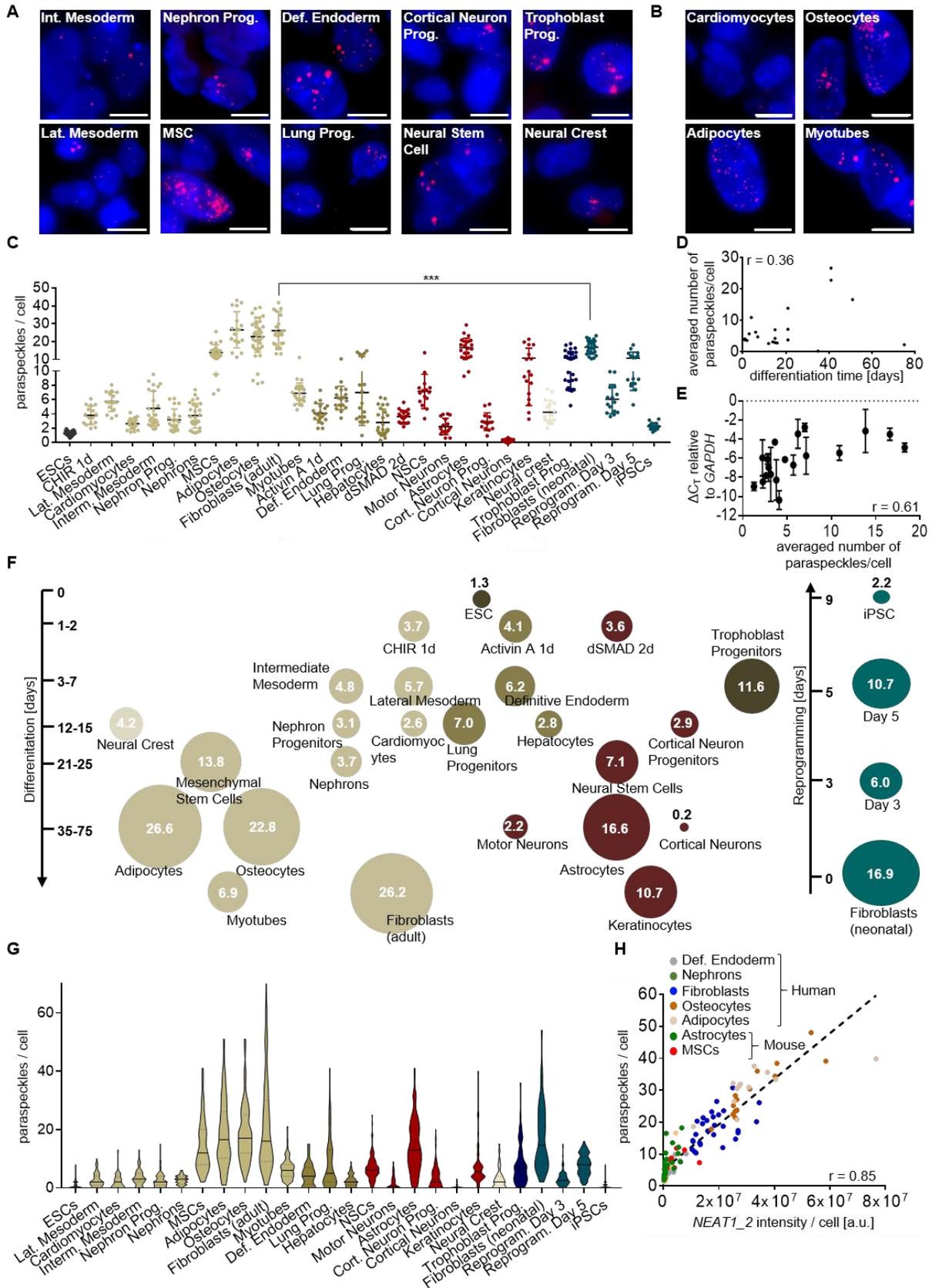
Figure 13: Analysis of paraspeckles in a panel of cell types and differentiated states. A, B) Representative images of *NEAT1_2* (red) in cells representing tissue progenitors (**A**), and terminally differentiated cells (**B**). **C)** The number of paraspeckles per cell in progenitors and differentiated cell types used to calculate the average number of paraspeckles in **F**. Each dot represents the average of one microscopic image displaying 20-150 cells. n=3 independent replicates using cells of different passages were analyzed with 5-7 images per replicate. Changes in number of paraspeckles are statistically significant for all cell types compared to ESCs (p < 0.0001, unpaired t-test, *** p < 0.001). **D)** Correlation of differentiation time as specified in the method section and averaged number of paraspeckles per cell type**. E)** RT-qPCR of *NEAT1_2* in 19 cell types and correlation with the averaged number of paraspeckles per cell indicated in **F**. RNA was obtained from 2-3 independent differentiations of cells in different passages. **F)** Summary of paraspeckle amounts in diverse developmental and terminally differentiated cell types, and during reprogramming of human neonatal fibroblasts. Size of circles corresponds to the average number of paraspeckles in the different cell types, which was quantified by automated spot (foci) detection in a total of 200 - 2000 cells per type representing 3 independent experiments **G)** Violin plots depicting the number of paraspeckles in 100 single cells from all tested human cell types based on **F**, black line represents mean value and dashed lines the quartiles. **H)** Analysis of the correlation of *NEAT1_2* total intensity and the number of paraspeckles per cell in representative human and mouse cell types. Each point represents a microscopic image. r in **D, E, H** is the Pearson`s correlation coefficient and dashed line in **H** is the linear regression line.

To evaluate the conversion of paraspeckle phenotypes in mammals, I analyzed mouse ESCs and primary mouse MSCs, cardiomyocytes, hepatocytes, and astrocytes, which represented the three germ layers. Similar to the equivalent cell types in the human, murine MSCs and astrocytes exhibited higher amounts of paraspeckles compared to cardiomyocytes and hepatocytes (**Fig. 14A, B**). However, the number of paraspeckles in murine MSCs and astrocytes was significantly lower compared to the human equivalent cell types (**Fig. 14A, B**). Finally, I verified that human astrocytes exhibited an exceptionally large amount of paraspeckles by analyzing primary cells of this type (**Fig. 14B**). Furthermore, I validated that the number of paraspeckles that I counted per cell was linearly correlated with the intensity of the *NEAT1_2* smFISH signal, which I used for measuring their amount (**Fig. 13H**). Taken together, I concluded that the tendency to exhibit higher or lower amounts of paraspeckles in specific cell types is evolutionarily conserved and that human cell types tend to have more paraspeckles compared to murine cells.

To corroborate the general conclusion that differentiation promotes an increase in the amount of paraspeckles, I analyzed paraspeckle kinetics during cellular reprogramming, which is the reverse process of differentiation. I utilized human fibroblasts, which exhibited some of the highest amounts of paraspeckles (**Fig. 13F**) and reprogrammed them into iPSCs by synthetic RNAs that encode the pluripotency factors *OCT4*, *SOX2*, *NANOG*, *KLF4*, *MYC*, and *LIN28A*. Analysis of paraspeckles revealed an oscillatory pattern as the amount declined and increased before declining again and settling on 1 - 2 paraspeckles per cell in the nascent iPSCs upon activation of the endogenous pluripotency genes (**Fig. 13C,**

F, Fig. 14C-E). This further confirmed the inverse correlation between pluripotency and paraspeckle formation.
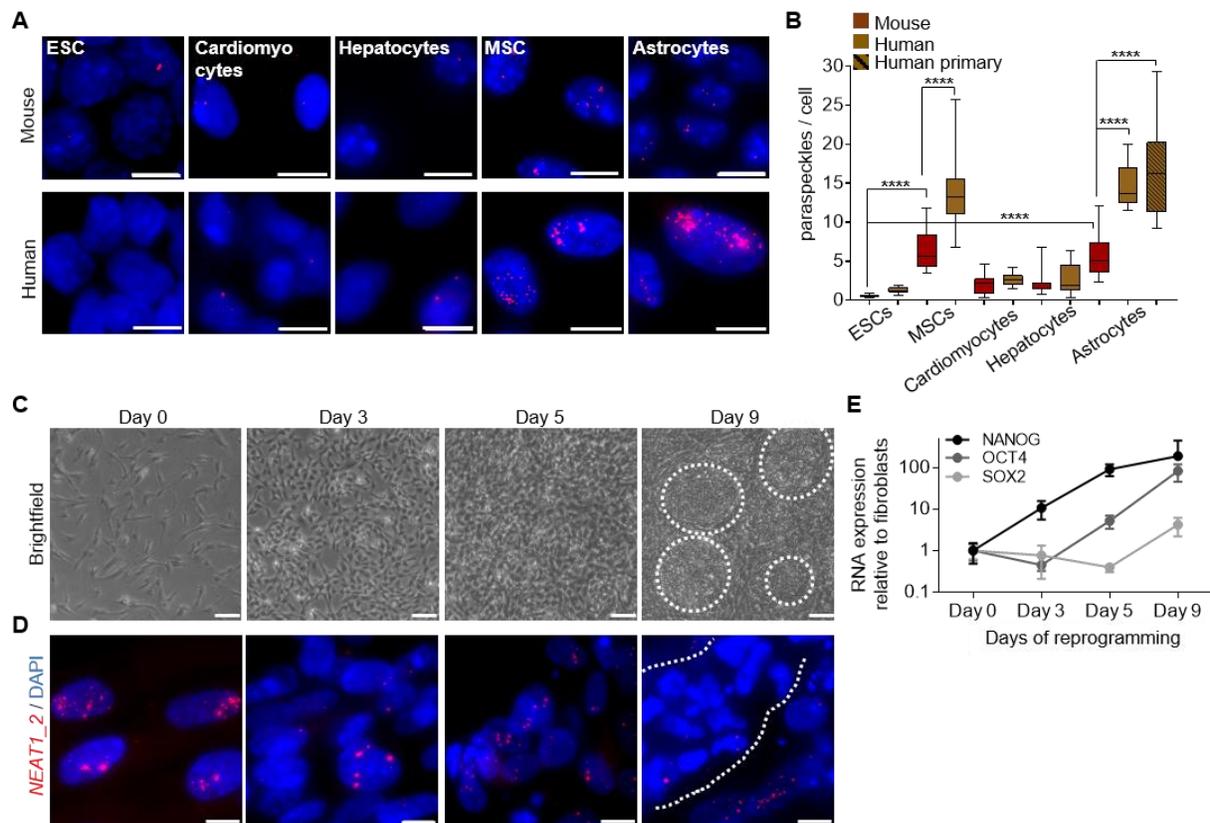
Figure 14: Characterization of paraspeckles in murine primary cell types and upon reprogramming of human fibroblasts. *A*) Representative images of *NEAT1_2* (red) in mouse ESCs and primary cardiomyocytes, hepatocytes, MSCs and astrocytes, next to the same cell types from the human. **B)** Quantification of paraspeckles in primary murine cell types (n=3 independent replicates using ESCs, or 3 different mice for the other cell types). **** p< 0.0001 unpaired t-test. **C, D)** Representative brightfield (**C**) and *NEAT1_2* (**D**) images taken during reprogramming of human neonatal fibroblasts. n=2 independent reprogramming experiments using cells of different passages were analyzed with 7 images per replicate; nascent iPSC colonies are marked with white circles. Quantification of paraspeckle numbers in **Fig. 13C, F**. **E)** Time-course RT-qPCR analysis of endogenous transcription of pluripotency factors *OCT4*, *SOX2* and *NANOG* during fibroblast reprogramming. n=2 independent reprogramming experiments.
Error bars represent standard deviation. DAPI staining in blue; scale bar is 10 µm in smFISH images and 50 µm in brightfield images.

3.2.2 Localization of *NEAT1_1* outside of paraspeckles is developmentally regulated

Both short and long *NEAT1* isoforms are generally confined to paraspeckles [191], however recently it was demonstrated that a subpopulation of *NEAT1_1* is localized outside of paraspeckles in the surrounding nucleoplasm in HEK293 cells where it primarily forms small foci containing 1-3 *NEAT1_1* molecules [95]. Supporting this notion, I observed *NEAT1_1* foci outside of paraspeckles in human and mouse ESCs,

however, the *NEAT1_1* foci disappeared upon prolonged differentiation (**Fig. 15A, B**). To interrogate general principles of *NEAT1_1* foci expression trajectories, I extended the paraspeckle atlas (**Fig. 13F**) by the analysis of *NEAT1_1* in those cell types and found that while most cell types were devoid of *NEAT1_1* foci, myotubes, neural crest cells, cardiomyocytes and cells committed to germ layer differentiation for 1-2 days exhibited *NEAT1_1* localization outside of paraspeckles **(Fig 15C, D)**. Interestingly, the number of *NEAT1_1* foci outside of paraspeckles was inversely correlated with the number of mature paraspeckles (**Fig. 15E**), indicating that increased expression of *NEAT1_2* promotes trans-localization of *NEAT1_1* to paraspeckles. I concluded that the formation of *NEAT1_1* foci is developmentally regulated and anti-correlated to the number of paraspeckles in mature cell types.
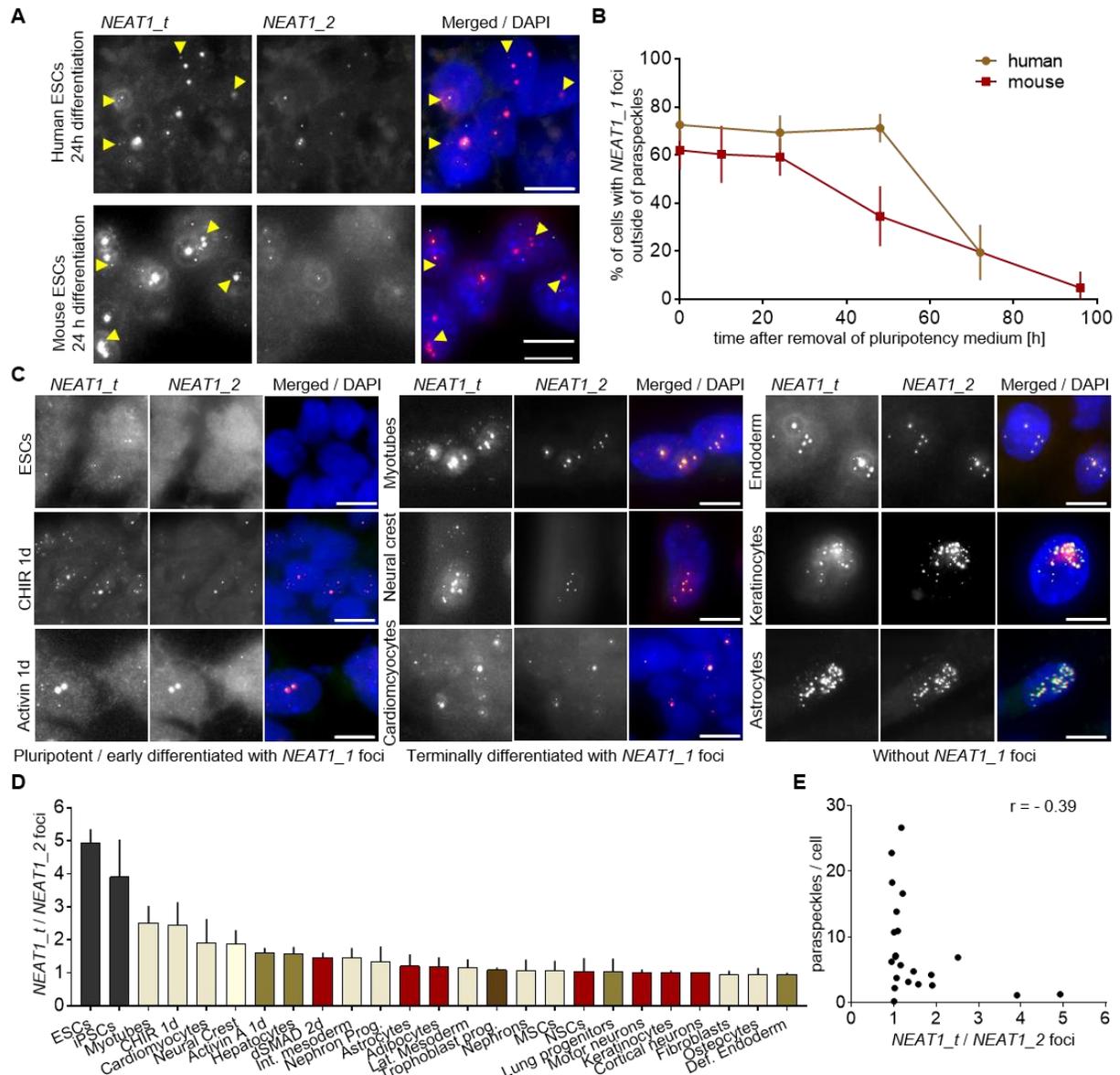
Figure 15: Characterization of *NEAT1_1* foci in ESCs and differentiated cells. A) Representative images of dual smFISH with probes for *NEAT1* 5`end targeting both *NEAT1* isoforms (*NEAT1_t(otal)*, red) and probes for the *NEAT1* middle segment targeting *NEAT1_2* (green) in mouse and human ESCs after 24 hours of differentiation. Arrowheads indicate *NEAT1_1* foci outside of *NEAT1_2*/paraspeckles. **B)** Quantification of *NEAT1_1* foci in mouse and human ESCs during spontaneous differentiation. n=6 images per time point. **C)** Representative images of cell types with or without *NEAT1_1* foci. Imaging as in **A**. **D)** Ratio of number of *NEAT1_t* and *NEAT1_2*-only foci. A value of 1 indicates the absence of *NEAT1_1* foci. Cell types and image analysis as in **Fig. 13C**. **E)** Correlation of *NEAT1_t* / *NEAT1_2* ratio and *NEAT1_2* foci in all cell types. r represents Pearson`s correlation coefficient.
Error bars represent standard deviation. DAPI staining marks nuclei in blue; scale bar: 10 μm.

3.2.3 Paraspeckle amount correlates with nucleus size

The paraspeckle atlas revealed that drastic changes in the amount of paraspeckles took place during differentiation without overt patterns of cell lineages or timing. I therefore asked what other cellular features might correlate with the amount of paraspeckles based on the cell atlas database. I noticed first a strong positive correlation between the amount of paraspeckles and the size of nuclei for individual cells within neonatal and adult fibroblasts (**Fig. 16A, B**). This prompted me to investigate whether nuclei size is in general a predictive factor for paraspeckle quantity. Strikingly, analyzing nuclei size in all cell types of the atlas revealed lineage correlation with the number of paraspeckles (**Fig. 16C**). Moreover, I found that the oscillating pattern of paraspeckle formation during reprogramming was correlated to changes in average nuclear size (**Fig. 16D**). This led me to hypothesize that the differences in paraspeckle amount between human and mouse astrocytes and MSCs (**Fig. 14B**) could be also explained by nucleus size differences. Indeed, adjusting the number of paraspeckles to the size of the nucleus in mouse MSCs and astrocytes showed that the corrected values of paraspeckles are similar (**Fig. 16E, F**). Moreover, differences in paraspeckles numbers between neonatal and adult fibroblasts (**Fig. 13C, F**) could be explained by the nucleus size **(Fig. 16G, H).** This provided a cell-intrinsic explanation for the high degree of variability in the number of paraspeckles observed between cells of the same type and for different types of cells.

Next, to assess whether the size of the nucleus determines the amount of paraspeckles or vice versa, I analyzed *NEAT1$^{-/-}$* and *NEAT1ΔpA* hESCs (introduced in section 3.4), which are either devoid of paraspeckles or exhibited 2-fold increase in the amount of paraspeckles due to the deletion of the internal polyA site [47] (**Fig. 16I**). Analyzing the size of nuclei did not reveal differences between *NEAT1*-modified cell lines compared to wildtype (**Fig. 16J**), thus I concluded that it is the nucleus size that determines the amount of paraspeckles.
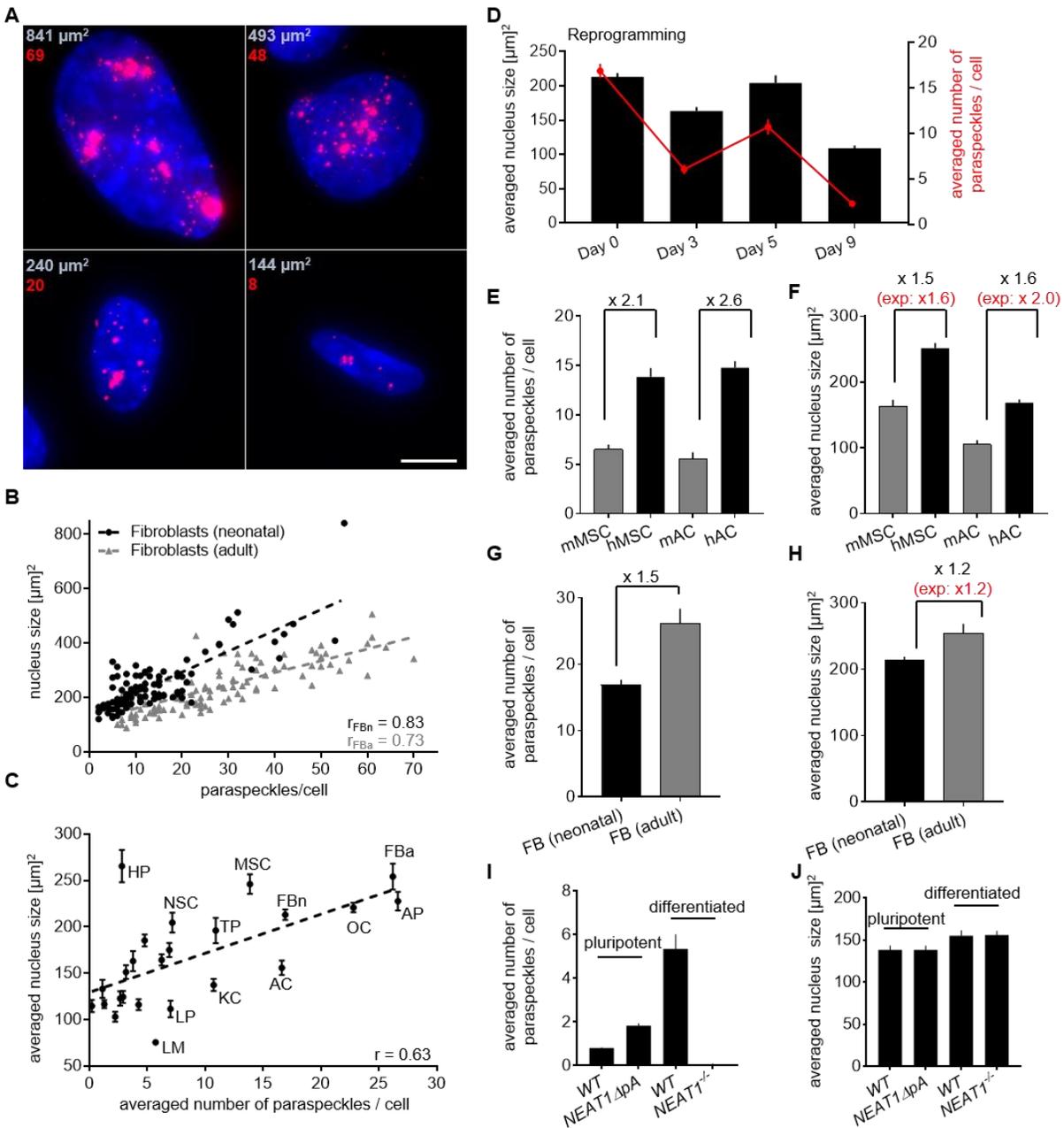
Figure 16: Paraspeckle formation correlates with the nuclear size within and across different cell types.
A) Images and quantification of nuclear area (μm²) by DAPI staining (blue) and number of *NEAT1_2* foci analyzed by smFISH (red) of representative human neonatal fibroblasts that exhibited different sizes (Scale bar: 10 μm). **B)** Analysis of the correlation between the number of paraspeckles and nucleus size of 100 human adult and neonatal fibroblasts. **C)** Analysis of the correlation between the averaged number of paraspeckles and averaged nucleus size per cell in 24 cell types analyzed in the atlas database represented in **Fig. 13F**. **D)** Averaged nucleus size (black) and number of paraspeckles (red; based on **Fig. 13F**) analyzed during reprogramming of human neonatal fibroblasts. **E-G)** Averaged number of paraspeckles per cell (**E, G**) based on **Fig. 13F** and averaged nuclear size (**F, H**) in mouse (grey) and human (black) MSCs and astrocytes (AC), as well as in adult (grey) and neonatal (black) fibroblasts. Numbers on top are the fold changes between the respective cell types from the human and the mouse. The numbers in red represent

predicted fold changes based on the slope of the regression line in **C. I, J)** Averaged number of paraspeckles per cell (**I**) and averaged nuclear size (**J**) of *NEAT1[-/-]*, *NEAT1ΔpA* and *WT* hESCs in pluripotent condition or differentiated by RA addition for 3 days.

Nucleus size in **C**, **F**, **H**, **J** represents the averaged value of 7-14 images per cell type from 2 independent experiments with 10-100 cells per image (details in methods). Error bars represent standard error of the mean (SEM). r in **B**, **C** represents the Pearson`s correlation coefficient and dashed line is the linear regression line.

3.2.4 TDP-43 regulates paraspeckle formation

Nuclear size seems to be a predictive factor for paraspeckle formation (**Fig. 16**), but the molecular mechanism regulating developmental paraspeckle formation remains unknown. The following findings were generated together with my colleague Miha Modic and were recently published, hence are described in this section only very briefly and the following figure was adapted from the publication [47].

It was shown by individual-nucleotide resolution Cross-Linking and Immunoprecipitation (iCLIP) that the multifunctional RBP TDP-43 binds to *NEAT1* [192], which prompted us to investigate its role in the regulation of *NEAT1* transcription. Indeed, we observed *TDP-43* down-regulation during exit from pluripotency, which was correlated with the expression of *NEAT1_2* (**Fig. 17A-C**). Strikingly, the number of h/mESCs with expression of *NEAT1_2* foci increased significantly upon depletion of *TDP-43*, similar to the deletion of the *NEAT1* polyA site, which inhibits *NEAT1_1* processing and promotes *NEAT1_2* expression (**Fig. 17D, E**). RNA sequencing confirmed the up-regulation of *NEAT1_2* in h/mESCs after down-regulation of *TDP-43* (**Fig. 17F**). Next, we performed iCLIP which showed enrichment of TDP-43 at UGUG repeats upstream of the *NEAT1* polyadenylation site (**Fig. 17G**). Importantly, when deleting this region, we observed an increase in the number of *NEAT1_2* foci (**Fig. 17H**), which led us to conclude that TDP-43 inhibits processing of *NEAT1_2* in pluripotent cells by binding and regulating the usage of its polyadenylation site. These findings were recently validated by others that linked paraspeckle formation to the depletion of nuclear TDP-43 in spinal cord neurons of ALS patients [128].
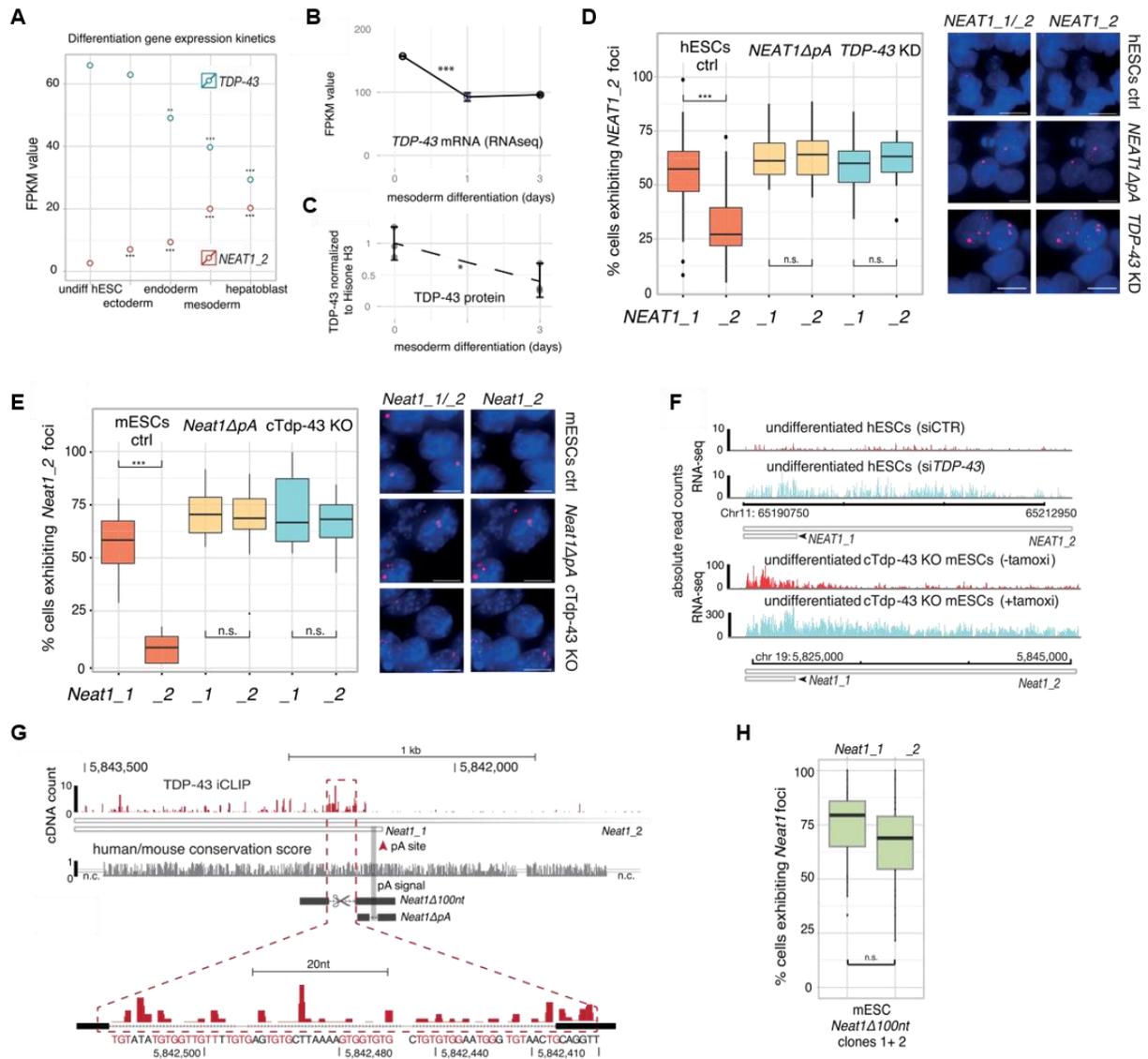
Figure 17: TDP-43 inhibits *NEAT1_2* processing by regulating *NEAT1* polyadenylation. **A)** Expression analysis of *NEAT1_2* and TDP-43 based on transcriptome data of hESCs and differentiated cells [193]. **B, C)** TDP-43 mRNA (**B**) and protein (**C**) expression during mesoderm differentiation induced by daily addition of 10 μM CHIR to hESCs. **D, E)** Percentage and representative images of human (**D**) and mouse (**E**) ESCs that are labeled with *NEAT1_2* probe (red). *NEAT1_1* and *NEAT1_2* foci were counted separately in 200 cells per condition. **F)** Representative mapping of *NEAT1/Neat1* RNA sequencing reads in murine and human ESCs and after depletion of T*DP43/Tdp-43*. **G)** TDP-43 cross-linked positions (red bars) based on iCLIP in mESCs. Sequence conservation is indicated as grey bars. Zoom into the region that was deleted for paraspeckle quantification in **H**. **H)** Quantification of *NEAT1* expression in mESCs with deletion of the TDP-43 binding site, as indicated in **G**. Cells were counted as in **D, E**. Mann-Whitney U test; **p < 0.001, ***p < 0.0001. Figure was adapted from [47].

3.3 DNA accessibility is required for paraspeckle assembly

The broad range in the amount of paraspeckles in cells with different sizes of nuclei led me to interrogate what common traits could regulate their structural similarity across different types of cells. Because *NEAT1_2* can form RNA-DNA triple helix structures, similarly to many other lncRNAs such as *HOTAIR* and *MALAT1* [111], I hypothesized that changing the accessibility of the major groove could perturb paraspeckle formation. Actinomycin D (ActD) is a small molecule that binds the minor groove of double-strand DNA (dsDNA) and changes its conformation [194], [195]. I therefore tested whether ActD treatment can promote the disassembly of paraspeckles. Strikingly, I noted a sharp increase and spreading of *NEAT1_2* foci in diverse types of cells that were treated by ActD for one hour, including trophoblast progenitors, NSCs and endoderm progenitors that were derived from hPSCs, as well as in primary astrocytes and neonatal fibroblasts (**Fig. 18A, D**). Importantly, the number of *NEAT1_2* foci was reflected by the number of paraspeckles in respective cell types of the atlas and peaked between 1 and 2 hours after ActD treatment (**Fig. 18F**). Contrarily, the paraspeckle core proteins SFPQ and NONO localized to perinucleolar caps after addition of ActD (**Fig. 18C**), which is in line with previous observations that reported perinucleolar localization of paraspeckle proteins after transcriptional inhibition [82] and during cell division [133] when *NEAT1_2* is down-regulated. This suggested that the *NEAT1_2* foci arising upon ActD treatment are not functional. Moreover, the intensity distribution of resulting *NEAT1_2* speckles indicated a high proportion of single *NEAT1_2* molecules after ActD treatment (**Fig. 9**), based on which I was able to estimate the amount of *NEAT1_2* molecules per cell. According to the *NEAT1_2* intensity profile after ActD addition, I estimated that, dependent on the cell type, cells contained on average between 20 and 50 *NEAT1_2* molecules (**Fig. 18E**), which is in the range of the 26 *NEAT1_2* molecules estimated by RT-qPCR in U2OS cells [95]. I concluded that treatment by ActD induced the disintegration of *NEAT1_2* foci to single molecules.

Disintegration of paraspeckles was observed before after 4 hours of ActD treatment in HeLa cells, however, the authors argued that this is due to ActD-mediated global inhibition of transcription [85]. To test whether this is true or alternatively, the disintegration of paraspeckles occurred as a result of the disruption of dsDNA, I treated trophoblast progenitors with α-Amanitin, which selectively inhibits RNA polymerases [196]. In parallel, the cells were treated by Hoechst 33342 and Mithramycin A which, similarly to ActD, are known to bind to the minor groove of dsDNA and disrupt its confirmation [197], [198]. Importantly, I observed similar kinetics of accumulation and decay of *NEAT1_2* foci upon Hoechst 33342 and Mithramycin A treatment as with ActD, but not following α-Amanitin, which ruled out inhibition of RNA polymerases as the mechanism of the immediate paraspeckle disintegration by ActD (**Fig. 18B, F**). Of

note is that all treatments including α-Amanitin led to *NEAT1_2* down-regulation after several hours of treatment (**Fig. 18A, B, F**), which is in agreement with previous observations and likely due to the decay of *NEAT1_2* [199].
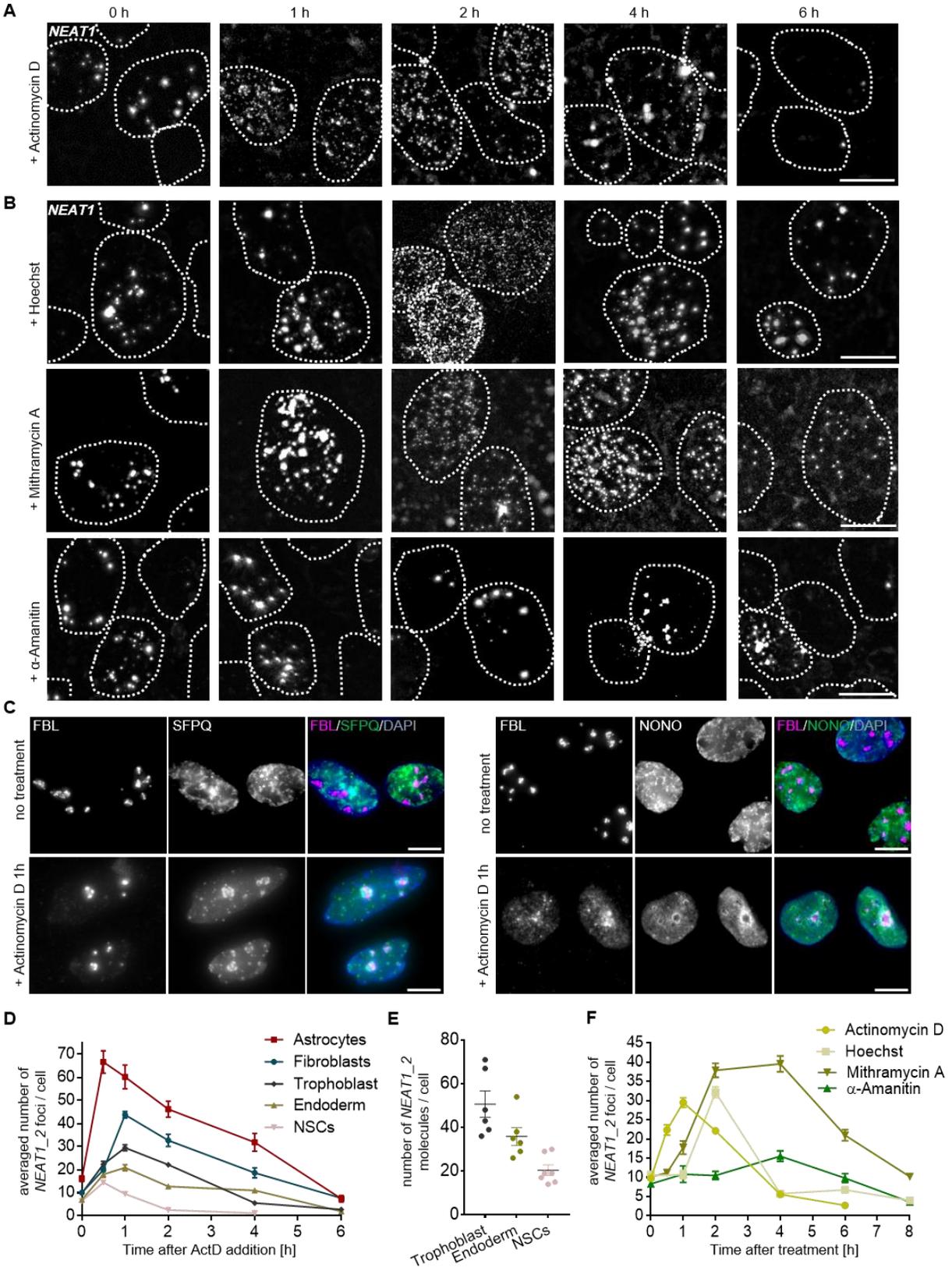
Figure 18: Treatment by DNA-binding small molecules promotes paraspeckle disassembly. **A, B)** Representative images of *NEAT1_2* smFISH after treatment of cells by 2 μM Actinomycin D (ActD) (**A**), 100 μg/ml Hoechst 33342, 5 μM Mithramycin A and 50 μM α-Amanitin (**B**) in trophoblast progenitors produced by 3 days of BMP4 treatment of hESCs. **C)** Immunocytochemistry of nucleolar protein fibrillarin (FBL) and paraspeckle proteins SFPQ and NONO in untreated trophoblast progenitors and after treatment by 2 μM ActD for 1 hour. **D)** Analysis of the averaged amount of *NEAT1_2* foci following ActD treatment in 5 different cell types **E)** The number of *NEAT1_2* molecules per cell, calculated based on the averaged intensity of a single *NEAT1_2* molecule (**Fig. 9**). **F)** Analysis of the averaged amount of *NEAT1_2* foci in trophoblast progenitors following treatment by the four chemicals shown in **A, B**. Error bars in **D**, **F** represent SEM and standard deviation in **E**. 7 images were analyzed in **E** and 14 in **D**, **F** representing 2 independent replicates using cells of different passages.

It is known that small DNA binding molecules can induce DNA double-strand breaks [200], which might be the underlying reason for paraspeckle disintegration. Indeed, after ActD treatment, I observed a significant increase in DNA double-strand breaks, which were analyzed by the appearance of γ-H2A.X foci [201], however, the treatment by Hoechst or Mithramycin A did not change the percentage of cells expressing γ-H2A.X (**Fig. 19A, B**) and thus I concluded that paraspeckle disintegration is not induced by DNA damage.

ActD and Mithramycin A are part of chemotherapeutic regimens to inhibit tumor growth [202] and I was interested whether paraspeckle disintegration could be induced by other chemotherapeutic reagents. I tested this by treating the cells with the microtubule inhibitor Vincristine [203], the DNA intercalator Doxorubicin [204], the topoisomerase II inhibitor Etoposide [205] or with Flavopiridol, an inhibitor for cyclin-dependent kinases [206]. Indeed, I observed a significant increase in *NEAT1_2* foci after treatment by Etoposide and Flavopiridol, but not by Vincristine (**Fig. 19C**) and since only the first two molecules were shown to bind dsDNA [207], [208], this supported the conclusion that DNA binding by small molecules induces paraspeckle disintegration (**Fig. 19E**). An exception was Doxorubicin, which, added to the cells, induced complete degradation of *NEAT1_2*, whereas *NEAT1_1* foci disintegrated similarly to *NEAT1_2* foci in the other treatments (**Fig. 19C, D**). Strikingly, this was a very concentration-sensitive effect, observed already by increasing the concentration of Doxorubicin from 4.3 μM to 5.7 μM. It seemed that Doxorubicin is detrimental for paraspeckle formation but possibly has a severe effect on *NEAT1_2* stability.

Lastly, to test whether the disintegration of nuclear lncRNAs is a general phenomenon, I analyzed *MALAT1* by smFISH after ActD treatment. Strikingly, I found that *MALAT1* speckles disintegrate with similar kinetics as paraspeckles (**Fig. 19F, G**). I concluded that dsDNA binding serves as structural basis for assembly and maintenance of paraspeckles and other nuclear lncRNAs.
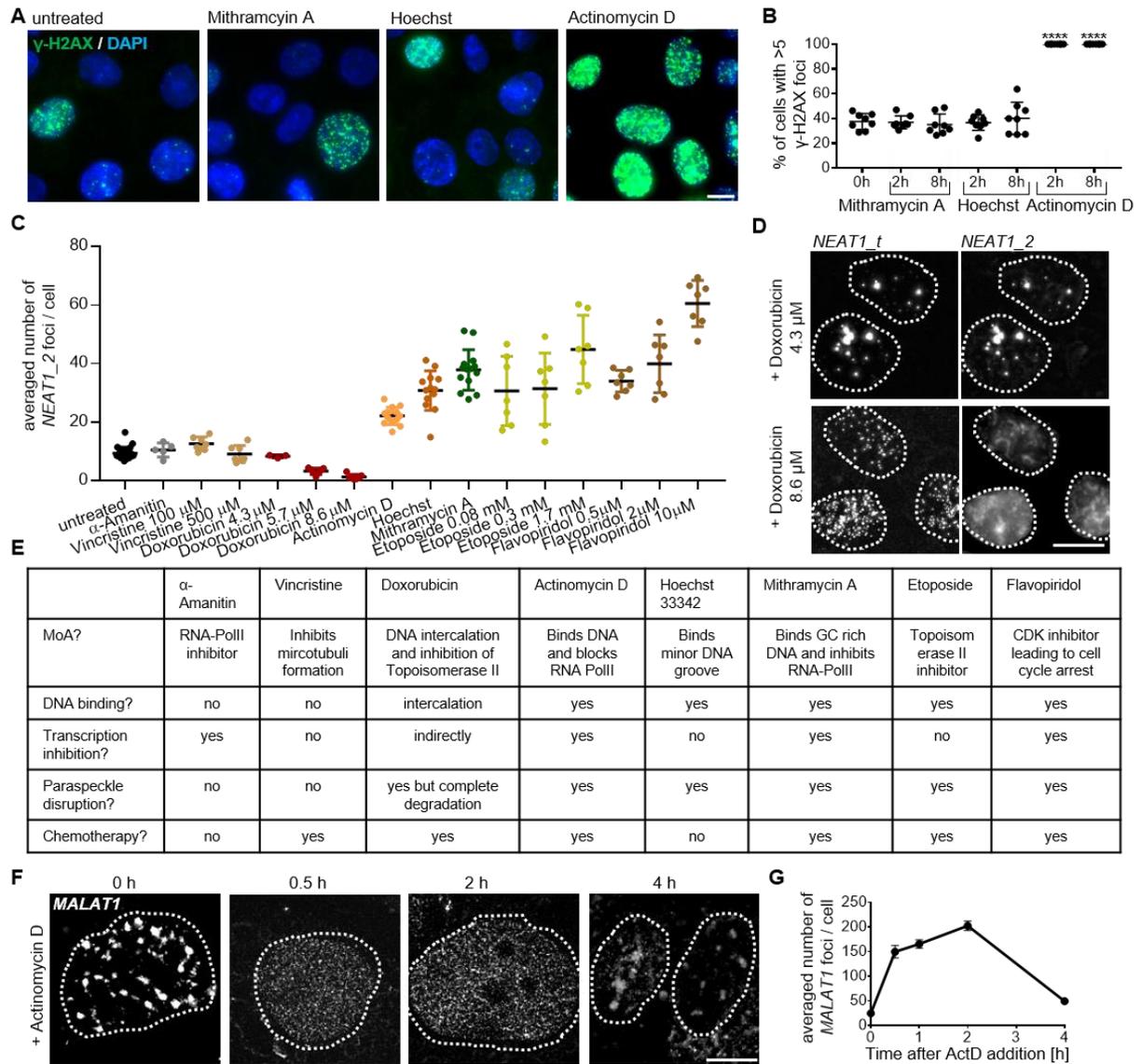
Figure 19: Chemotherapeutics disintegrate *NEAT1_2* foci. A, B) Representative images (**A**) and quantification (**B**) of γ-H2AX foci after addition of small DNA binding molecules. n=8 images. Note that already untreated hESCs exhibited γ-H2A.X foci, which were observed in mESCs, before [209]. **C)** Averaged amount of *NEAT1_2* foci following 2 h of treatment by the chemicals in **Fig. 18F** and different concentrations of the chemotherapeutic reagents Vincristine, Doxorubicin, Etoposide and Flavopiridol. **D)** SmFISH of *NEAT1* 5`end indicating the presence of both *NEAT1* isoforms and of *NEAT1_2* in trophoblast progenitors treated by different concentrations of Doxorubicin for 2 hours. **E)** A table with the mode of action (MoA) of small molecules used in this study and their ability to bind DNA, to inhibit transcription and to disintegrate paraspeckles. **F, G)** Representative images (**F**) and quantification (**G**) of *MALAT1* smFISH in cells treated by ActD as above. n=2 independent replicates with 7 images per replicate.

Cells were differentiated to trophoblast progenitors by addition of BMP4 for 3 days. Dashed lines in **D, F** depict the locations of the borders of the nuclei.

3.4 *NEAT1_2* but not *NEAT1_1* regulates exit from pluripotency

3.4.1 Manipulation of *NEAT1* expression reveals cell type-specific regulation of paraspeckle formation

Both isoforms of *NEAT1* exhibited drastic expression changes during germ layer differentiation (**Fig. 13, 15**), however, their function during this process remains unknown. To interrogate the role of *NEAT1* in differentiated cells, I generated hESCs with deletions of different functional parts of the *NEAT1* gene (**Fig. 20A**). To reduce *NEAT1* expression, I targeted its promoter by deleting 1150, 2300 and 6000 base pairs (bp) surrounding the *NEAT1* TSS. This led respectively to 10, 50 and 99% reduction of *NEAT1* expression (**Fig. 20B**) and a complete absence of paraspeckles for the biggest deletion (**Fig. 20C, D**), which therefore was referred to as *NEAT1^{-/-}* hESCs. Alternatively, I inserted a *polyA-YFP* stop cassette directly after the *NEAT1* TSS according to a previously published study in HEK293 cells [95]. Interestingly, this integration did not abrogate paraspeckle formation (**Fig. 20E**), indicating that *NEAT1* exhibits cell type-specific regulation and has an alternative TSS in hESCs. To test this, I deleted 1150 bp surrounding the *NEAT1* TSS in HEK293 cells and strikingly, whereas this deletion reduced *NEAT1* expression only by 10% in hESCs, paraspeckle formation and *NEAT1* expression was completely abolished in HEK293 cells (**Fig. 20B, F**). I concluded that *NEAT1* has one or multiple alternative TSSs in hESCs, but not in HEK293 and possibly other somatic cells. The *NEAT1Δ1150* HEK293 cells were used by collaborators to show that the $(G_4C_2)_n$ RNA arising from the C9ORF72 gene locus forms foci which co-localize with paraspeckle proteins, a process that is not influenced by *NEAT1* [210].

Additionally, I targeted the processing sites of *NEAT1_1* and *NEAT1_2*, by deleting the genomic region encoding for the polyadenylation signal or the triple helix, respectively. While a complete down-regulation of *NEAT1_2* and paraspeckles was observed in HAP1 cells upon deletion of the triple helix [211], I observed a reduction in *NEAT1_2* expression by 90% and number of paraspeckles by 50% in differentiated *NEAT1ΔTH* hESCs (**Fig. 20B-D**). In contrast, the deletion of the *NEAT1* polyA site resulted in the disappearance of *NEAT1_1* foci in undifferentiated cells (**Fig. 20G**) and consequently to increased number of paraspeckles due to elevated levels of *NEAT1_2* in pluripotent condition (**Fig. 20B, H**). Importantly, over-expression of paraspeckles was significant only in pluripotent condition, whereas differentiated *NEAT1ΔpA* hESCs did not exhibit substantially more paraspeckles compared to parental control cells (**Fig. 20H**). To summarize, I generated hESCs with reduced expression of short (*NEAT1ΔpA*), long (*NEAT1ΔTH*) or both *NEAT1* isoforms (*NEAT1^{-/-}*).
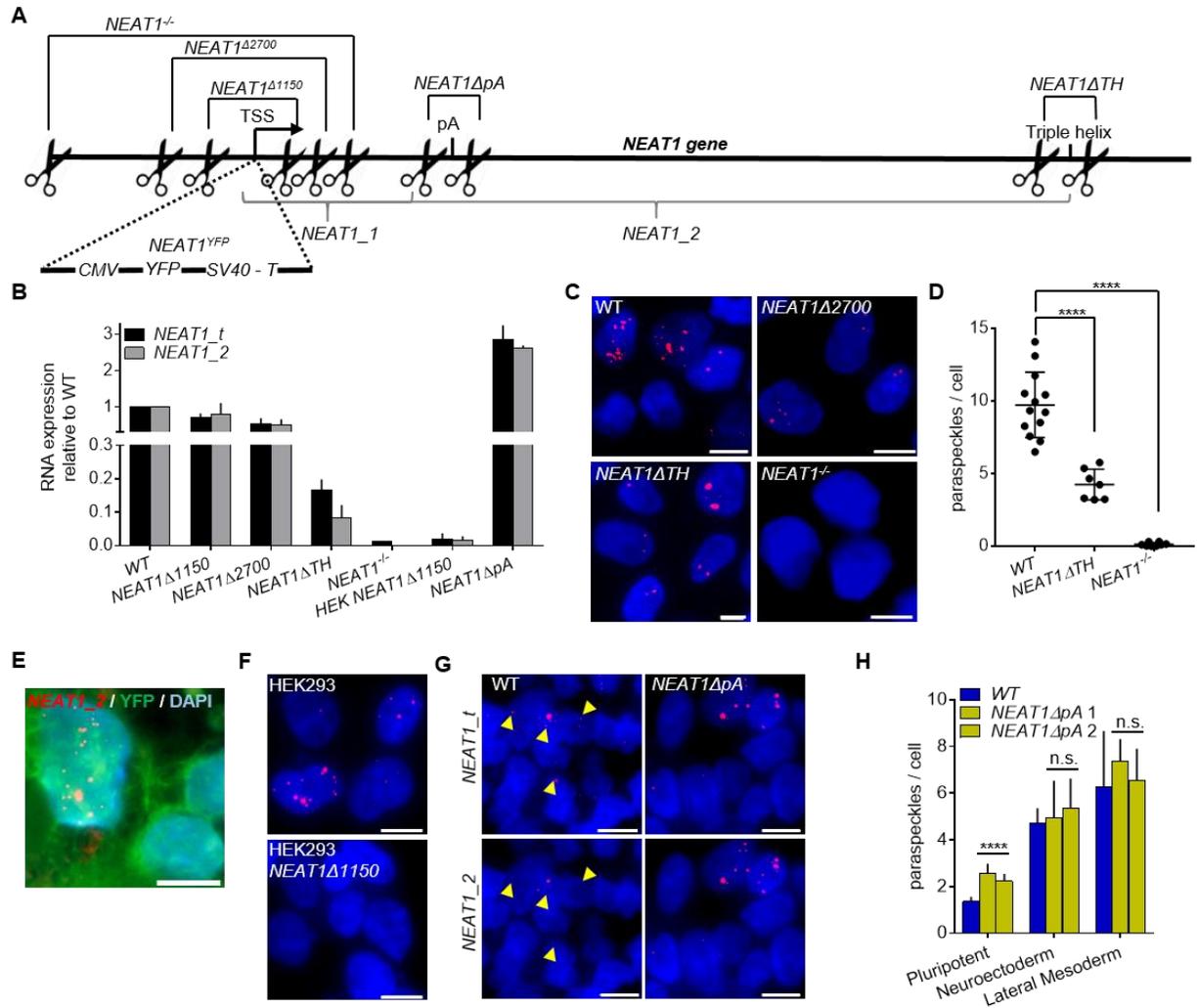
Figure 20: Generation of *NEAT1_1* and *NEAT1_2* knock-out cell lines. A) The strategy of manipulating *NEAT1* expression by CRISPR/Cas9. **B)** RT-qPCR of *NEAT1* in cell lines presented in **A**. n=3 experiments of cells in different passages. Cells were differentiated by removal of pluripotency medium for 3 days. *NEAT1ΔpA* hESCs were kept in pluripotency medium. **C, D)** Representative images (**C**) and quantification (**D**) of *NEAT1_2* foci (red) in *NEAT1Δ2700*, *NEAT1ΔTH* and *NEAT1-/-* hESCs differentiated for 3 days by addition of RA. **E)** Representative merged image of *NEAT1_2* smFISH in *NEAT1YFP* hESCs differentiated for 3 days with RA. **F)** Representative images of *NEAT1_2* foci in parental HEK293 cells and upon depletion of 1150 bp surrounding the *NEAT1* TSS. **G, H)** Images (**G**) and quantification (**H**) of paraspeckles in pluripotent *NEAT1ΔpA* hESCs. Yellow arrows mark *NEAT1_1* foci outside of paraspeckles.
Scale bar is 10 μm; DAPI in blue marks nuclei, *NEAT1_2* in red indicates paraspeckles. n =7-14 images in **D**, **H** representing 2 independent experiments. Error bars represent standard deviation. **** p<0.0001, unpaired t-test.

3.4.2 *NEAT1_2* inhibits spontaneous and neural differentiation

NEAT1$^{-/-}$ and *NEAT1ΔTH* hESCs represented cells with complete or 50% reduction of paraspeckles and hence were used to address whether *NEAT1_2*/paraspeckles are important for hESCs differentiation. By analyzing the expression of developmental markers, I found that the pluripotency characteristics of both cell types in pluripotent condition were intact, same as in wildtype with exception of premature up-regulation of *FOXA2* and *PAX6* **(Fig. 21G)**. Remarkably, the induction of spontaneous differentiation accelerated the down-regulation of pluripotency transcription factors *OCT4* and *NANOG* and cell surface markers *TRA1-60* and *SSEA5* in *NEAT1$^{-/-}$* and *NEAT1ΔTH* cells compared to the parental cell line after three days (**Fig. 21A-C**). A similar acceleration was observed during neuroectoderm differentiation after 4 days, but not prior, at the time that paraspeckles started to form **(Fig. 21D-F)**. Interestingly, when I used directed differentiation protocols that involved stimulatory cytokines, I did not observe overt acceleration of differentiation (**Fig. 21H-L**). I concluded that paraspeckles are functionally important in hESCs when they form during spontaneous or neural differentiation by slowing down the process, but the cells can compensate for their loss when treated by differentiation stimuli that can accelerate differentiation compared to normal development [212], [213].
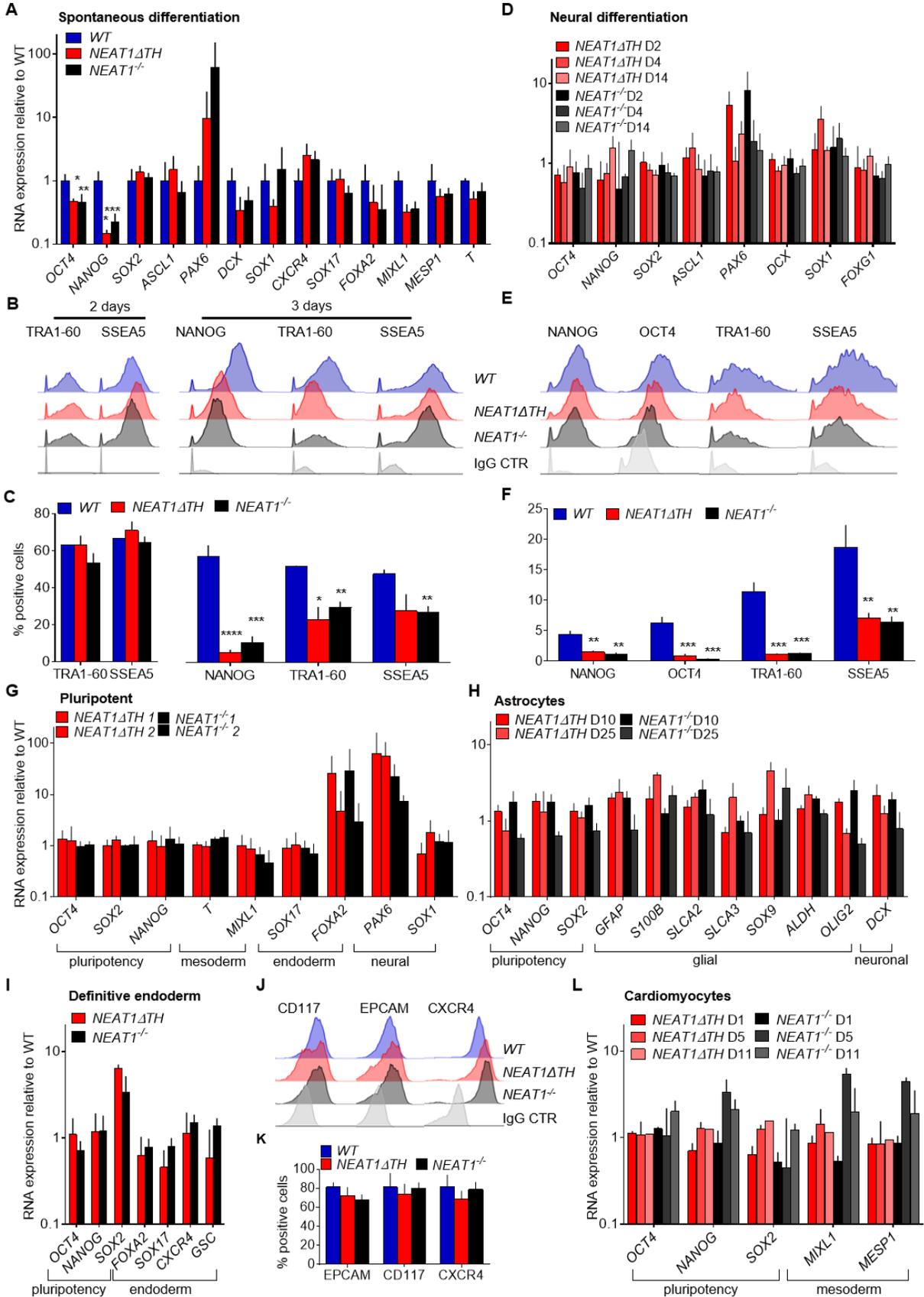
Figure 21: *NEAT1_2* **knock-out hESCs exhibit enhanced spontaneous and neural differentiation potential. A-F)** RT-qPCR of pluripotency and differentiation markers (**A, D**), and flow cytometry of pluripotency markers after *NEAT1ΔTH* and *NEAT1⁻/⁻* hESCs were spontaneously differentiated for 3 days (**A-C**) or after 4 days of neuroectoderm differentiation according to the protocol to generate NSCs (**D-F**). **G, H)** RT-qPCR of pluripotency and differentiation markers of undifferentiated (**G**) *NEAT1⁻/⁻* and *NEAT1ΔTH* hESC clones and cells differentiated to astrocytes (**H**). **I-K)** RT-qPCR of pluripotency and endoderm markers (**I**) and flow cytometry analysis of endoderm surface markers (**J, K**) after 6 days of endoderm differentiation. **L)** RT-qPCR of pluripotency and mesoderm markers during cardiomyocyte differentiation.
n=2 independent experiments using cells of different passages and with 3 knock-out clones per cell line. Forward and side scatter gating was employed to gate out debris and cell clumps. Error bars represent standard deviation, except in **C, F** where the SEM is shown. * p < 0.05, ** p < 0.01, *** p < 0.001, **** p< 0.0001, unpaired t-test.

3.4.3 Generation of *NEAT1⁻/⁻* hESCs by CRISPR knock-in confirms differentiation phenotype

To exclude potential confounding effects due to the genomic deletions in *NEAT1⁻/⁻* human ESCs, I sought to validate the phenotype observed above by an independent editing strategy. Therefore, I inserted a stop cassette (sequence is of proprietary knowledge of the Westmeyer lab; Helmholtz Zentrum Munich) to block transcription ~ 1500 bp after the *NEAT1* TSS. These *NEAT1^STOP* hESCs exhibited reduced expression of *NEAT1_2* and absence of paraspeckles (**Fig. 22A-C**). Of note is that *NEAT1_1* foci were still observed in *NEAT1^STOP* hESCs, indicating that 1500 bp of *NEAT1_1* is enough to form aggregates (**Fig. 22A**), even though these aggregates are somewhat more diffuse compared to full-length *NEAT1_1* foci (**Fig. 15**).

Importantly, integration of the stop cassette did not change the pluripotent character of the cells, as they expressed pluripotency and differentiation genes similar to the parental control cells (**Fig. 22D-F**). After spontaneously differentiating *NEAT1^STOP* hESCs for 3 days, a significant reduction of the pluripotency surface markers SSEA4 and TRA1-60, and the pluripotency master transcription factors OCT4 and NANOG was observed (**Fig. 22G, H**), which confirmed that *NEAT1* depletion promotes the exit from pluripotency. Interestingly, *NEAT1^STOP* hESCs exhibited up-regulation of neural transcription factors *PAX6* and *SOX1* indicating that *NEAT1^STOP* hESCs are primed for the neuroectoderm lineage (**Fig. 22I**). To test this, I induced neural differentiation and observed a significant up-regulation of *PAX6*, together with other neural markers, concomitant with down-regulation of *OCT4* and *NANOG* in *NEAT1^STOP* hESCs compared to the control cell line (**Fig. 22J**). I concluded that the knock-out of *NEAT1_2* promotes exit from pluripotency, possibly by priming neural differentiation.
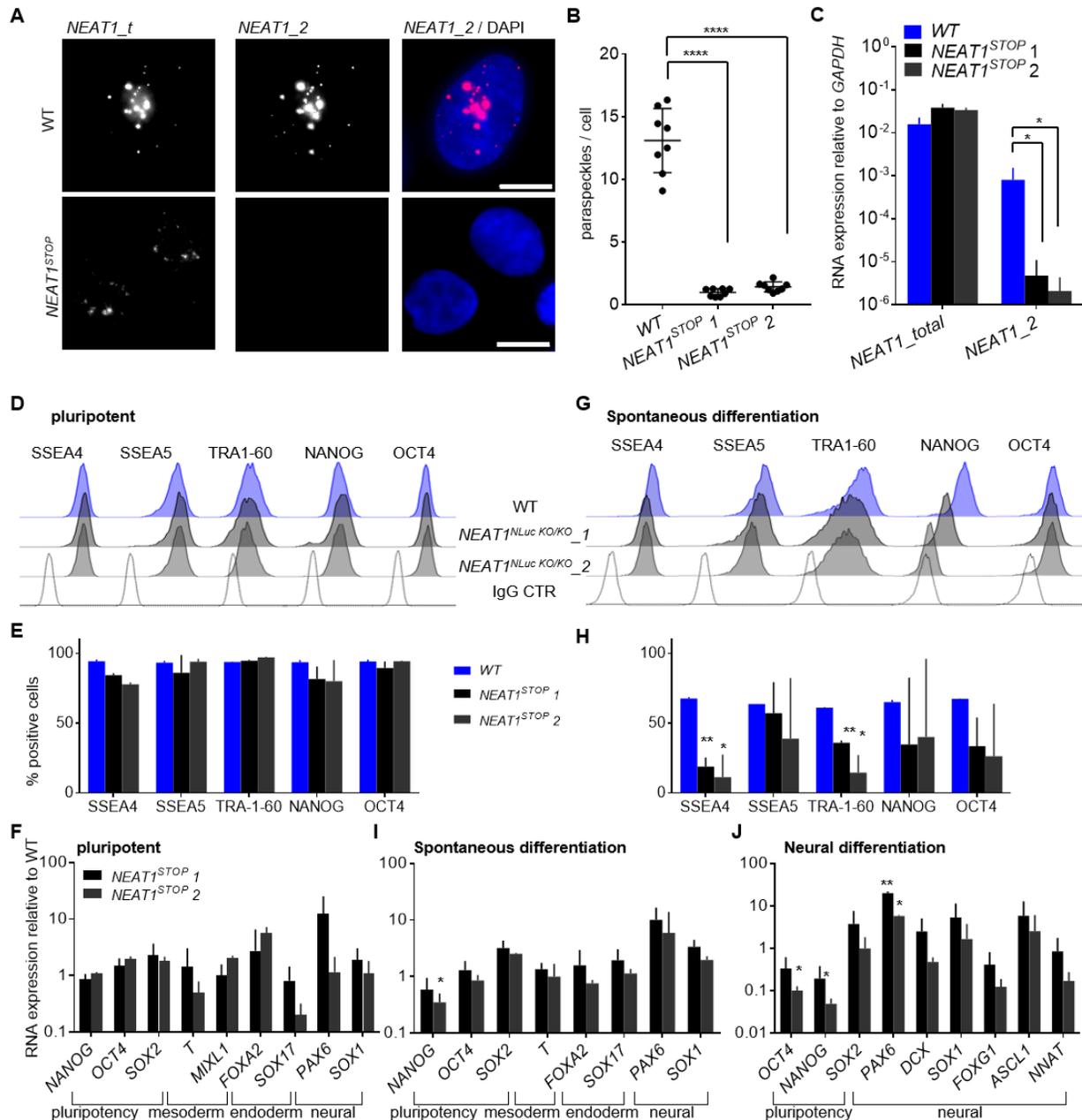
Figure 22: *NEAT1^{STOP}* **hESCs exhibited increased differentiation potential. A-C)** SmFISH (**A**) and RT-qPCR (**C**) of *NEAT1_t(otal)* and *NEAT1_2* (red), and quantification of paraspeckles in *NEAT1^{STOP}* hESCs, which were spontaneously differentiated (**C**) or treated by BMP4 for 3 days (**A, B**). DAPI marks nuclei, scale bar is 10 μm. 7 images were analyzed per condition. **D-J)** Flow cytometry analysis of pluripotency surface markers and transcription factors together with RT-qPCR of pluripotency and differentiation genes of *NEAT1^{STOP}* hESCs in pluripotent condition (**D-F**) and upon spontaneous differentiation for 3 days (**G-I**), together with RT-qPCR of neural genes (**J**) upon neuroectoderm differentiation for 4 days according to the protocol to generate NSCs.

n=2 independent experiments using cells of different passages. Forward and side scatter gating was employed to gate out debris and cell clumps. Error bars represent standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, unpaired t-test.

3.4.4 Over-expression of endogenous *NEAT1* impairs exit from pluripotency

Based on the results shown above, I hypothesized that *NEAT1* over-expression induces up-regulation of the pluripotency machinery. *NEAT1_2* is a very long RNA, hence it is hardly amenable for plasmid transfection-based over-expression, which is why I targeted it by CRISPR-mediated gene activation [214]. First, I generated stable hESCs expressing defective Cas9 (dCas) fused to the SunTag complex with the transcriptional activator VP64 [215] under control of a doxycycline-inducible promoter according to a previous study [180] (**Fig. 23A**). These cells, from here on referred to as *Suntag* hESCs, readily expressed the dCas9 construct upon doxycycline addition (**Fig. 23B**). Importantly, *Suntag* hESCs can be used to over-express any gene-of-interest, simply by transfecting a gRNA targeting the transcription start site. I screened a panel of five gRNAs, including some that were tested before [211], to achieve transient over-expression of *NEAT1_2* by 5-10 fold (**Fig. 23C**). This was concomitant with significant up-regulation of pluripotency factors *OCT4*, *NANOG*, *GDF3* and *NODAL* (**Fig. 23C**), the latter two being genes of the TGF-β pathway which are important to maintain pluripotency [216], [217]. The increase in pluripotency was confirmed by flow cytometry for pluripotency transcription and surface markers which were up-regulated compared to the *SunTag* cell line without *NEAT1* gRNAs (**Fig. 23D, E**). To summarize, I generated hESCs with stable expression of a dCas9-SunTag-VP64 trans-activator complex, which exhibited increased expression of pluripotency factors after transfection of *NEAT1* gRNAs, indicating that *NEAT1* over-expression impaired exit from pluripotency.
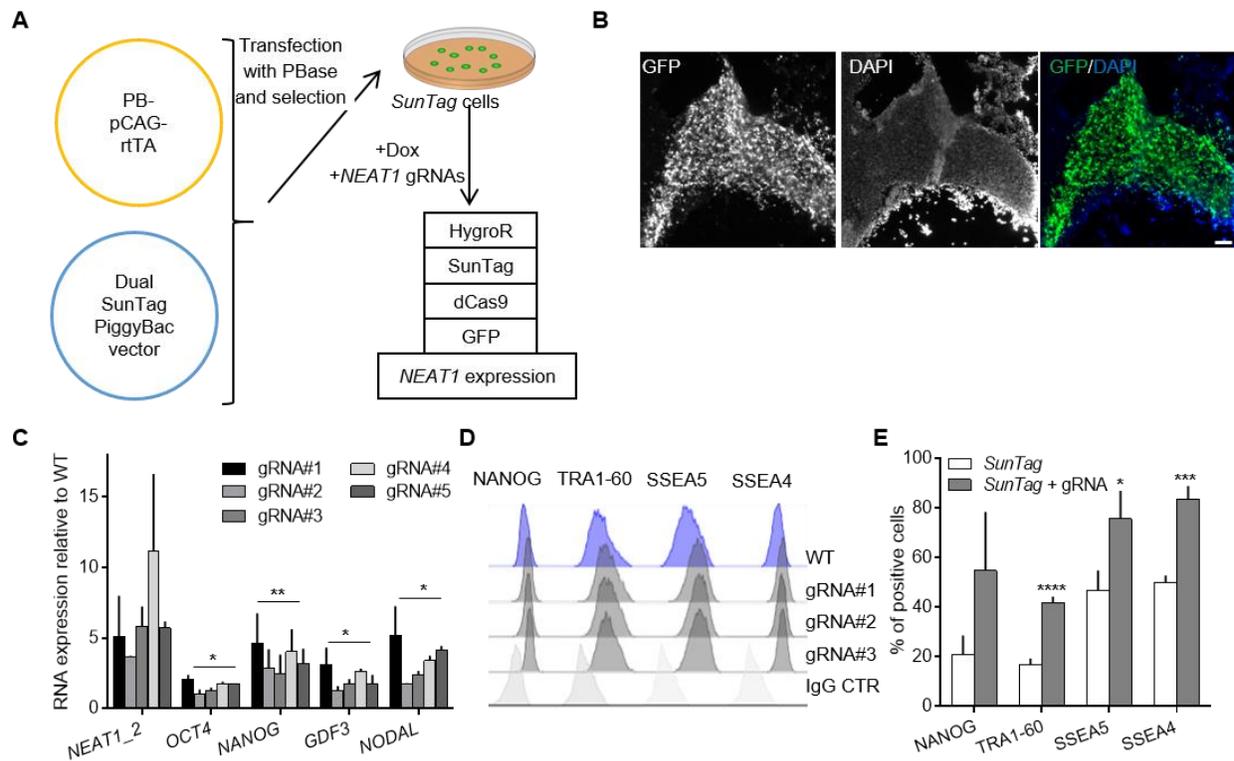
Figure 23: CRISPR-mediated activation of *NEAT1* induced pluripotency retention. A) Workflow for the generation of stable *SunTag*-expressing hESCs according to [180] **B)** *GFP* expression (green) indicating the production of the SunTag-VP64 transactivator complex in *SunTag* hESCs after doxycycline treatment. DAPI staining in blue; scale bar: 50 μm. **C-E)** RT-qPCR of *NEAT1_2* and pluripotency genes (**C**) and flow cytometry of pluripotency markers (**D**) in *SunTag* hESCs transfected with gRNAs targeting the *NEAT1* TSS. An average of three gRNAs is shown in **E**. Cells were induced by doxycycline for 3 days and spontaneously differentiated for 2 days. n=2 independent experiments using cells of different passages. Error bars represent standard deviation. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$, unpaired t-test.

3.4.5 *NEAT1* regulates *NANOG* expression

Gene expression analysis indicated that the expression of the pluripotency transcription factor *NANOG* was the most susceptible to changes in *NEAT1* expression, indicating that *NANOG* is regulated by paraspeckles. To validate this. I knocked-down *NEAT1_2* by transfection of *NEAT1*-directed antisense oligonucleotides (ASOs), which indeed prompted the down-regulation of *NANOG* (**Fig. 24A**), an effect that was slightly elevated by additional transfection of ASOs against *MALAT1*, the lncRNA that is in close proximity to *NEAT1* at the genome. Next, I established a protocol for flow cytometry-based cell sorting of *NEAT1*-expressing cells, stained by smFISH, according to a published protocol [178]. To test the efficacy of the protocol, I analyzed pluripotent and RA-treated hESCs and observed, as expected, a shift in *NEAT1* expression in the differentiated cell population (**Fig. 24B**). I sorted the top and lowest 5% of *NEAT1*-

expressing cells upon differentiation, which exhibited higher *NANOG* expression in the *NEAT1*-positive cell population (**Fig. 24C, D**). To conclude, *NANOG* expression correlated with *NEAT1* levels, which validated findings from above by two orthogonal methods.
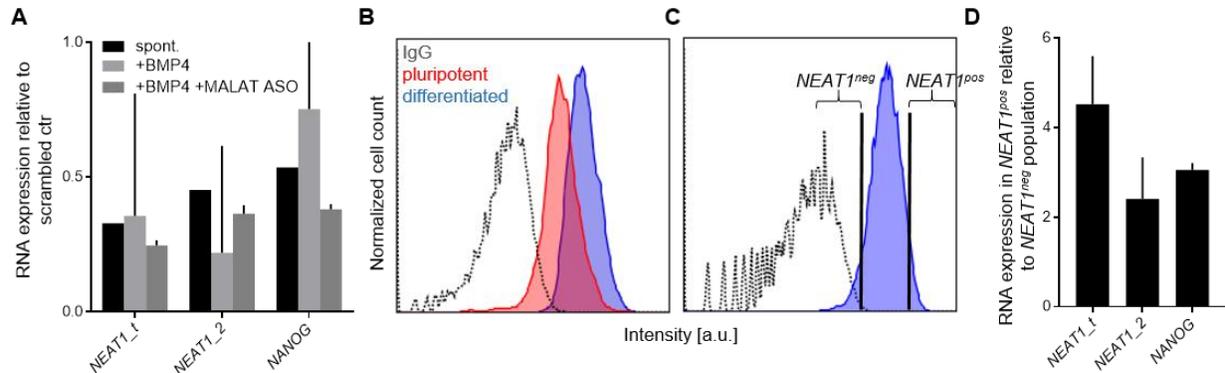
Figure 24: *NEAT1* regulates *NANOG* expression. A) RT-qPCR of hESCs transfected with ASOs for *NEAT1* and spontaneously differentiated for 2 days with or without the presence of BMP4 to induce trophoblast progenitor differentiation. For the latter, ASOs for *MALAT1* were added in a separate experiment. RNA expression was normalized to the scrambled control ASO. **B)** Flow cytometry-based analysis of single cells stained by smFISH for *NEAT1* in undifferentiated cells (red) and cells differentiated by addition of RA for 4 days (blue). **C, D)** Flow cytometry-based single-cell sorting of the top (*NEAT1^pos^*) and lowest (*NEAT1^neg^*) 5% of *NEAT1*-expressing cells that were spontaneously differentiated for 2 days. RT-qPCR of *NEAT1* positive population normalized to the negative population in **D**.
n=2 independent experiments, except for spontaneous differentiation in **A** with only one replicate. Error bars represent standard deviation.

3.4.6 *NEAT1_1* is dispensable for germ layer differentiation

Finally, I utilized the genetically edited *NEAT1* hESCs to address the function of the short isoform, *NEAT1_1*, which was previously shown to form microspeckles outside of paraspeckles [95]. Since I did not observe different phenotypes for the *NEAT1^-/-^* and *NEAT1ΔTH* human cell lines (**Fig. 21**), although the latter produced the short isoform (**Fig. 20B**), I hypothesized that *NEAT1_1* is dispensable for the differentiation of germ layer progenitors. Indeed, analyzing the differentiation of *NEAT1ΔpA* hESCs that harbor a deletion of the internal polyadenylation did not reveal a difference in the up- and down-regulation of differentiation and pluripotency genes **(Fig. 25A-C)**. In undifferentiated cells however, an increase in germ layer markers, most notably *T* and *PAX6* was observed (**Fig. 25D**), concomitant with increased expression of *NANOG* by 10-fold (**Fig. 25D**), which is reminiscent of the phenotype after *NEAT1* over-expression (**Fig. 25E**) and a reciprocal expression pattern compared to *NEAT1^-/-^* hESCs (**Fig. 21, 22**). To summarize, only the architectural isoform of *NEAT1*, *NEAT1_2*, was required for differentiation, whereas *NEAT1_1*, which in

hESCs, was localized outside of paraspeckles (**Fig. 15A, B**) did not affect the formation of germ layer progenitors.
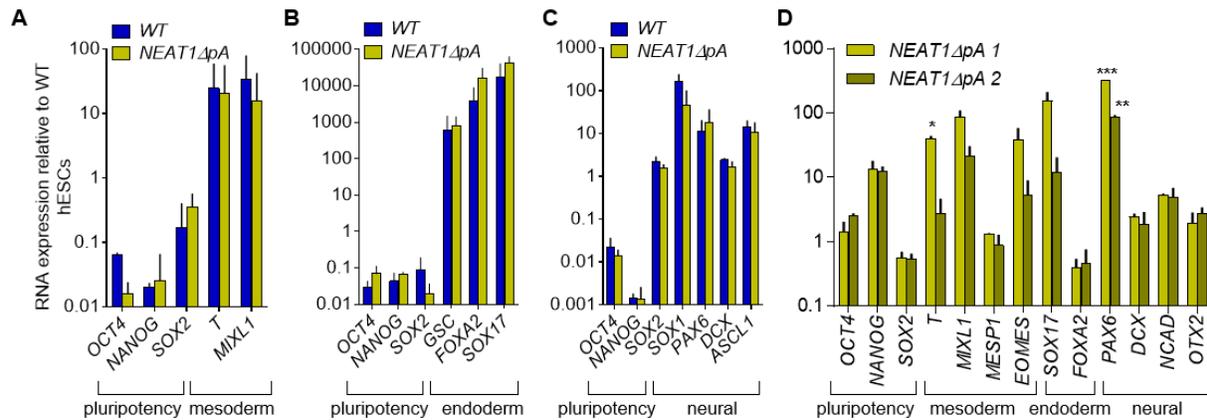
Figure 25: Characterization of gene expression changes in *NEAT1ΔpA* hESCs. A-C) RT-qPCR analysis of *NEAT1ΔpA* hESC clones differentiated to lateral mesoderm (**A**), definitive endoderm (**B**) and neuroectoderm by 4 days differentiation of NSCs (**C**). n=2 biological replicates of cells in different passages and 4 replicates representing two different clones for *NEAT1ΔpA* hESCs. **D)** RT-qPCR of *NEAT1ΔpA* clones in pluripotent conditions. n=2 biological replicates of cells in different passages. Error bars represent standard deviation.

3.5 Functional assays to determine the mode-of-action of paraspeckles in differentiated PSCs

Paraspeckles are a hub for proteins [83] and RNA [84] and *NEAT1*-dependent sequestration of both was shown to have an impact on cell fate [191]. To achieve first mechanistic insights into the molecular mechanisms executed by paraspeckles in differentiated hESCs, I sought to interrogate changes in RNA retention and protein composition in paraspeckle-depleted cell lines.

3.5.1 Global changes in nuclear RNA retention after depletion of paraspeckles

First, to analyze how the nuclear and cytosolic RNA landscape was changed in *NEAT1*[-/-] hESCs, a subcellular fractionation method for the subsequent isolation of nuclear and cytoplasmic RNAs was established. Process control was performed by western blot with enrichment of histone H3 and almost complete absence of β-actin (ACTB) in the nuclear fraction (**Fig. 26A**), and by RT-qPCR with cytoplasmic enrichment of GAPDH and mitochondrial MT-CYB mRNAs and the lncRNAs *NEAT1*, *MALAT1* and *SCARNA10* in the nucleus (**Fig 26B**). Next, I performed polyA-enriched RNA sequencing of differentiated *NEAT1*[-/-] and parental control cells. Enrichment of nuclear lncRNAs in the nucleus and mRNAs in the cytoplasm indicated successful fractionation (**Fig. 26C**). Amongst the ~10000 detected transcripts, 32% changed their location after paraspeckle depletion (p<0.05) (**Fig. 26D**). By comparing the nuclear-to-

cytoplasmic ratio of transcripts in parental and $NEAT1^{-/-}$ cells, I identified transcripts of 76 genes that were exported to the cytoplasm and 135 genes that were retained in the nucleus by more than 2-fold after paraspeckle depletion (**Fig. 26E**). The nuclear retained transcripts are generally produced by cytoplasmic genes with metabolic and translational functions, whereas genes whose transcripts are exported to the nucleus after paraspeckle knock-out exhibited functions in RNA processing (**Fig. 26F**). To summarize, the transcript localization landscape was drastically changed upon loss of paraspeckles, which could lead to changes in the protein amount that might affect stem cell differentiation. However, no mRNAs of differentiation-associated genes were found to change their localization, which is why mRNA retention probably only indirectly contributes to the paraspeckle phenotype described above.
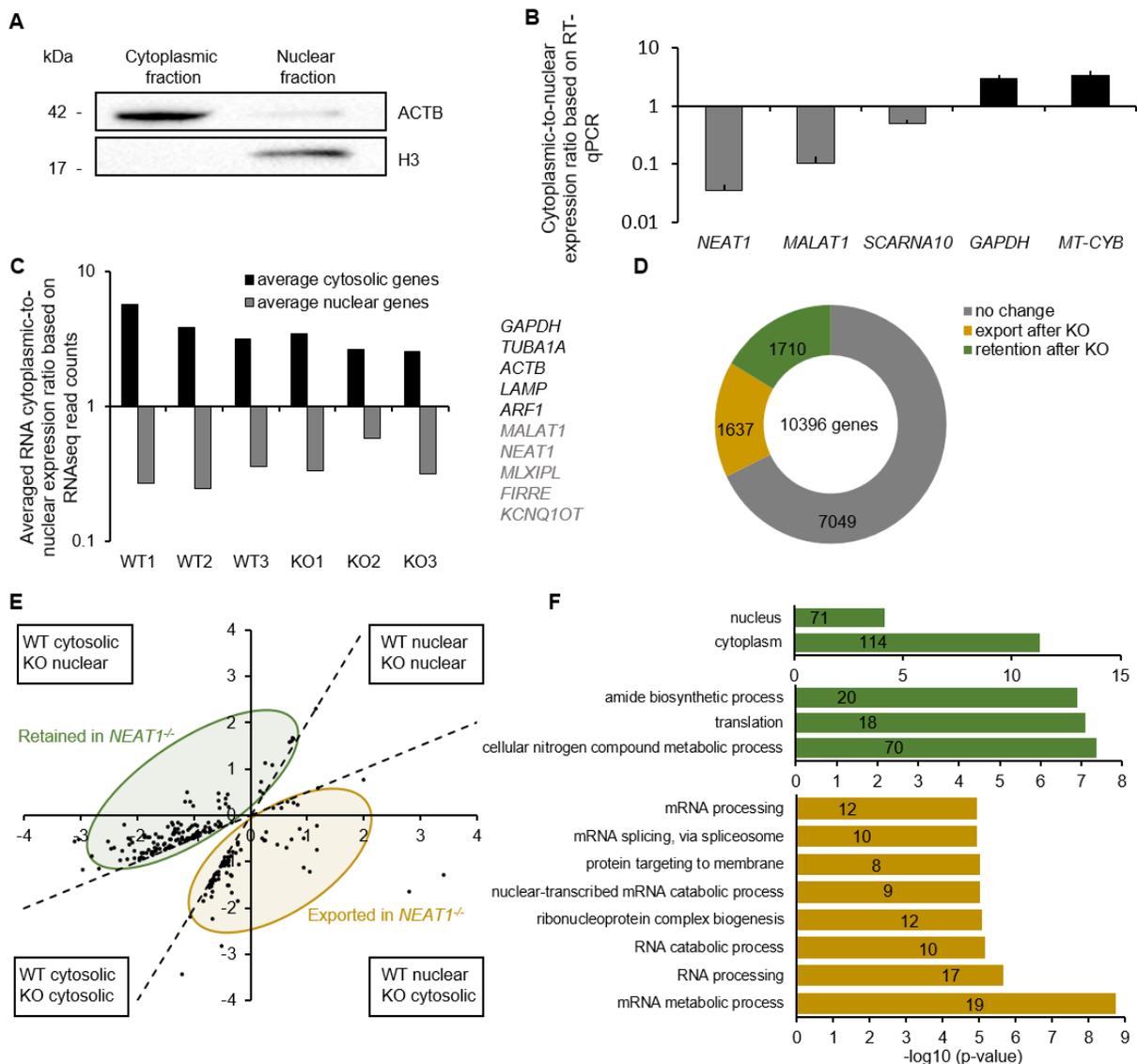
Figure 26: Subcellular fractionation combined with RNA sequencing determines changes in RNA localization upon *NEAT1* depletion. **A, B)** Process control of cytoplasmic-nuclear fractionation by western blot for cytoplasmic β-actin (ACTB) and nuclear histone protein H3 (**A**) and RT-qPCR of nuclear lncRNAs (grey) and cytoplasmic transcripts (black) (**B**). Error bars represent standard deviation. **C)** Averaged cytoplasmic-to-nuclear ratio of cytoplasmic (black) and nuclear (grey) RNAs after RNA sequencing of differentiated wildtype and *NEAT1^-/-^* hESCs. Differentiation was performed for 4 days according to the protocol to generate NSCs. **D)** Number of transcripts that exhibited localization changes after *NEAT1* depletion (p<0.05) **E)** Scatterplot depicting the nuclear-to-cytoplasmic ratio of transcripts in differentiated wildtype (x-axis) and *NEAT1^-/-^* (y-axis) hESCs. **F)** GO enrichment analysis of genes whose transcripts are retained in the nucleus (green) or exported to the cytoplasm (yellow) in *NEAT1^-/-^* hESCs according to the string database. Numbers in bars represent the number of genes found per GO term.

3.5.2 Analyzing the cross-talk between RBPs of paraspeckles and the RNA polymerase II

Many RNA binding proteins are associated to paraspeckles [83], similar to the RNA polymerase II (RNA PolII) to which a plethora of transcriptional and co-transcriptional factors are bound [218]. I hypothesized that the repertoire of RBPs attached to RNA PolII is influenced by paraspeckles. To address this, I performed a pull-down of RNA PolII using an antibody for the phosphorylated Serine 2 residue of its C-terminal domain (CTD), a marker of the elongating PolII [219]. I employed a recently published protocol for chromatin immunoprecipitation with selective isolation of chromatin-associated proteins (CHiP-SICAP) [184] in parental and *NEAT1^-/-^* hESCs after 4 days of neural induction. By mass spectrometry, 34 proteins were identified to be significantly differentially associated to the wildtype or *NEAT1^-/-^* RNA PolII CTD (p<0.05, **Fig. 27A**), amongst which 13 were splicing proteins (p=1.3x10^-11^) and 6 proteins that influence DNA conformational change (p=9.6x10^-5^) (**Fig. 27B**). As expected, 33/34 proteins were nuclear and 30/34 regulate gene expression, which indicates that the experiment was free of cytoplasmic contaminants. I concluded that there is a cross-talk between paraspeckles and RNA polymerase and that depletion of paraspeckles changes the repertoire of RBPs attached to the RNA PolII, mostly by depleting splicing factors. Further experiments are required to untangle the connection between paraspeckles and splicing proteins and how this might be involved in promoting the pluripotency exit.

A

Gene	Log2 (WT/KO)	P-value
S100A9	4.321928095	0.00056447
SUB1	4.321928095	0.00056447
RPS8	4.321928095	0.00056447
PHF6	4.321928095	0.00056447
HIST1H1D	4.236459597	0.00240227
SAP18	4.321928095	0.0101433
HNRPDL	4.321928095	0.0101433
HNRNPR	4.321928095	0.0101433
SNRNP200	4.321928095	0.0101433
U2AFBP	4.321928095	0.0101433
DDX5	4.321928095	0.0101433
PTBP1	4.321928095	0.0101433
RPL3	4.321928095	0.0101433
CBX5	4.321928095	0.0101433
RNPS1	4.321928095	0.0101433
ZNF280C	4.321928095	0.0101433
SRRT	4.321928095	0.0101433
SLTM	4.321928095	0.0101433
H2AFY2	4.321928095	0.0101433
SALL4	4.321928095	0.0101433
HIST1H1B	3.154433528	0.0152072
PNN	1.183074669	0.02245921
XRCC6	3.99112519	0.02508661
UBA52	3.376375272	0.03304082
VIM	3.948697237	0.04024355
HMGB1	3.830158202	0.04171047
U2AF2	3.826133932	0.04446722
HNRNPK	3.511533393	0.05354376
ENO1	3.908150137	0.06126336
HMGN1	2.608882757	0.06131647
HNRNPA1	2.7481093	0.08346627
H1FX	1.934210716	0.09094365
PPIA	1.518452577	0.09839229
THRAP3	0.676001041	0.10208124

Figure 27: ChIP-Sicap reveals changes in RNA-PolII CTD repertoire after *NEAT1* depletion. **A)** A table depicting genes that were significantly enriched in wildtype compared to *NEAT1* knock-out RNA PolII, which was pulled down using an antibody for phosphorylated Serine 2 residue of the CTD. Proteins were identified by mass spectrometry. Differentiation was performed for 4 days according to the protocol to generate NSCs. **B)** Protein interaction network extracted from the string database (medium confidence setting) for significantly changed genes indicated in **A**. Splicing proteins are shown in red and proteins enriched with functions in changing the DNA conformation in blue.

3.6 DBHS proteins regulate exit from pluripotency

Members of the DBHS protein family are an integral part of paraspeckles [82] and often, the mode-of-action of paraspeckles can be explained by sequestration of DBHS proteins. In humans, three DBHS proteins are expressed, namely SFPQ, NONO and PSPC1, all of which are produced in hESCs at a level similar to the pluripotency genes and substantially higher than differentiation markers (**Fig. 28A, B**). In the following sections, I dissected the dynamics and effect of DBHS proteins on ESC maintenance and differentiation.

3.6.1 DHBS proteins localize to paraspeckles at the onset of differentiation

NEAT1 up-regulation correlates with lineage-independent exit from pluripotency (**Fig. 11**) but it is unknown whether the expression of DBHS proteins exhibits a similar pattern. To address this, I performed differentiation of hESCs into the germ layers by differentiation towards cardiomyocytes, neural rosettes and definitive endoderm and analyzed the expression of *SFPQ*, *PSPC1* and *NONO*. Moreover, I measured expression changes during reprogramming of hESCs that are generally in a primed state back to their naïve state. The latter will be outlined in detail in section 3.6.5. With exception of a minor down-regulation of *PSPC1* during mesoderm and endoderm specification, the expression of DBHS proteins did not change drastically after exit from pluripotency and also not after reprogramming to the naïve state (**Fig. 28C**). However, as expected, DBHS proteins formed aggregates in differentiated cells but not in hESCs that are devoid of paraspeckles (**Fig. 28D, E**). I concluded that the expression of DBHS proteins and *NEAT1* is uncoupled, but paraspeckles could act as sponge for DBHS proteins and thereby regulate their function.
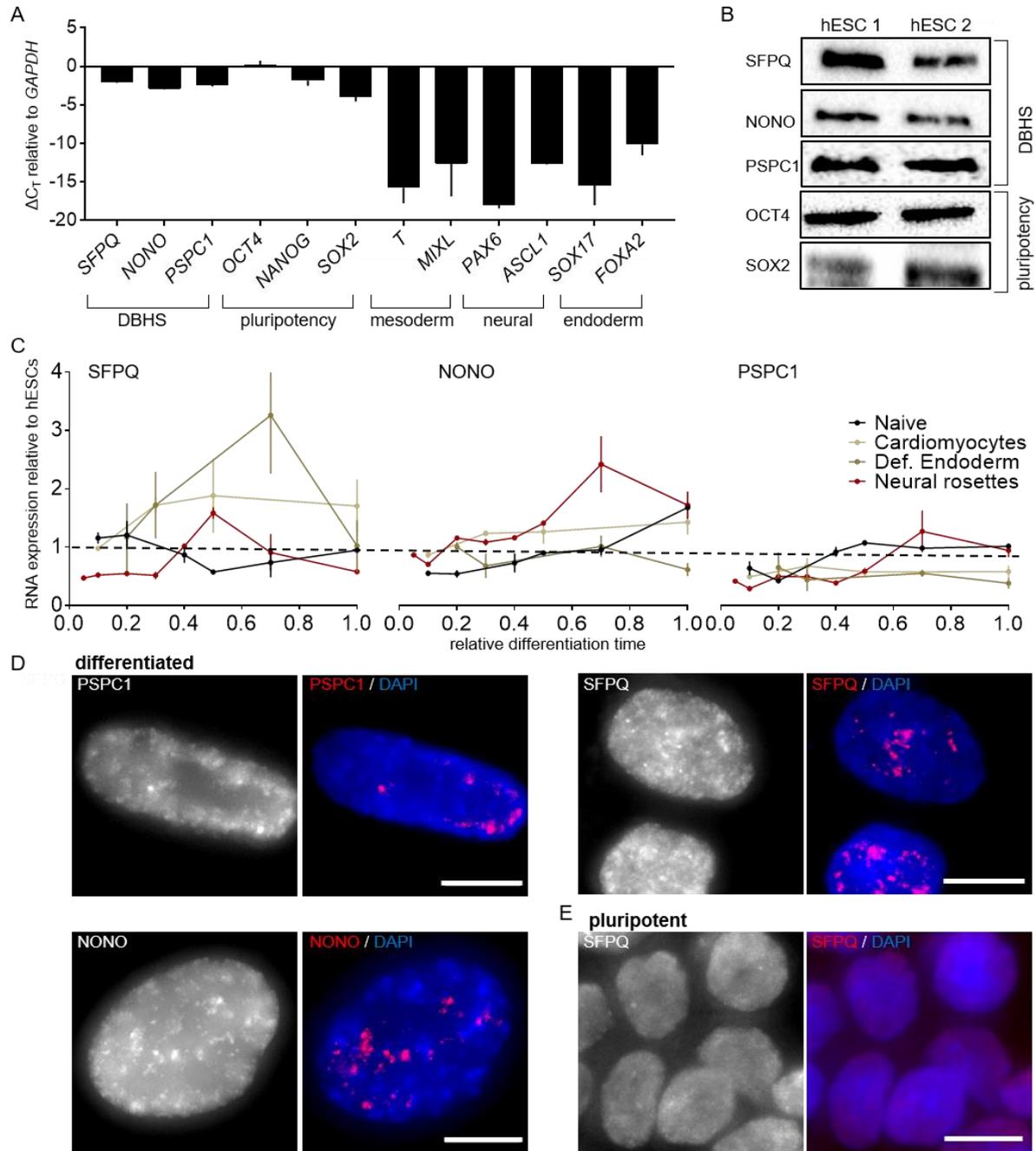
Figure 28: Transcriptional and post-translational regulation of DBHS proteins in differentiated hESCs.
A) RT-qPCR analysis of DBHS, pluripotency and differentiation genes in undifferentiated hESCs. **B)** Western blot depicting the expression of DBHS and pluripotency proteins in undifferentiated hESCs in 2 replicates. **C)** RT-qPCR of DBHS genes during naïve conversion and differentiation of hESCs to cardiomyocytes, definitive endoderm and neural rosettes, for 70, 9, 6 and 14 days respectively. Differentiation time was normalized to the endpoint. Dashed line marks base line expression level in hESCs. **D, E)** Representative immunocytochemistry images labeling DBHS proteins (red) in differentiated **(D)** and undifferentiated **(E)** hESCs. DAPI (blue) marks nuclei and scale bar is 10 μm. Cells were differentiated for 3 days with retinoic acid.
n=2 experiments using cells of different passages. Error bars represent SEM.

3.6.2 DBHS proteins PSPC1 and NONO maintain pluripotency in human ESCs

To dissect the developmental functions of human DBHS proteins, I disrupted the first exons of *NONO*, *PSPC1* and *SFPQ* in hESCs using CRISPR/Cas9 (**Fig. 29A**) and confirmed the functional depletion of the genes by analyzing the respective genomic sites, transcripts and proteins (**Fig. 29B-H**). Although deletion of the first exon of *SFPQ* was efficient, it was apparent predominantly in detached dead cells, and in accordance, none of the clones exhibited *SFPQ* mutant alleles that were detected in detached cells (**Fig. 29D**). Nevertheless, I could knock-down 60-80% and 40-50% *SFPQ* transcript and protein using siRNAs (**Fig 29I, J**).

I analyzed first, whether depletion of DBHS proteins was compensated by up-regulation of other DBHS family members, however, I did not observe any significant changes in pluripotent hESCs (**Fig. 29K**). To analyze the involvement of DBHS proteins in maintaining the pluripotent, undifferentiated state, I quantified the expression of key reprogramming - pluripotency genes. This showed that pluripotency is down-regulated by 40-60% in *PSPC1*[-/-] hESCs, but not significantly perturbed by removal of SFPQ and NONO (**Fig. 29L**). Moreover, I analyzed genes that drive the specification of the primary early germ layers, and genes involved in EMT, which characterizes some of the first embryonic developmental transitions [220]. Interestingly, I noted significant up-regulation in the basal expression of canonical germ layer genes primarily in *NONO*[-/-], but also to some extent in *PSPC1*[-/-] hESCs. This included *PAX6*, *BRACHYURY* (*T*) and *SOX17*, which drive the specification of ectoderm, mesoderm and endoderm progenitors, respectively (**Fig. 29M**). I concluded that paraspeckle-independent functions of DBHS proteins PSPC1 and NONO is to respectively prime the multilineage differentiation and promote pluripotency. Knock-down of *SFPQ* did not significantly affect the expression of differentiation genes (**Fig. 29L, M**), thus I concluded that perturbation of the pluripotency circuit is not a plausible mechanism for the lethal phenotype as a result of *SFPQ* disruption. Collectively, this indicates that DBHS proteins regulate pluripotency-differentiation balance and that *SFPQ* has house-keeping functions through mechanisms that are independent of key developmental genes.
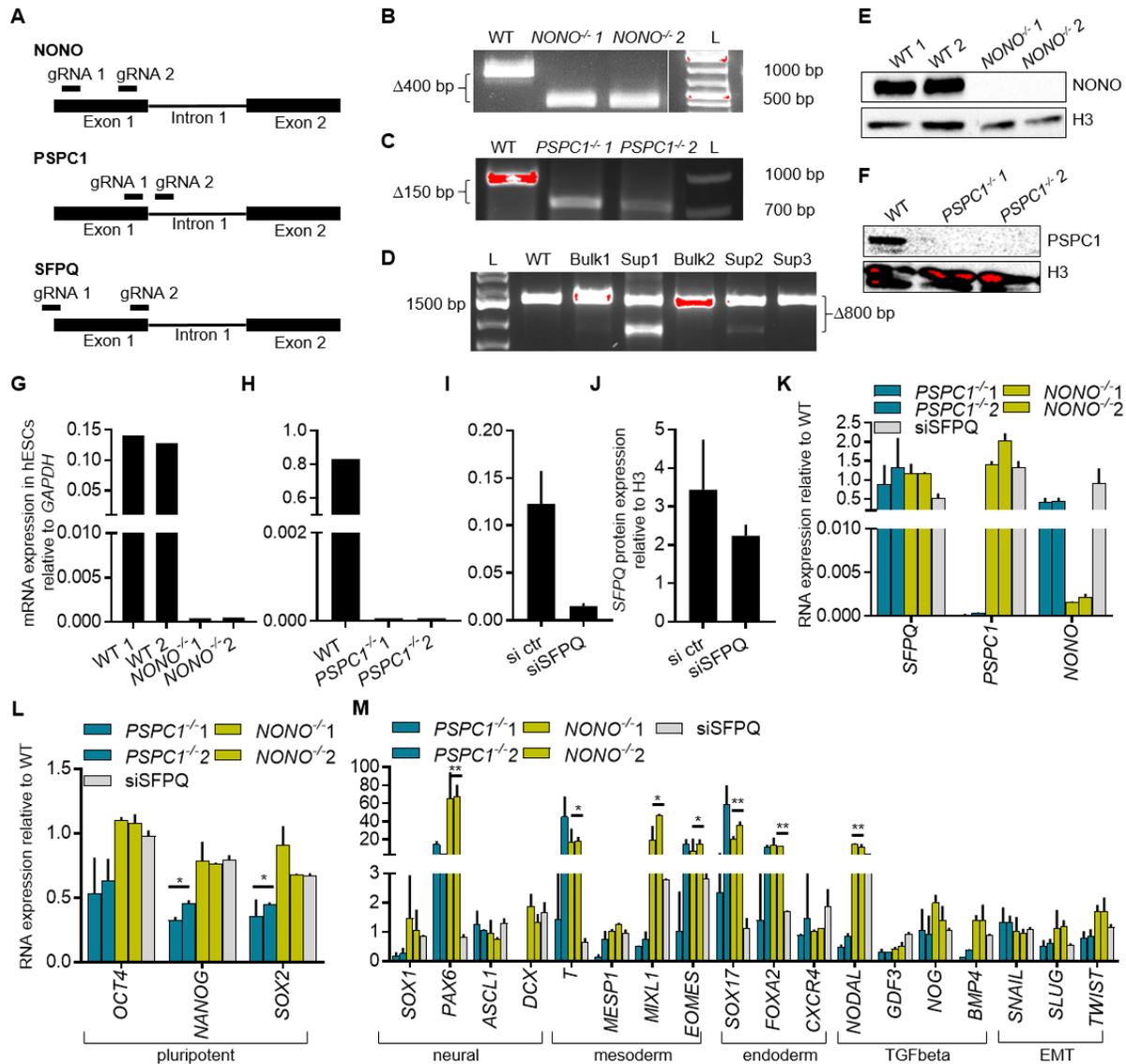
Figure 29: Generation and analysis of DBHS-depleted hESCs. **A)** A scheme for the generation of *NONO*, *PSPC1*, and *SFPQ* knock-out hESCs by CRISPR/Cas9. **B, C)** PCR screening of genomic deletion in the *NONO* (**B**) and *PSPC1* (**C**) coding sequence for two isolated clones. **D)** PCR screening of genomic deletion in the *SFPQ* coding sequence. Bulk represents living cells that attached while supernatant contained dead, detached cells. Bulk1, Sup1, Bulk 2, Sup2 and Sup3 were analyzed 2, 7 and 14 days after Cas9/gRNA transfection, respectively. **E, F)** Western blot for NONO (**E**) and PSPC1 (**F**) protein in two isolated clones. H3 was used as loading control. **G, H)** RT-qPCR analysis of NONO (**G**) and PSPC1 (**H**). **I, J)** SFPQ mRNA and protein expression two days after siRNA transfection analyzed by RT-qPCR (**I**) and western blot (**J**), respectively. Protein levels were normalized to H3. **K-L)** RT-qPCR analysis of DBHS proteins (**K**), pluripotency transcription factors (**L**) and differentiation markers (**M**) in *NONO*[-/-] and *PSPC1*[-/-] and *SFPQ* knock-down hESCs. RNA expression was normalized to *GAPDH* and is depicted relative to wildtype (WT) levels with n=3 biological replicates. * p<0.05, ** p<0.01, unpaired t-test.

3.6.3 NONO regulates spontaneous differentiation via metabolic control

To further characterize how NONO promotes pluripotency independently of paraspeckles, I performed RNA sequencing of *NONO^-/-^* and parental undifferentiated hESC samples. Importantly, this revealed the down-regulation of 320 genes (p<0.05) that classified functional categories involved in the regulation and synthesis of cholesterol and related metabolic products (**Fig. 30A**). Cholesterol is produced from acetyl-CoA, which also serves as a precursor for histone acetylation and thereby priming of differentiation genes in PSCs [221]. Therefore, reduction of cholesterol production could explain up-regulation of 188 genes (p<0.05) implicated in developmental processes (**Fig. 30B**), including *HOX* genes that are important for body axis development [56] and genes which have been implicated in neural differentiation and CNS formation (*COL3A1, EPHA4, HES1, EGR1, ID4, FOXJ1*) (**Fig. 30C**). I confirmed the down-regulation of cholesterol synthesis pathway by RT-qPCR for enzymes that actively take part in cholesterol metabolism and found that in pluripotent condition, all of them were down-regulated by 40-70% compared to parental control (**Fig. 30D**). Collectively, this implicates NONO in the paraspeckles-independent maintenance of pluripotency by regulating enzymes that shift metabolite synthesis, and thereby prime the activation of genes that promote germ layer development.

Next, I assessed whether down-regulation of cholesterol synthesis has an impact on differentiation. As expected, I observed significant up-regulation of many developmental genes, such as the endoderm markers *FOXA2, CXCR4* and the neural transcription factors *FOXG1* and *PAX6* in *NONO^-/-^* hESCs upon removal of the pluripotency medium (**Fig. 30E**). Importantly, this is in contrast to a previous study in mouse ESCs that showed an increase of the pluripotency gene expression after *Nono* knock-out, concomitant with reduced differentiation potential, which the authors demonstrated by impaired formation of embryonic bodies and more importantly, by impaired neuronal differentiation [222]. Strikingly, when differentiating *NONO^-/-^* hESCs to motor neurons, I did not observe differences in the gene expression profile of neuronal markers compared to the parental control cell line (**Fig. 30F, G**), which led me to conclude that the function of NONO diverges from its mouse orthologue in human ESCs.
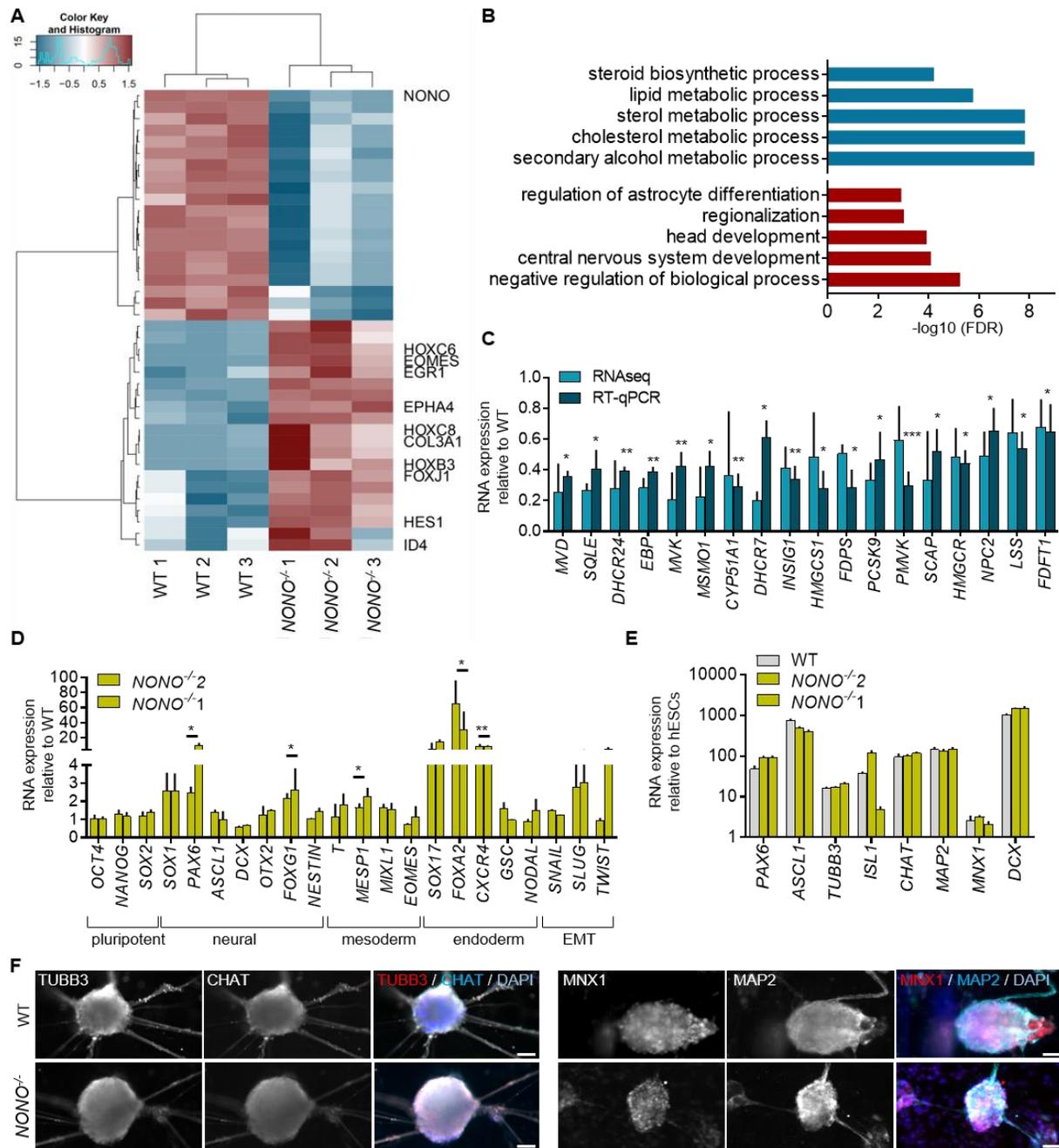
Figure 30: NONO-mediated gene expression changes in pluripotent and differentiated conditions. A) Heat map clustering of the top 20 highest and lowest significantly differentially expressed genes in *NONO⁻/⁻* hESCs compared to parental cells. **B)** Gene ontology (GO) term analysis of up- and down-regulated genes in *NONO⁻/⁻* hESCs compared to WT. Analysis was performed on the String platform **C)** RT-qPCR analysis of genes of the cholesterol synthesis pathway in *NONO⁻/⁻* hESCs and comparison with RNA sequencing data. **D)** RT-qPCR of pluripotency-differentiation markers of *NONO⁻/⁻* hESCs spontaneously differentiated for 3 days. **E, F)** RT-qPCR (**E**) and immunostaining (**F**) of neuronal markers in *NONO⁻/⁻* and WT motor neurons differentiated for 75 days. Scale bar is 50 μm.

n=4 biological replicates including two *NONO⁻/⁻* clones. * p<0.05, ** p<0.01, *** p<0.001, unpaired t-test. Error bars represent standard deviation.

3.6.4 PSPC1 depletion impairs *in vitro* adipogenesis

Besides NONO, PSPC1 has been implicated in murine development including mouse adipogenesis, which is regulated by *Pspc1* [223]. Since *NONO*-depleted hESCs exhibited a different phenotype compared to *NONO*$^{-/-}$ mESCs, I tested whether this is the case for PSPC1 as well. To address this, I sought to recapitulate the murine *Pspc1* knock-out phenotype by differentiating *PSPC1*$^{-/-}$ human ESCs *in vitro* to adipocytes via MSCs. Indeed, there were morphologic differences in the MSC morphology with *PSPC1*$^{-/-}$ MSCs forming a loose monolayer of cells in contrast to MSCs of the parental cell line that exhibited dense colony-like morphology (**Fig. 31A**). This was concomitant with reduced expression of MSC surface markers *CD73* and *CD90* (**Fig. 31B**), of EMT markers *SLUG* and *TWIST*, a hallmark of functional MSCs, and of *SIM2*, an MSC-specific transcription factor (**Fig. 31C**). Further differentiation of MSCs to adipocytes resulted in cells with lower expression of the adipocytes marker *FABP4* (**Fig. 31D**) and importantly, Oil Red O staining of lipid droplets was markedly reduced in *PSPC1*$^{-/-}$ adipocytes (**Fig. 31E**). Nevertheless, expression of other markers of *PSPC1*$^{-/-}$ MSCs and adipocytes was not affected suggesting that PSPC1 only to some extent is required for *in vitro* human adipogenesis, which is in contrast to the phenotype observed in mice [223].

85
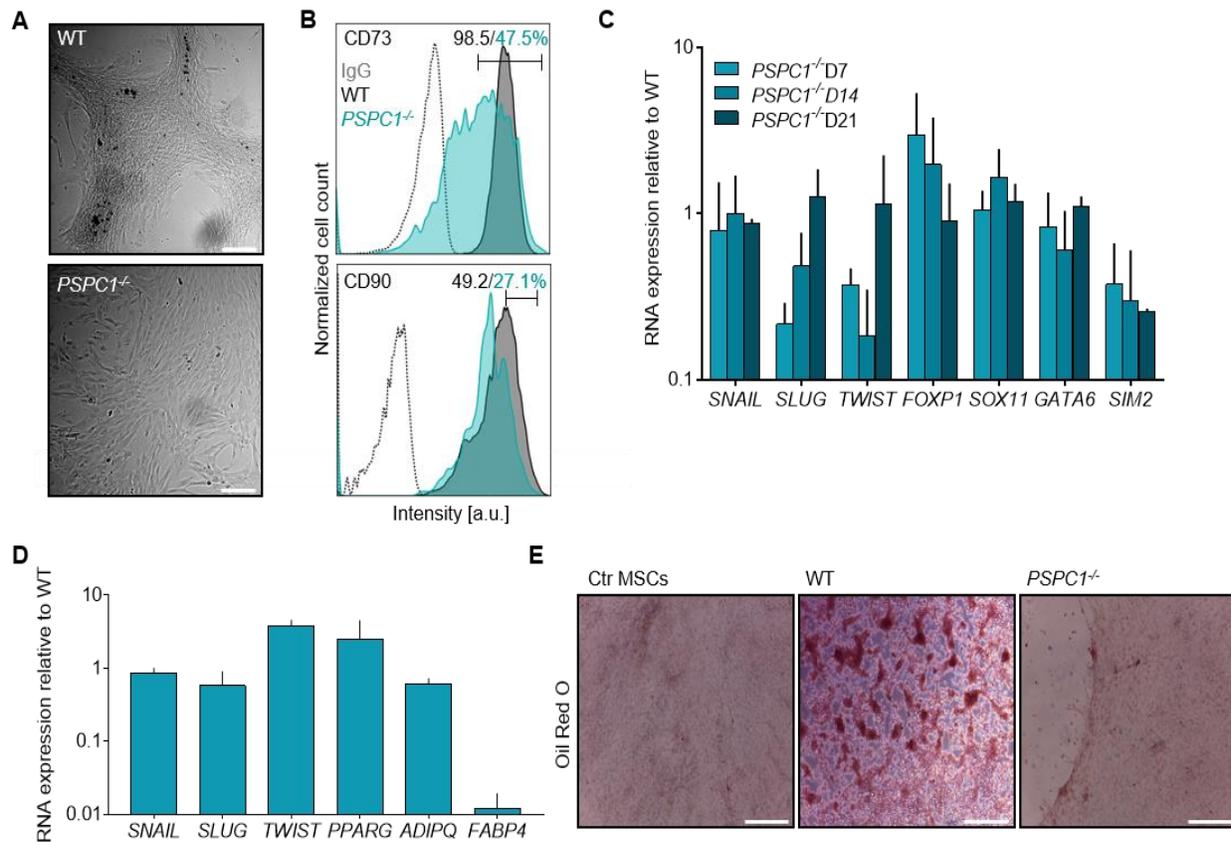
Figure 31: Phenotypic analysis of *PSPC1*-/- MSCs and adipocytes. A) Morphology of *PSPC1*-/- and parental (WT) MSCs. **B)** Flow cytometry analysis of MSC surface markers CD73 and CD90 in WT (black) and *PSPC1*-/- (cyan) MSCs. IgG control in grey. **C)** Time course RT-qPCR analysis of EMT markers and MSC transcription factors in *PSPC1*-/- MSCs. RNA expression was normalized to WT on the respective days. **D)** Gene expression analysis of adipocyte and EMT markers for *PSPC1*-/- adipocytes differentiated for 56 days. **E)** Oil Red O staining of lipid droplets in WT and *PSPC1*-/- adipocytes with WT MSCs as negative control. Magnification in phase-contrast images is 5x with a scale bar of 500 μm. n=3 biological replicates for WT and 4 replicates including 2 different clones for *PSPC1*-/- MSCs or adipocytes. Error bars represent standard deviation.

3.6.5 NONO and PSPC1 inhibit naïve conversion of human ESCs

Next, I asked whether PSPC1 or NONO are involved in the early developmental transitions of hESCs. To address this, I differentiated DBHS protein-depleted hESCs to intermediate mesoderm, definitive endoderm and neural rosettes and performed gene expression analysis for a panel of lineage-specific marker genes. No significant changes in the gene expression profile of lineage markers was observed upon depletion of *PSPC1* and *NONO* indicating that both genes are not required for germ layer specification (**Fig. 32**).
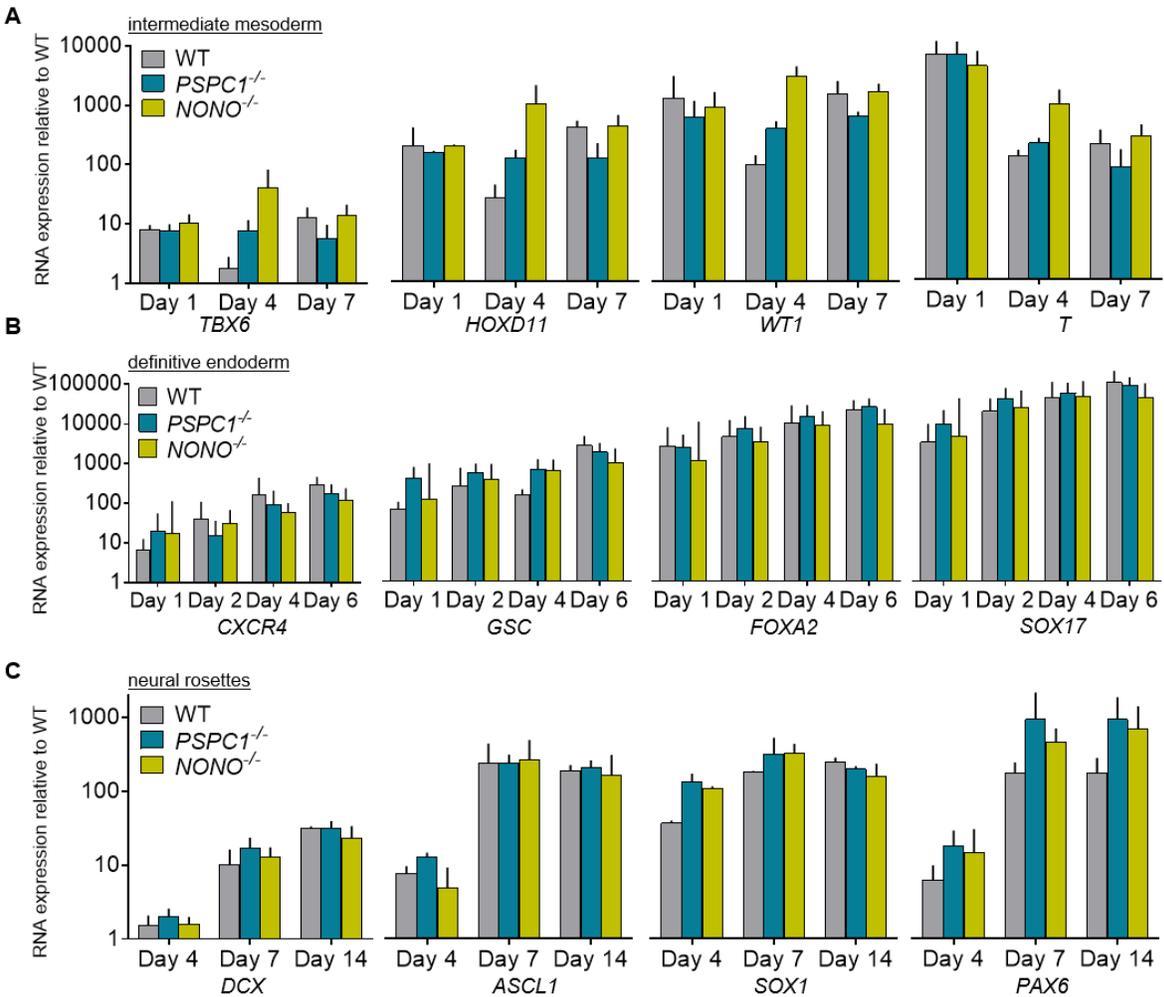
Figure 32: Gene expression analysis of *PSPC1*⁻/⁻ and *NONO*⁻/⁻ hESCs committed to germ layer specification. **A-C**) Time course RT-qPCR for intermediate mesoderm (**A**), definitive endoderm (**B**), and neural rosette (**C**) differentiation of *PSPC1*⁻/⁻ and *NONO*⁻/⁻ hESCs. n=2 biological replicates for WT and 4 replicates including two different clones for knock-out cell lines. Error bars represent standard deviation.

It is widely accepted by now that mouse ESCs represent a ground, or naïve state of pluripotency, whereas the transcriptional profile and other phenotypic features of human ESCs are more similar to the primed, epiblast-like stage of pluripotency [26]. Recently, it was demonstrated that primed hESCs can be converted to the naïve stage, which prompted me to investigate whether DBHS proteins might influence this process. I employed a commercially available medium to generate naïve cells for 10 passages. While parental and *NONO-* or *PSPC1*-depleted hESCs exhibited similar dome-shaped colony formation (**Fig. 33A**), the flow cytometry analysis of primed surface markers CD24 and CD90 revealed a slightly accelerated loss of primed pluripotency in *PSPC1*⁻/⁻ hESCs, which was more pronounced in *NONO*⁻/⁻ cells (**Fig. 33B, C**). I verified successful conversion to the naïve stage by RT-qPCR analysis of naïve markers *NANOG, LBP9,*

KLF4, *KLF2* and *DNMT3L*, which increased in expression with continued passaging. Strikingly, the expression of naïve markers was even further elevated in *PSPC1*[-/-] or *NONO*[-/-] hESCs for a majority of genes and at all time points (**Fig. 33D**). I concluded that both NONO and PSPC1 perturb the naïve-to-primed transition of human ESCs. Interestingly, when differentiating naïve *PSPC1*[-/-] and *NONO*[-/-] hESCs, I observed changes in their transcriptional profile compared to primed, *PSPC1*- or *NONO*-depleted hESCs (**Fig. 33E, F**), which indicated that the DBHS-related phenotype in PSCs depends on their developmental state.
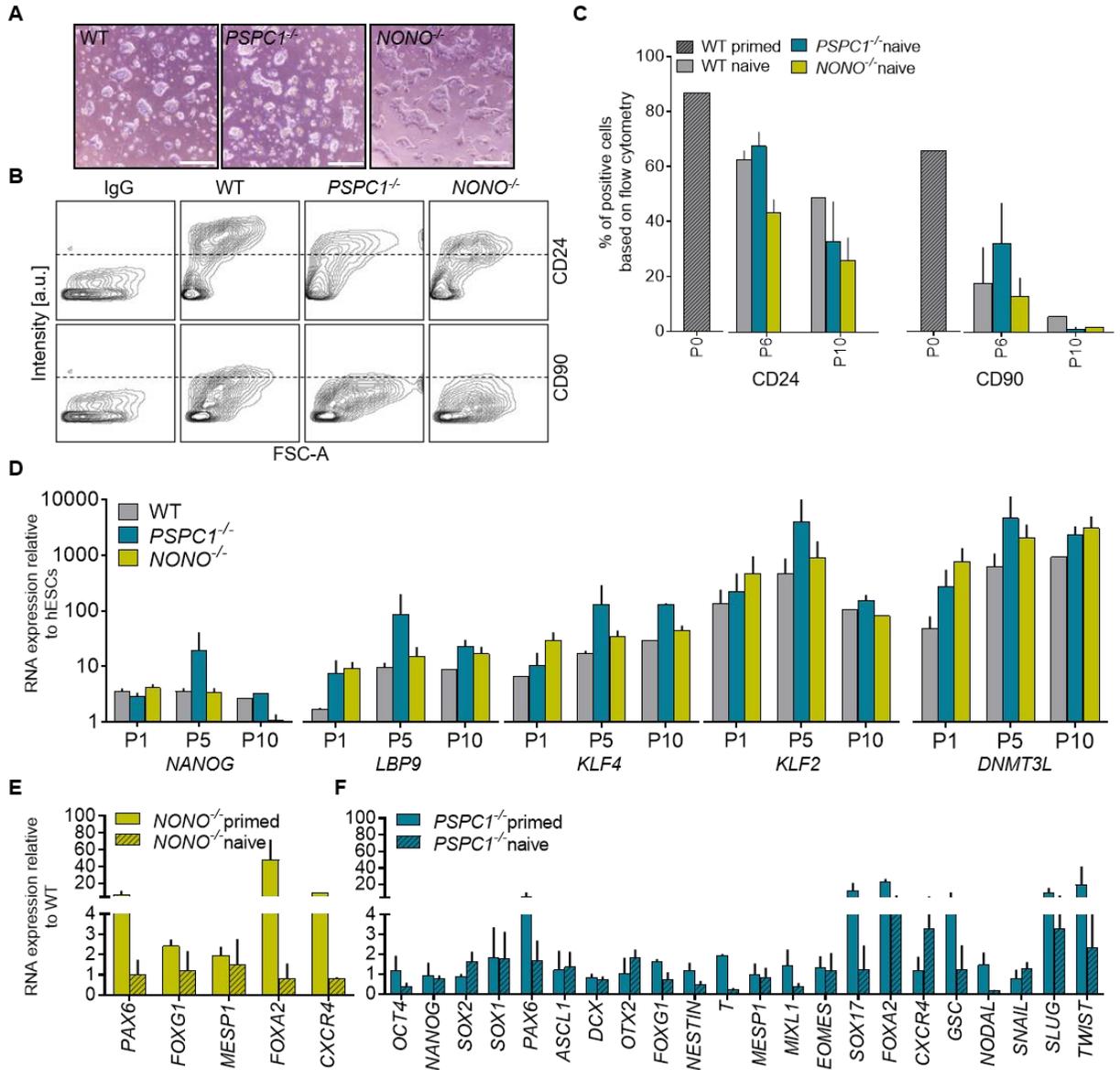
Figure 33: Generation of *NONO* and *PSPC1* knock-out naïve human ESCs. A) Brightfield images of naïve WT, *PSPC1*[-/-] and *NONO*[-/-] hESCs at passage 10 (P10). Scale bar: 500 μm. **B, C)** Flow cytometry analysis of primed hESCs surface markers CD24 and CD90 in primed WT and naïve WT, *PSPC1*[-/-] and *NONO*[-/-] hESCs. Representative blots in **B** and quantification of 2 independent experiments with 2 knock-out clones in **C**. **D)** RT-qPCR of naïve markers in WT, *PSPC1*[-/-] and *NONO*[-/-] hESCs at passages 1, 5 and 10 of naïve

conversion. Expression was normalized to primed hESCs with 2 biological replicates of WT and 4 replicates including 2 different clones for *PSPC1* and *NONO* knock-out hESCs. **E, F)** RT-qPCR of pluripotency and differentiation genes after 3 days of spontaneous differentiation of naïve and primed *NONO*[-/-] (**E**, primed values from **Fig. 31D**) and *PSPC1*[-/-] (**F**) hESCs.
n=2 biological experiments. Error bars represent standard deviation.

3.6.6 SFPQ mediates cellular homeostasis by association with the polyadenylation machinery

Finally, I aimed to understand the function of the third DBHS protein, SFPQ, in the maintenance of PSCs. A full knock-out of SFPQ proved to be lethal for hESCs (**Fig. 29D**), however I could reduce its expression by RNA interference (**Fig 29I, J**). I performed RNA sequencing two days after siRNA transfection in pluripotent condition and found >700 differentially expressed genes (p<0.05). Amongst them were the top up-regulated genes *FOXJ1* and *NODAL* (**Fig. 34A**), two proteins that are important for development and which were up-regulated in *NONO*[-/-] hESCs as well (**Fig. 29L, 30A**). Nevertheless, the overall overlap of differentially expressed genes after *SFPQ* and *NONO* down-regulation was minimal (**Fig. 34B**) indicating that both proteins control different cellular processes in hESCs. Down-regulated genes were classified with functions in the organization of organelles and cell cycle regulation, processes that are critical for cell survival (**Fig. 34C**).

SFPQ regulates both, RNA- and DNA-related processes, dependent on the cellular background [146]. I performed co-immunoprecipitation (co-IP) of SFPQ to identify potential interacting partners, which would provide a first hint of how SPFQ exerts its pivotal function in hESCs. By mass spectrometry, 89 proteins were identified to be significantly enriched over IgG control (p<0.05; enrichment >5-fold). With NONO, PSPC1, FUS, HNRNPM and MATR3, most of the known interactors of SFPQ were identified (**Fig. 34D**). Moreover, the majority of immunoprecipitated proteins had annotated function in RNA processing, including splicing proteins but also 3`end processing (**Fig. 34E**). This includes almost the complete polyadenylation machinery that was co-purified together with SFPQ (**Fig. 34F**). Together, this data indicates that in hESCs, SFPQ controls gene expression by interaction with the splicing and polyadenylation machinery resulting in regulation of genes that have critical functions in cell homeostasis.
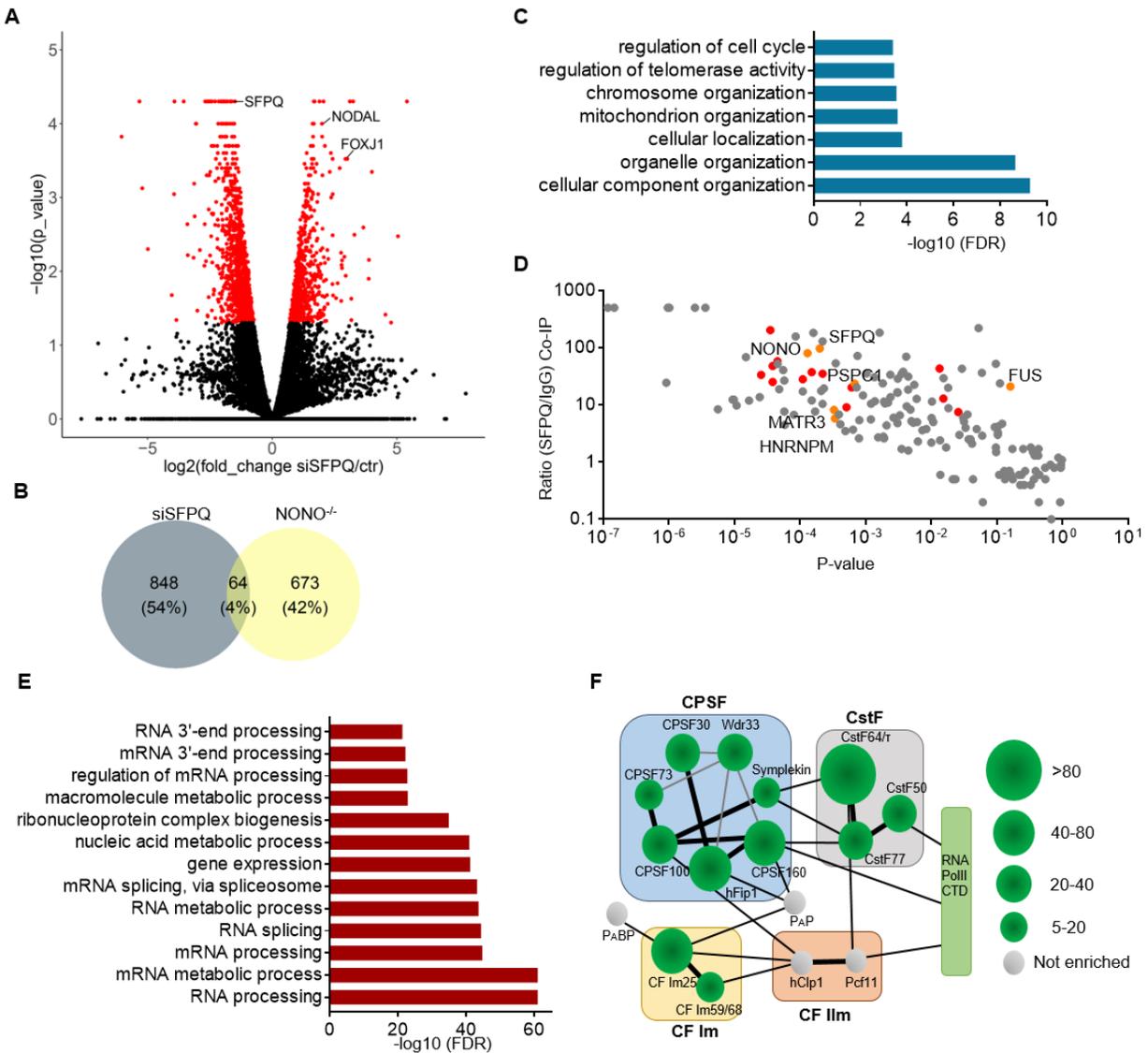
Figure 34: Characterization of SFPQ binding partners and regulated genes. **A)** Volcano plot of differentially expressed genes detected by RNA sequencing after *SFPQ* knock-down by siRNA transfection in hESCs. p<0.05 in red. n=3 independent experiments. **B)** Venn diagram showing the overlap between differentially expressed genes (p<0.05) upon depletion of *SFPQ* and *NONO*. **C)** GO term analysis of significantly down-regulated genes (p<0.05) after siSFPQ treatment. **D)** Mass spectrometry analysis of co-immunoprecipitated proteins using SFPQ as bait protein. n=3 independent experiments of cells in pluripotent condition. Known SFPQ interactors are marked in orange (based on the SFPQ interactome published on the string database), proteins of the polyadenylation machinery in red. **E)** GO term analysis of significantly enriched proteins after SFPQ Co-IP. **F)** Proteins of the polyadenylation machinery. Green circles indicate proteins that were co-immunoprecipitated together with SFPQ. Circle circumference depicts fold enrichment compared to IgG control. Line thickness indicates the degree of interaction. Figure was adapted from Chan *et al*. [224].

3.6.7 NONO is important for paraspeckle integrity but dispensable for *NEAT1_2* stability

Lastly, I sought to interrogate the connection between DBHS proteins and paraspeckle formation. It was shown in many studies, that NONO but not PSPC1 is critical for paraspeckle assembly, possibly by stabilizing *NEAT1_2*, the architectural backbone of paraspeckles [191]. However, these studies were carried out mostly in somatic cells and employed conventional FISH to analyze *NEAT1* and thus paraspeckle formation [85]. Here, I performed smFISH to measure paraspeckle amounts in differentiated hESCs that lack expression of either *PSPC1* or *NONO*. I treated the cells with retinoic acid, a potent inducer of paraspeckles [47] and in agreement with other studies [135], [190], *PSPC1*$^{-/-}$ hESCs exhibited similar amounts of paraspeckles as the parental cell line (**Fig. 35A, B**). In contrast, while previous literature suggested a complete absence of paraspeckles in *NONO*-depleted cells [85], they were still present in differentiated *NONO*$^{-/-}$ hESCs (**Fig. 35A, B**). Strikingly, next to paraspeckles, differentiated *NONO*$^{-/-}$ hESCs exhibited many *NEAT1_2* foci with low intensity, which fell below the intensity threshold for automated paraspeckle detection (**Fig. 35E**). I termed these foci *NEAT1_2* microspeckles, analogous to microspeckles formed by *NEAT1_1* that have been identified, recently [95]. *NEAT1_2* microspeckles were detected in 84% of *NONO*$^{-/-}$ cells while only 16% exhibited a reduced amount or zero paraspeckles (**Fig. 35C**). A cross-section of *NEAT1_2* microspeckles revealed that they displayed a relatively uniform intensity profile and hence contain most likely only one or a few *NEAT1_2* molecules (**Fig. 35D, E**). I concluded that NONO is important for paraspeckle integrity, however, its depletion is not sufficient for degradation of *NEAT1_2* transcripts when treating the cells with retinoic acid. Moreover, the knock-out of *NONO* leads to the appearance of *NEAT1_2* microspeckles similar to the single *NEAT1_2* molecules formed after addition of DNA-binding molecules (**Fig. 18, 19**)
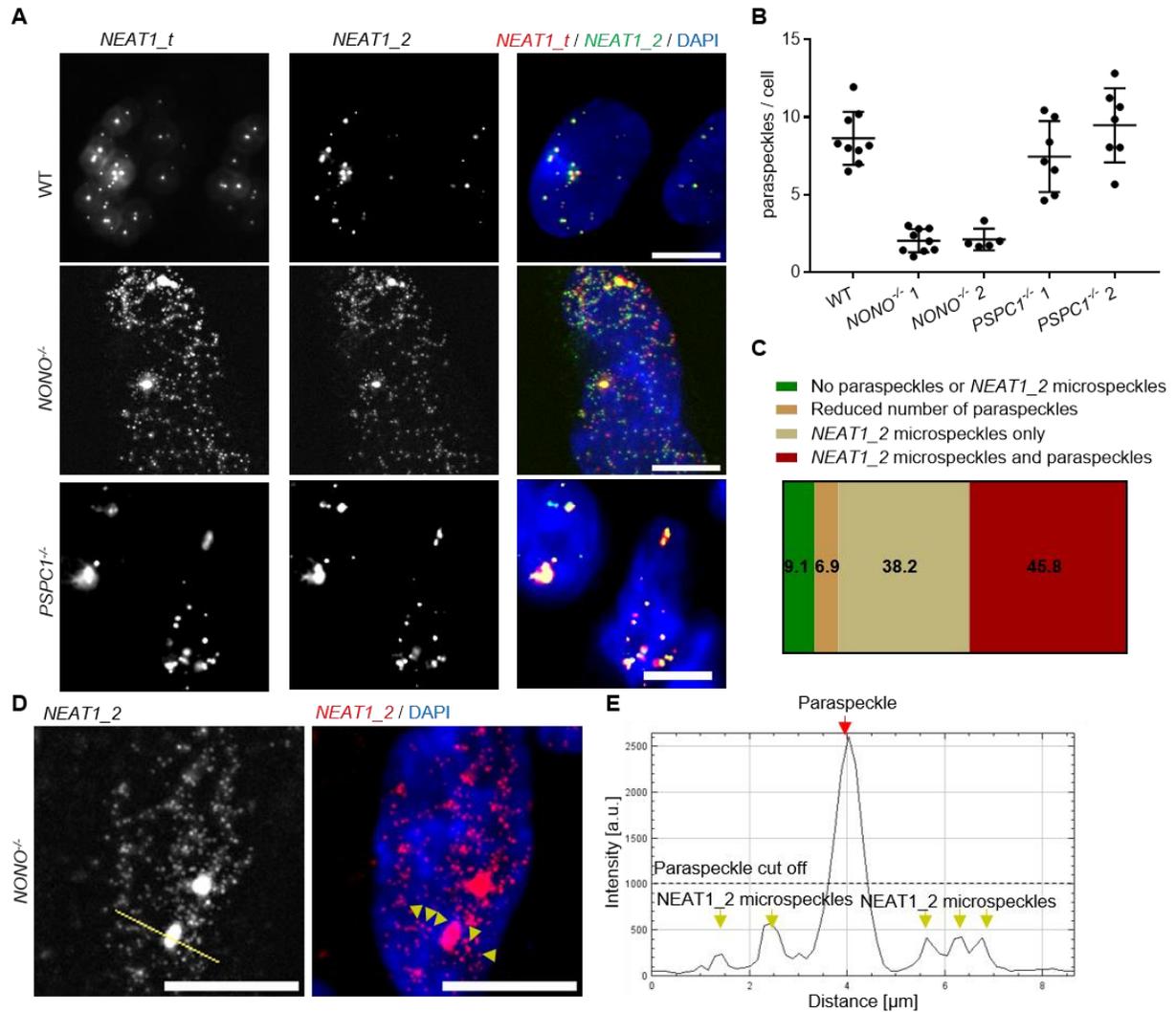
Figure 35: Paraspeckle formation in *PSPC1*-/- and *NONO*-/- hESCs. A) SmFISH of *NEAT1_total* (red) and *NEAT1_2* (green) in WT, *PSPC1*-/-, *NONO*-/- and parental hESCs after treatment by RA for 3 days. **B)** Quantification of the number of paraspeckles based on *NEAT1_2* staining depicted in **A**. n=2 biological replicates with 3-4 images per condition. **C)** Quantification of 200 *NONO*-/- hESCs based on their number of paraspeckles and *NEAT1_2* microspeckles. **D, E)** *NEAT1_2* intensity profile along the yellow line. Yellow arrows depict *NEAT1_2* microspeckles along this line, which also crossed a paraspeckle, marked by the red arrow in **E**.

Scale bar is 10 µm. DAPI in blue marks nuclei.

4. Discussion

4.1 Identification of developmentally regulated nuclear lncRNAs

Nuclear lncRNAs represent an additional layer of gene regulation, in part by binding to chromatin and attracting chromatin-modifying complexes [225]. Here, I showed that the expression pattern of many nuclear lncRNAs is changed during germ layer differentiation (**Fig. 11**), indicating their pivotal role in the pluripotency-differentiation transition. Importantly, besides lncRNA *lncPRESS1*, which has been shown to control pluripotency maintenance [69], many dynamically regulated lncRNAs have not been annotated so far in the context of stem cell differentiation. This includes the top two up-regulated lncRNAs *PINCR* and *LINC00472*, which in colorectal cancer cells, are respectively up- or down-regulated [226], [227], and the tumor suppressor lncRNA *PVT1*, which regulates *MYC* expression in breast cancer cells [228]. *MYC* is crucial for PSC maintenance [229] and promotes the generation of iPSCs [230], hence it is likely that down-regulation of *PVT1* during exit from pluripotency results in *MYC* down-regulation, thereby inducing stem cell differentiation. LncRNA *MEG3* is another tumor suppressor gene, which is significantly down-regulated in all germ layer lineages (**Fig. 11**), but also in solid tumors such as colorectal and breast cancer [231]. Interestingly, one of its target genes is BMP4 [232], a cytokine that is crucial for the development of many tissue progenitors [233]. Including *RMRP* [234], six out of the eight most up- and down-regulated lncRNAs are tumor suppressors or oncogenes, which supports the connection of stem cell self-renewal and cancer cell proliferation. Studies on protein-coding genes have shown that both processes are regulated by common pathways [235] and the data presented here hints that this notion applies also for the non-coding part of the transcriptome.

Besides lineage-independent regulation of lncRNAs, I uncovered lncRNAs with a lineage-specific expression profile including the lncRNA *HOTAIR*, which was up-regulated in MSCs in agreement with its function in MSC differentiation [236], but also *H19*, which was highly up-regulated during lung progenitor differentiation, a process to which it has not been connected before (**Fig. 11**). *H19* is a paternally imprinted gene [237], and its transcript was shown to sequester *let-7* miRNA during muscle differentiation [61]. *H19* is expressed in some fetal organs such as heart and muscle [238] but it is severely down-regulated after birth [236], and only re-expressed in cancer cells where it acts as sponge for miRNAs or interacts with polycomb-group proteins [239]. Considering that miRNAs of the *let-7* family have crucial functions in embryonic development [240], it would be interesting to interrogate their regulation by *H19* and how this controls differentiation specifically to lung progenitor cells. A promising avenue is to analyze the connection to LIN28A, which targets *let-7* miRNAs for degradation [241] and has recently been shown to regulate lung development [242].

In summary, this study represents a comprehensive analysis of lncRNA expression trajectories during the process of human ESC differentiation, which serves as basis for an in-depth analysis of lncRNAs-of-interest for instance by performing a CRISPR screen to assign functionality during germlayer differentiation. Moreover, the intracellular localization of many developmentally-regulated lncRNAs remains unknown. It was recently demonstrated that several lncRNAs (including some of this study) form nuclear foci similar to paraspeckles [243], which might bind to multiple genomic loci, attract many RBPs or sequester miRNAs and thus regulate gene-expression and consequently cell fate transitions.

4.2 The function of paraspeckles in human cell types

Architectural lncRNAs (arcRNAs) are crucial for the stability of ribonucleoprotein aggregates [244], most prominently exemplified by *NEAT1_2*, the architectural backbone of paraspeckles [245]. Besides *NEAT1*, four other arcRNAs are known to date but only two are expressed in *H.Sapiens*, namely nucleolar intergenic spacer lncRNAs and satellite III RNAs of the nuclear stress body [244]. Amongst these, *NEAT1* was studied the most, which is due to its dynamic expression pattern in development and disease [81], however up to now, no comprehensive analysis of paraspeckle formation in human cell types was performed, which likely is due to the scarcity of material. Here, I employed PSCs to generate 8 mature and 13 progenitor cell types which were analyzed for paraspeckle content together with 3 somatic cell types. I observed that many human cell types exhibited *NEAT1_2* expression and thus paraspeckle formation as they progressed through a stage of multipotency and lineage specification (**Fig. 13**). Multipotent cells are generally restricted to differentiate into a subset of cells [246] for instance neural stem cells that can differentiate amongst others to motor neurons or astrocytes (**Fig. 12**). While the former exhibited a low amount of paraspeckles, the latter had many paraspeckles and it is interesting to speculate that the heterogeneity in paraspeckle numbers generally observed in all cell types, including neural stem cells (**Fig. 13G**), determines the cell fate transition, for instance between neurons and glial cells.

In mouse tissues, paraspeckle expression is restricted to certain organs, most notably the gut and the reproductive system [247]. Contrarily, in human differentiated cells, I identified several cell types that readily formed paraspeckles, often more than their mouse counterparts (**Fig. 14B**). This indicates a cell-type and species-specific function of paraspeckles. Astrocytes, a type of glial cells, contained many paraspeckles, which is supported by a recent study in murine brain sections that demonstrated high expression of paraspeckles in glial cells and low expression in neurons [248]. Moreover, oligodendrocytes, another glial cell type, exhibited high numbers of paraspeckles and down-regulation of *NEAT1* impaired oligodendrocyte formation [249]. I did not observe any differentiation defects of *NEAT1*$^{-/-}$ hESCs towards

astrocytes (**Fig. 21H**), however, this might be a caveat of the *in vitro* system that is saturated with differentiation cytokines, which potentially overrides a *NEAT1*-related phenotype. It is also possible that paraspeckles take on different functions in astrocytes, and for instance, are required to maintain their immunogenic functions [250]. The function of *NEAT1* in mediating the immune response is well established and many studies showed its up-regulation upon sensing an aberrant amount of dsRNA or dsDNA [129]. Moreover, *Neat1* induces the activation of inflammasomes in macrophages, thereby mediating the immune response [251]. Besides astrocytes, other paraspeckle-expressing cell types including mesenchymal stem cells, adipocytes and osteocytes are implicated in locally regulating the immune response [252]–[254] and it is plausible that paraspeckles are involved in sensing of pathogens and mediating immunogenic functions in tissues occupied by these cells. Importantly, while paraspeckles are generally up-regulated upon viral infection, those cell types already exhibit a high amount of paraspeckles and thus might be able to react faster to the pathogen invasion. It would be interesting to challenge $NEAT1^{-/-}$ cell types with a virus and to analyze their behavior compared to the parental cell line. To conclude, I have identified many human cell types with a high number of paraspeckles, which could enable a fast immune response of those cells when encountering a pathogen.

4.3 Paraspeckles could serve as potential markers for nuclear size

Cell and nucleus size are interconnected features and highly dynamic during embryonic development [255], however how size changes confer phenotype is currently not known. I showed that nucleus size correlates with the number of paraspeckles within and across different cell types (**Fig. 16B, C**). This trend, while being intuitive, has not been observed for paraspeckles before and also other subnuclear aggregates do not necessarily follow this trend. For instance, the number of nucleoli decreases with increased neuronal cell body size, whereas the number of Cajal bodies follows a similar positive correlation [256]. Paraspeckles are one of only few nuclear bodies whose formation is directly affected by RNA expression [244]. It is known that bigger nuclei exhibit higher overall transcriptional activity [257], which could explain the increase in the number of paraspeckles. Contrarily, there are cell types that do not follow this trend, including hepatocytes with large nuclei and low numbers of paraspeckles and vice versa in keratinocytes and astrocytes, indicating that there are other variables that determine paraspeckle formation. One such factor could be the proliferation rate, which is high for keratinocytes and astrocytes and low for hepatocytes. Supporting this notion, it was shown that paraspeckles are crucial for cell cycle progression [258]. How paraspeckle numbers in different cell types can be explained on molecular level remains largely unknown, however, there are probably multiple pathways and regulatory proteins involved at different

stages of development, including, as I showed, TDP-43 that inhibits paraspeckle formation in pluripotent cells (**Fig. 17**). Up-regulation of the pluripotency factors during fibroblast reprogramming had an immediate effect on paraspeckle dissolution (**Fig. 13F**) indicating that the pluripotency machinery is upstream of a paraspeckle-inhibitory network, likely acting via a TDP-43-related mechanism. To summarize, I have identified factors that can predict paraspeckle formation, which include differentiation traits, size of the nucleus and potentially proliferation rate (**Fig. 36A**).

Nuclear size correlates with paraspeckle formation, however, paraspeckles themselves do not determine nuclear size (**Fig. 16I, J**), which in turn is mainly affected by the meshwork of nuclear lamina [255], evidenced by the fact that changes in laminin expression is accompanied by reduction of nuclear size during mouse and frog development [259], [260]. Laminins are intermediate filament proteins that polymerize between the inner nuclear membrane and the peripheral chromatin [261]. Since the nuclear lamina is a tightly packed and highly interconnected network, it is plausible that paraspeckle formation is hindered in this environment due to spatial constraints. The thickness of the nuclear lamina is likely uncoupled from the nuclear size, hence cells with smaller nuclei have smaller relative nuclear inner volume (without volume occupied by nuclear lamina) compared to cells with bigger nuclei (see calculations in **Fig. 37B**). This means that the spatial constraints are even more pronounced in cells with smaller nuclei, which could explain why smaller changes in nuclear size (~2-fold change between smallest and biggest nuclei, **Fig. 16C**) could lead to dramatic changes in the number of paraspeckles (~9-fold change between lowest and highest number of paraspeckles, **Fig. 16C**).

Nuclear size is also connected to cancer and is used by cytopathologists as parameter for prognosis and to predict its stage of progression [262]. To date, the expression of only a hand-full of proteins are known to correlate with nuclear size, amongst others the laminin-like proteins LINC1 and LINC2 in *Arabidopsis thaliana* [263] and importin α during *Xenopus* development [264], however, mammalian factors are not known. I propose that the number of paraspeckles can be used as a determinant of mammalian nuclear size. Especially since *NEAT1* expression and number of paraspeckles are well correlated (**Fig. 13E**), nuclear size could be predicted after assessment of single-cell RNA sequencing data, given that the long version of *NEAT1* was measured. This way, a wealth of information regarding mammalian nuclear size in development and disease can be generated, thereby providing new insight into the connection between nuclear architecture and crucial biological processes.
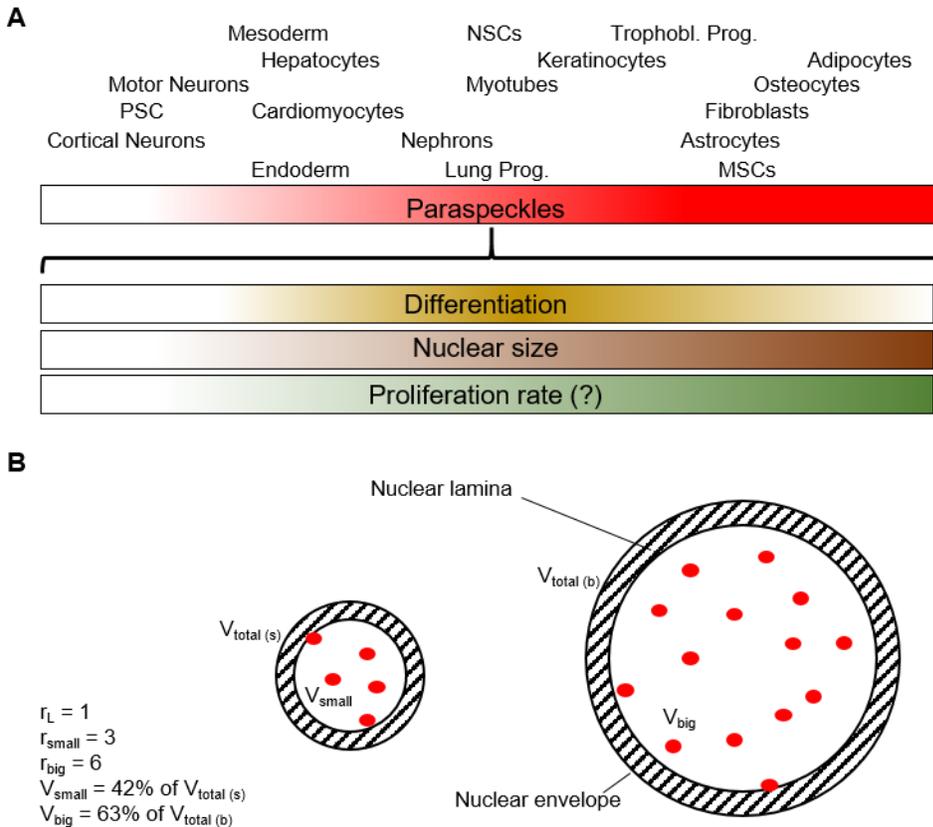
Figure 36: Paraspeckle formation can be predicted by differentiation traits and nuclear size. A) Factors that predict paraspeckle formation. Proliferation rate seems to be positively correlated with paraspeckle formation, however, a detailed analysis is missing. Cell types are positioned according to their number of paraspeckles from left (low amount) to the right (high amount). **B)** Connection between nuclear size, lamina and paraspeckles. Red circles represent paraspeckles. Assuming a spherical shaped nucleus with a constant radius of nuclear lamina (r_L), the inner volume of small nuclei that paraspeckles can occupy is only 42% of the actual nuclear volume, whereas it is 63% in big nuclei with doubled inner radius.

4.4 Double-strand DNA stabilizes lncRNA foci

Paraspeckle formation depends on interaction of *NEAT1_2* and essential paraspeckle core proteins, which bind and stabilize the *NEAT1_2* transcript [81], however, the connection of paraspeckle ribonucleoprotein particles to dsDNA is not well understood. Evidence for co-transcriptional assembly of paraspeckles stems from the fact that a subpopulation of paraspeckles is found at the *NEAT1* genomic locus [265], and that artificial tethering of *NEAT1* to another locus is sufficient for the assembly of paraspeckle proteins [266]. A dimer of the core paraspeckle proteins SFPQ and NONO binds first to a subdomain of *NEAT1_2*, before inducing polymerization, likely via their coiled-coil domain [136], to cover the whole *NEAT1_2* lncRNA. Other essential paraspeckle proteins such as FUS and RBM14 are recruited, a process that is driven by liquid-liquid phase separation [245]. Paraspeckle integrity may also depend on RNA-RNA interactions,

which have been recently demonstrated to be crucial for granule formation *in vitro* [267]. Indeed, structural mappings revealed numerous RNA-RNA interactions within *NEAT1_2*, which could occur in paraspeckles [268]. To summarize, while the paraspeckle assembly at the *NEAT1* locus is known reasonably well, it is not entirely clear, how the *NEAT1*:RBP complex relocates and becomes embedded elsewhere in the chromatin.

NEAT1 possesses putative DNA binding domains that were implicated in forming DNA-RNA triple helix structures in the major groove [269]. By addition of small DNA binding molecules that cause conformational changes of the dsDNA [194], I sought to analyze the connection between paraspeckle formation and DNA integrity. Intriguingly, these molecules were able to disintegrate paraspeckles and splicing speckles within 30 minutes. A similar observation has been made by Sasaki and colleagues after 4 hours of treatment by Actinomycin D, however, the authors assumed that this is due to transcriptional inhibition [85], whereas I showed that it is the DNA binding ability of the molecule that causes the disintegration (**Fig. 18**). This provides further evidence of the direct interaction between lncRNAs and the dsDNA. Moreover, this indicates that in addition to RNA and proteins, DNA itself is fundamentally important for the aggregation of chromatin-embedded lncRNAs. It cannot be excluded that the reagents directly disrupt the RNA-RNA interaction between *NEAT1_2* molecules and thereby induce paraspeckle disintegration, however so far there is no evidence that ActD, Hoechst, Flavopiridol and others bind to RNA. Interestingly, paraspeckles in differentiated *NONO*$^{-/-}$ hESCs tend to be much smaller and look somewhat similar to *NEAT1_2* foci after treatment by small DNA binding molecules. It is possible that the displacement of NONO and other essential paraspeckle proteins to the perinucleolar caps after treatment by DNA-binding molecules (**Fig. 18C**) is the cause for paraspeckle disintegration. It remains to be tested, whether this displacement is due to the disruption of the *NEAT1*:dsDNA triplex structure.

These findings also suggest a tantalizing connection to chemotherapy treatments given to cancer patients. Many chemotherapeutic reagents bind to DNA often as a byproduct when high concentrations are used, as shown for the CDK inhibitor Flavopiridol [207]. Actinomycin D, Etoposide, and Mithramycin A, which readily disintegrate paraspeckles, are commonly used chemotherapies for example in the treatment of osteosarcoma [202], [270]. I have implicated these reagents here for the first time in the dissociation of lncRNAs from the chromatin, which plausibly is a general phenomenon. This raises the possibility that the mechanism-of-action and, or, the side effects of common chemotherapies that bind dsDNA are connected to genome-wide disruption of lncRNA architecture.

4.5 *NEAT1* **has alternative, developmentally regulated transcription start sites**

It is well established that *NEAT1* has alternative transcription termination sites that produce either the short isoform by polyadenylation, or the long isoform by RNase P cleavage [84]. By genomic deletion of parts of the *NEAT1* promoter, I showed that a consecutive increase in size of the deleted region led to a step-wise decrease of *NEAT1* transcription in differentiated hESCs (**Fig. 20B**), indicating that besides alternative termination sites, *NEAT1* also contains alternative transcription start sites. Deletion of similar regions in HEK293 cells induced stronger down-regulation of *NEAT1* expression (**Fig. 20B, E**) indicating that alternative *NEAT1* start sites exist in hESCs, but not in somatic HEK293 cells. Moreover, in contrast to HEK293 cells [95], the insertion of a *YFP-polyA* stop cassette directly after the *NEAT1* TSS was not sufficient to eliminate paraspeckles (**Fig. 20C**). On the other hand, insertion of a stop cassette 1500 bp after the TSS fully down-regulated paraspeckles in hESCs (**Fig. 22**), which is evidence that the alternative *NEAT1* transcription start site(s) is located within the first 1500 bp of *NEAT1* gene. For protein-coding genes, alternative transcription start sites result in differences in the length of the first exon [271], which respectively gives rise to different protein isoforms or has an impact on translational control [272]. Alternative transcription of *NEAT1* results in shortening of both isoforms, which might not be meaningful for the 23 kb long, architectural isoform, but might be relevant for the processing of the 3.7 kb short isoform. Indeed, the aggregation potential of *NEAT1_1* seems to be impaired in *NEAT1^STOP^* hESCs, which express only a 1.5 kb long isoform of *NEAT1_1* that is able to form foci, which however, appeared more dispersed than full-length *NEAT1_1* foci (**Fig. 22A**). It would be interesting to analyze if, besides the canonical short and long isoforms, hESCs express other isoforms of *NEAT1*, or whether the usage of alternative start sites is caused by the genomic deletion of the canonical TSS. Moreover, the use of alternative *NEAT1* TSSs could exhibit tissue-specificity, which was shown for protein-coding genes due to the usage of different promoters and enhancer regions [273], [274]. In all, hESCs exhibit alternative *NEAT1* transcription start sites that might be absent in somatic cells.

4.6 Paraspeckles exhibit phenotypic differences in mouse and human ESCs

Paraspeckles are up-regulated in various developmental processes and in general during cellular stress including proteasome inhibition, hypoxia and viral infection (**Fig. 6**), however, its function during these processes remains enigmatic. Here, I showed that *NEAT1_2* was up-regulated during human germ layer differentiation (**Fig. 11B**) concomitant with the formation of paraspeckles (**Fig 13F**). Interestingly, *NEAT1* is one of only 18 genes that is up-regulated no matter what differentiation-inducing cytokine is added to the cell [47], which is why I initially reasoned that *NEAT1* has a lineage-independent effect on

differentiation. A similar up-regulation was observed in differentiated mouse ESCs, which is the system that we (in collaboration with my PhD colleague Miha Modic) tackled first to analyze a potential developmental phenotype of paraspeckles. We generated mouse ESCs that harbored a deletion of the triple helix located at the 3`end of *NEAT1_2*, which had a ~60% reduction of paraspeckles, similar to *NEAT1ΔTH* human ESCs (**Fig. 20**). Strikingly, differentiated murine *NEAT1ΔTH* ESCs exhibited elevated levels of SSEA1, a pluripotency surface marker in mESCs, and also minor up-regulation of *Nanog*, indicating that paraspeckle depletion impairs exit from pluripotency [47]. Moreover, by performing a 2n-4n aggregated mouse complementation assay [275], embryos arising from *NEAT1ΔTH* mESCs exhibited defects in primitive streak formation [47]. Unexpectedly, I observed the opposite trend in differentiated human ESCs that harbored a similar deletion in the triple helix where pluripotency markers were down-regulated relative to wildtype. These findings were validated in *NEAT1^-/-* hESCs with a complete absence of paraspeckles, which had the same phenotype as *NEAT1ΔTH* hESCs (**Fig. 21**). Furthermore, paraspeckle depletion by insertion of an expression stop cassette also induced down-regulation of pluripotency genes upon differentiation (**Fig. 22G-J**), in all indicating that the phenotype of human *NEAT1* is reversed compared to murine ESCs. In general, it is not unusual that the functions of mouse and human orthologues diverge, which has been shown for many protein-coding genes [276] and might be true for non-coding transcripts that generally evolve faster than protein-coding genes [50]. It is well established that mouse ESCs are in a different developmental stage compared to human ESCs [26], which means that the conditions supporting self-renewal and differentiation differ greatly. It is also plausible that the paraspeckle composition of human and mouse ESCs is fundamentally different, and until today no systematic analysis of paraspeckle content in murine cells was executed. Furthermore, it was shown that the half-life of mouse and human *NEAT1_2* is quite different (1 h vs 5 h, [199]) indicating different processing of *NEAT1_2* in both species. The relatively unstable murine *NEAT1_2* transcript could result in paraspeckles with a high turn-over rate where paraspeckle proteins are sequestered only transiently, which could lead to differences in the pluripotency-differentiation transition. Finally, it has to be pointed out that a thorough analysis of paraspeckle-depleted mESCs is still missing and should be done side-by-side with *NEAT1^-/-* hESCs in similar conditions to untangle species-specific functions of paraspeckles.

In general, it is unexpected that paraspeckles delay the exit from pluripotency instead of promoting it, as their expression profile would suggest. It is important to point out that the phenotype was observed only after initial differentiation when a few paraspeckles had already formed. I hypothesize that paraspeckles serve as a control mechanism for the cell to fine-tune its differentiation, however, their loss can be compensated by hESCs after prolonged differentiation, which is in line with the fact that *Neat1^-/-*

mice do not exhibit an overt developmental phenotype [247]. Furthermore, the phenotype was primarily observed in spontaneous or neural differentiated cells but not in other germ layers, indicating a germ layer-specific function of *NEAT1*. It is possible that the cytokines added to the various differentiation protocols mask a potential *NEAT1*-related phenotype *in vitro* and only the neural lineage, which was induced not by adding activating cytokines, but only inhibitory molecules of the TGFbeta pathway resulted in an enhanced exit from pluripotency. It would be interesting in the future to analyze the paraspeckle content across different germ layers, which I speculate could be quite different.

Two genes seemed most affected by changes in *NEAT1_2* expression, namely *NANOG* and *PAX6*, which are respectively down- or up-regulated in differentiated *NEAT1_2* knock-out hESCs. *NANOG* is a core pluripotency transcription factor, which, in high levels, allows the feeder-free culture of hESCs [277]. Its expression is regulated by the pluripotency factors OCT4 and SOX2 [278], which are activated by the FGF2 and ACTIVIN/NODAL signaling pathways [23]. Paraspeckles, on the other hand seemed to be inhibited by the activation of those pathways and up-regulated by other signaling cascades such as the Wnt3A and BMP4 pathways. How paraspeckles affect *NANOG* expression remains unclear. It is possible that they bind in close vicinity of the *NANOG* locus and attract transcriptional activators to counteract to some extent the down-regulation of *NANOG* induced by the loss of pluripotency signaling pathways. Another possibility is that paraspeckles contain transcriptional inhibitors of *NANOG*, which are released in *NEAT1*$^{-/-}$ hESCs to induce down-regulation of *NANOG* compared to wildtype. PAX6, on the other hand, is a transcription factor that induces neuroectoderm differentiation of hESCs and its up-regulation is correlated with down-regulation of *NANOG* [279]. Whether *PAX6* expression is directly inhibited by NANOG is not known but is possible that *PAX6* up-regulation is the result of reduced *NANOG* expression induced by the loss of paraspeckles. Interestingly, the signaling pathways that maintain *NANOG* expression in mESCs are very different compared to hESCs, which in mice is maintained primarily by the LIF/STAT3 and BMP4 signaling cascades [280]. This might be another explanation for the phenotypic differences that arise upon paraspeckle knock-out.

To summarize, in humans, *NEAT1_2* is modulating the early onset of differentiation by mediating the coordinated down-regulation of the pluripotency machinery. The mechanism behind this phenotype remains to be analyzed in future studies to understand the differences between paraspeckle biology in mouse and human development.

4.7 The function of the short *NEAT1* isoform

While *NEAT1_2* foci are generally up-regulated during differentiation, *NEAT1_1* foci exhibit a reciprocal trend and their disappearance correlates with the onset of differentiation (**Fig. 15B**). This is probably due to the formation of paraspeckles by *NEAT1_2*, which, when in high abundance, sponge and remove *NEAT1_1* foci from the nucleoplasm. Supporting this notion, cell types with high numbers of paraspeckles generally do not have *NEAT1_1* foci (**Fig. 15E**). The existence of *NEAT1_1* outside of paraspeckles was recently demonstrated in U2OS cells that form *NEAT1_1* foci containing 1-3 molecules, termed microspeckles [95]. By eye, *NEAT1_1* foci in PSCs seemed much bigger and to contain many more molecules, however a thorough analysis of *NEAT1_1* content was not performed, yet.

To analyze the function of *NEAT1_1*, I generated *NEAT1ΔpA* hESCs, which do not express *NEAT1_1* (**Fig. 20G, H**) due to the removal of the *NEAT1_1* polyadenylation site that is required for *NEAT1_1* production. While *NEAT1_1* is dispensable for paraspeckles as shown by the fact that *NEAT1ΔpA* hESCs readily formed paraspeckles, it is nevertheless possible that *NEAT1_1* functions as a "helper molecule", maybe by acting as seeding nucleus for the aggregation of *NEAT1_2* in differentiated cells. This could be addressed by over-expression of *NEAT1_1* in *NEAT1ΔpA* hESCs. It has been demonstrated that the short isoform is able to recruit paraspeckle proteins [266], which in turn could attract *NEAT1_2* upon its expression in differentiated cells. The question remains what could be the function of the short *NEAT1* aggregates in pluripotent cells. *NEAT1ΔpA* hESCs exhibit both up-regulation of *NANOG* and of differentiation genes (**Fig. 25D**), however, it is difficult to assess, whether this is due to the absence of *NEAT1_1* or the up-regulation of *NEAT1_2* and paraspeckles. Since *NANOG* up-regulation was observed after over-expression of endogenous *NEAT1_2* (**Fig. 23C-E**), whereas differentiation genes were not affected (data not shown), the latter might have been caused by the absence of *NEAT1_1* foci in pluripotent cells. In mice, the 2n-4n complementation assay of *NEAT1ΔpA* mESCs did not compromise the primitive streak formation [47] indicating that over-expression of *NEAT1_2* and loss of *NEAT1_1* are of no obvious developmental consequence. This was recently supported by the generation of *NEAT1ΔpA* mice that did not exhibit any overt phenotypical changes compared to the wildtype strain [281], [282]. Even though a high amount of *NEAT1_1* foci was observed in the G1 cell cycle phase, the deletion of *NEAT1_1* did not affect the cell cycle or proliferation of murine cells. Moreover, only a few genes were differentially expressed in *NEAT1ΔpA* U2OS cells, which substantiates the fact that *NEAT1_1* on its own is not able to affect gene expression [281]. The authors suggest that *NEAT1_1* itself is not functional but the cells need to constantly express it to ensure a rapid isoform switch to *NEAT1_2* when the cells are under stress [281]. This could be similar during exit from pluripotency when the addition of differentiation cytokines triggers

an isoform switch and the onset of paraspeckle formation. To summarize, PSCs exhibit formation of *NEAT1_1* foci, which disappear in differentiated cells. Deletion of *NEAT1_1* induced up-regulation of differentiation genes in human ESCs, indicating that these cells are primed for differentiation, however, the differentiation propensity of *NEAT1_1*-depleted human and mouse ESCs was not affected. It is possible that *NEAT1_1* expression represents the "cost" of the cell to have an active *NEAT1* locus to react fast to changes in the environment by up-regulation of *NEAT1_2* and paraspeckles.

4.8 Developmental paraspeckles regulate global splicing by RNA and protein retention

Paraspeckles are very complex structures containing, besides *NEAT1*, a plethora of RNAs and many proteins that together mediate the function of paraspeckles [81]. Given that ~25% of the 40 paraspeckle proteins in HeLa cells [247] are not expressed in hESCs or their differentiated progeny (based on transcriptome analysis; data not shown), the composition of developmental, "nascent" paraspeckles is likely quite different compared to paraspeckles in tumor and other somatic cells. Due to the vicinity of paraspeckles and splicing speckles [109], many paraspeckle proteins have annotated function in splicing regulation [83] together with general functions in transcription regulation and polyadenylation (**Fig. 4A**). Interestingly, this is quite similar to the repertoire of RBPs bound transiently to the RNA-PolII to mediate co-transcriptional processes [283], [284]. I hypothesized that paraspeckles and the RNA-PolII share a common repertoire of RBPs, which could mean that paraspeckle-depleted hESCs exhibit changes in the composition of RBPs attached to RNA PolII. Indeed, I identified mainly splicing-associated proteins that are enriched in wildtype compared to *NEAT1*-/- RNA-PolII (**Fig. 27**), which indicates that the depletion of paraspeckles causes protein translocations. It would be intriguing to untangle the molecular connection between RNA-PolII and paraspeckles. *NEAT1* itself is transcribed by RNA-PolII [285] and given the fact that paraspeckles already assemble at the *NEAT1* locus, proteins may be transferred from the CTD of RNA PolII to the nascent *NEAT1_2* transcript. This transfer might be compromised in differentiated *NEAT1*-/- hESCs, which could explain the enrichment of splicing proteins in wildtype compared to paraspeckle-depleted cells. By chromatin pull-down studies, a subpopulation of *NEAT1* was found to overlap with H3K4me3 regions, which mark regions of actively transcribed genes [109]. Although the PolII complex itself was not identified in paraspeckles, it is plausible that the transcribing RNA-PolII encounters paraspeckles outside of the *NEAT1* locus and it would be intriguing to analyze the spatiotemporal relationship between those two macromolecular complexes during active transcription.

Several RNAs have been identified in paraspeckles and their retention represents a means of the cells to inhibit translation [84], [97]. It has been hypothesized that paraspeckles contain many other

transcripts, especially those with inverted-repeat *Alu* elements in their 3`UTR [286], which prompted me to analyze global changes in RNA localization upon paraspeckle depletion. Intriguingly, I identified ~30% of transcripts that exhibited changes in their cytoplasmic or nuclear localization in paraspeckle-depleted hESCs (**Fig. 26D**). Interestingly, the proportion of RNAs that were retained in the nucleus was similar to the number of RNAs exported from the nucleus in differentiated *NEAT1$^{-/-}$* hESCs, arguing against the hypothesis that paraspeckles are a global hub of RNAs. Of note is the high proportion of mRNAs that were generally enriched in the nucleus, which is in line with recent findings by Halpern *et al.* who identified mature mRNAs in the nucleus of B- and liver cells and proposed a function in the reduction of cytoplasmic gene expression noise [287]. RNAs that are more exported in *NEAT1$^{-/-}$* hESCs are associated with mRNA processing, mostly by regulation of splicing (**Fig. 26G**), which indicates that the splicing process is affected in differentiated *NEAT1$^{-/-}$* hESCs due to a) changes in splicing proteins associated with RNA-PolII (section 4.7) and b) changes in the localization of transcripts encoding for splicing proteins. The notion that paraspeckles regulate splicing was demonstrated by a recent study that showed changes in alternative splicing of genes involved in axonogenesis and neuronal homeostasis in the cerebellum of *Neat1$^{-/-}$* mice [248]. Moreover, it was shown in murine cells that *Neat1* associates with the splicing factor SRp40 to regulate alternative splicing of *Pparg*, an essential gene for adipogenesis [288].

In conclusion, after the depletion of paraspeckles, I identified changes in the subcellular localization of RNAs and proteins involved in splicing regulation but how this leads to the phenotype described above remains to be analyzed in detail. The fact that paraspeckles are such complex granules that can mediate gene expression by chromatin binding, RNA retention or protein sequestration (described in section 1.7) aggravates mechanistic studies on a molecular level. Further studies are required to identify the protein components and genomic binding sites of developmental paraspeckles and address how this is connected to the regulation of pluripotency genes.

4.9 The function of DBHS proteins in the pluripotency-differentiation transition

DBHS proteins are core paraspeckle components [191] that control transcription and RNA processing [135]. All three members of the DBHS family, SFPQ, NONO and PSPC1, are highly expressed in undifferentiated hESCs (**Fig. 28A, B**) indicating that they have a paraspeckle-independent function in pluripotency maintenance that has not been addressed, yet. Moreover, even though expression of these proteins does not overtly change during stem cell differentiation (**Fig. 28C**), the induction of paraspeckles induces the sequestration of DBHS proteins (**Fig. 28D, E**), which might affect their function in the nucleus. By genomic deletions, I sought to analyze the function of DBHS proteins in pluripotent and differentiated hESCs,

initially with the idea to pinpoint the paraspeckle phenotype described above to the misregulation of SPFQ, NONO and PSPC1.

4.9.1 SFPQ regulates ESCs homeostasis via post-transcriptional processes

SFPQ is a multifunctional RBP that is essential for the survival of thymocytes [145] and zebrafish embryos [289]. In line with these findings, I observed cell death in hESCs where coding parts of the gene were deleted (**Fig. 29D**). Knock-down of *SFPQ* was however tolerated and resulted in the down-regulation of genes that are essential for cellular homeostasis (**Fig. 34C**) without changing the pluripotency characteristics of the cell (**Fig. 29M**). Interestingly, the overlap of changes in the transcriptome after *SFPQ* knock-down and *NONO* knock-out hESCs was minimal (**Fig. 34B**), even though both proteins strongly interact with each other [135] and do so also in hESCs (**Fig. 34D**). This suggests that SFPQ has NONO-independent functions, which hints to the existence of a population of SFPQ protein that is not bound to NONO. It was shown before that the loss of one DBHS protein can be compensated by the up-regulation of another one [137], however here I did not observe any changes in the expression of DBHS genes upon depletion of SFPQ, NONO or PSPC1 (**Fig. 29K**). Nevertheless, it is possible that after down-regulation of SFPQ, PSPC1 or NONO, the two remaining DBHS proteins form complexes with different stoichiometry and functions, which could explain the low overlap in the transcriptional profile of *SFPQ* knock-down and *NONO*$^{-/-}$ hESCs. By co-immunoprecipitation, I observed an enrichment of splicing proteins, which was expected given that SFPQ was originally identified in a splicing complex [290]. To my surprise, however, I identified many proteins of the polyadenylation machinery that interact with SFPQ (**Fig. 34F**), which could represent another function of SFPQ that has been much less characterized. One hint that SPFQ promotes polyadenylation was provided recently by tethering SFPQ adjacent to a weak polyadenylation signal in the 3`UTR of the *COX-2* transcript, which was then polyadenylated [291], however, the mechanistic details are lacking. The pull-down of almost every member of the canonical polyadenylation machinery suggests that SFPQ regulates the polyadenylation globally or of a subset of genes. Interestingly, next to being enriched at the 5`and 3`splice sites, SFPQ seems to be bound to the 3`termination sites [292], [293], a fact that was neglected so far by the authors of these studies. It would be interesting to analyze changes in the polyadenylation profile upon SFPQ depletion and to integrate this data with SFPQ iCLIP data to analyze which genes might be polyadenylation targets of SFPQ. To summarize, SFPQ is an essential gene for the survival of human ESCs and might act via modulating the polyadenylation and 3`end processing of genes crucial for cell survival.

4.9.2 NONO maintains pluripotency by regulating the expression of cholesterol synthesis enzymes

NONO is a strong binding partner of SFPQ (**Fig. 34D**), however, the gene expression profile upon NONO depletion is quite different compared to differentially expressed genes after SFPQ down-regulation (**Fig. 34B**). In contrast to SFPQ, the knock-out of NONO affects the pluripotency state of hESCs, indicated by the up-regulation of lineage specification markers *PAX6*, *T*, *MIXL1*, *SOX17* and *FOXA2* (**Fig. 29M**). This is evidence that NONO is required for pluripotency maintenance of hESCs by inhibition of differentiation-inducing genes. Strikingly, an opposite phenotype was observed in mouse ESCs, which are more naïve and less likely to differentiate after the knock-out of *Nono* [222]. Here, I did not observe any changes in the neuronal differentiation of *NONO*-depleted hESCs (**Fig. 30E, F**) which was shown to be compromised in the mouse [222]. This suggests that similar to paraspeckles, deletion of *NONO* has the opposite outcome in human compared to mouse ESCs. Reasons for that could be the inherent differences between mouse and human ESCs [26], as discussed above, or that NONO has different species-specific binding partners and genomic binding preferences. On the other hand, naïve *NONO*⁻/⁻ hESCs, which should be more similar to mouse ESCs, did not recapitulate the phenotype of mESCs (**Fig. 33E**) suggesting that differences in the developmental state of human and mouse ESCs cannot explain the reverse phenotype.

A human-specific function of NONO could be its regulation of cholesterol synthesis genes in hESCs (**Fig. 30B, C**) and in cancer cell lines (Archa Fox; personal communication). Strikingly, all enzymes that are required for the cholesterol synthesis are down-regulated by approximately 2-fold suggesting that *NONO*-depleted hESCs produce less cholesterol compared to wildtype. Metabolic analysis of *NONO*⁻/⁻ hESCs is currently being performed to analyze changes in the cholesterol derivatives. Cholesterol is a main component of the cell wall [294] and a precursor for steroid hormones such as testosterone, estrogen and corticoids [295]. It was shown recently that these hormones are required for mesoderm specification of mESCs [296] and that cholesterol depletion resulted in neuronal cell death in the mouse cerebellum [297]. In humans, not much is known about the function of cholesterol in stem cell maintenance and differentiation. Moussaieff *et al*. reported that a biochemical inhibition of cholesterol-induced a small increase in pluripotency gene expression in differentiating hESCs [221]. This suggests that the phenotype of *NONO*-depleted hESCs is unrelated to the down-regulation of the cholesterol pathway. Nevertheless, the reduced expression of cholesterol could be masked by the stem cell maintenance medium that contains an excess of small metabolites such as sodium pyruvate, which is converted to acetyl-CoA, the starting molecule for the cholesterol synthesis [298]. It remains unknown whether the down-regulation of cholesterol synthesis is responsible for changes in the pluripotency characteristics of *NONO*⁻/⁻ hESCs.

Finally, others have shown that NONO is required for the stability of *NEAT1_2* and that down-regulation of *NONO* led to the complete absence of paraspeckles [85]. Here, I showed that differentiated *NONO*[-/-] hESCs exhibited down-regulation of paraspeckles by ~70% concomitant with their disintegration to microspeckles (**Fig. 35B**). It is possible that increased expression of differentiation-associated gene in differentiated *NONO*[-/-] hESCs (**Fig. 30D**) is in part due to the down-regulation of *NEAT1_2*, which in turn increased differentiation, mostly by down-regulation of pluripotency genes, but also by up-regulation of differentiation genes *FOXA2* and *PAX6*. Moreover, the phenotype of *Nono* and paraspeckle knock-out ESCs in mouse is similar [47], [222], which in all suggests that paraspeckles are a downstream mediator of NONO function and responsible for differences in phenotypic outcome. Given that paraspeckles itself are complex structures of 40 or more proteins, it is likely that their composition and possibly genomic binding preferences are quite different in mouse and human cells, which could explain the contrary phenotype observed in both species. It would be interesting to analyze whether other essential paraspeckle proteins such as RBM14, EWSR1, HNRNPK or DAZAP1 [81] exhibit phenotypic differences in mouse and human ESC differentiation.

In conclusion, NONO regulates expression of cholesterol synthesis genes in hESCs, which might be the cause for the developmental phenotype observed in *NONO*[-/-] hESCs. Both NONO and paraspeckles exhibit a complementary phenotype in mouse and human ESCs which indicates that paraspeckles are in part responsible for conveying NONO gene-regulatory functions.

4.9.3 The function of PSPC1 in PSCs and during adipogenesis

The third member of the human DBHS protein family is PSPC1, which in contrast to SPFQ and NONO, is dispensable for paraspeckle formation [191]. A screen for pluripotency and differentiation genes revealed the down-regulation of pluripotency genes *OCT4*, *SOX2* and *NANOG* by ~50-60%, concomitant with minor up-regulation of lineages-specifying markers in *PSPC1*[-/-] hESCs (**Fig. 29L, M**). Similarly, it was reported that the expression of many genes involved in developmental processes was changed in murine *PSPC1*[-/-] ESCs [299]. Interestingly, these genes are mostly enriched in markers of the 2-cell-like totipotent stage, including MERVL, a family member of endogenous retroviruses (ERV), which PSPC1 regulates together with the epigenetic DNA-modifying protein TET2. It is not known whether PSPC1 takes on a similar role in human ESCs. *PSPC1*[-/-] hESCs exhibited increased expression of naïve genes and less primed surface markers during the conversion towards naïve cells (**Fig. 33**) indicating that PSPC1 is required for the naïve-to-primed conversion that happens during the development of the blastocyst to the epiblast stage embryo [300]. Whether this transition is also mediated by interaction with TET2 remains to be analyzed. While the

differentiation characteristics of *PSPC1*[-/-] hESCs remain largely intact upon initial germ layer commitment (**Fig. 32**), I observed impaired MSC and adipocyte differentiation (**Fig. 31**), which is in line with recent findings that demonstrated the importance of PSPC1 during murine adipogenesis [223]. However, while this publication reported down-regulation of master adipocyte regulators *Pparg*, *aP2* (*Fabp4*) and *Adiponectin*, *PSPC1*[-/-] human adipocytes were mostly similar in their gene expression profile to the parental cell line except for the down-regulation of the fatty-acid chaperone *FABP4*. Besides *NEAT1_2* and NONO, this is the third paraspeckle component whose function diverges to some extent compared to the mouse orthologue. In summary, PSPC1 represents another paraspeckle protein that influences human stem cell differentiation, mainly towards adipocytes.

4.10 Conclusion and outlook

Paraspeckles are complex structures that have attracted a lot of interest over the last years because they represent phase-separated, dynamic granules that are up-regulated when cells or the organism are subjected to developmental changes or stress. Germ layer differentiation is accompanied by paraspeckle formation which raised the question about their function in this process. A thorough dissection of ESCs that exhibited down-regulation or complete absence of *NEAT1* isoforms or the paraspeckle core protein components revealed that *NEAT1_2*, NONO and PSPC1 generally are required for coordinated exit from pluripotency whereas *NEAT1_1* and SPFQ do not change the gene expression of pluripotency and differentiation genes (**table 1**). Strikingly, the phenotype of *NEAT1_2*, NONO and PSPC1 is different compared to the mouse orthologues, which raises the question whether this is due to the *in vitro* model of development or whether these paraspeckle components truly exhibit human-specific functions. A more humanized model of development, which are currently being developed [301] is needed to address this question. Given that deletion of paraspeckle components can be compensated by prolonged differentiation and that mature cell types did not depict an overt phenotype, it is likely that paraspeckles are not essential for human development but might be needed when the organism is challenged by external stress such as pathogen infection.

Paraspeckles represent membraneless organelles that can exchange their content with the environment in a dynamic manner [81]. Principles of granule formation are objective of many studies since certain membraneless aggregates were hypothesized to drive the pathogenesis of neurodegenerative diseases and cancer [302]. These studies focused mostly on the contribution of proteins [76] and RNAs [267] that are sought to come together by phase separation. Here, I showed that small DNA binding molecules disintegrate paraspeckles and other DNA-bound granules, which puts emphasis on a previously neglected

molecule for the formation of granules. DNA likely provides a scaffold for the aggregation of paraspeckles, which bind directly via triple helix formation of *NEAT1* with the double-strand DNA [111], [303]. It is interesting now to analyze this phenomenon in the context of tumor treatment, which is mostly achieved by administration of a chemotherapeutic cocktail containing the same or similar DNA binding molecules that have been shown here to disintegrate paraspeckles. It is possible that the disintegration of chromatin-bound lncRNAs is the first event that leads to subsequent cell death.

PSCs represent a powerful tool to investigate developmental processes or molecular mechanisms of diseases. In this study, I have generated human pluripotent cell lines that exhibited deletions in many functional parts of *NEAT1* or paraspeckle core proteins. While the exit from pluripotency and germ layer differentiation was analyzed in detail in those cells, due to the versatility of PSCs, it is possible now to study paraspeckles and their components in any cell type-of-interest, which represents a valuable asset for the paraspeckle community. In all, this study represents a comprehensive analysis of lncRNA trajectories during germ layer differentiation with focus on paraspeckles where I demonstrated novel mechanisms of formation, function and principles of their regulation in human PSCs, mature cell types and upon reprogramming.

Table 1: Effect of paraspeckle components on hESC maintenance and differentiation.

Paraspeckle components	Pluripotency genes in KOs	Differentiation genes in KOs
NEAT1_1	Up-regulated	Up-regulated
NEAT1_2	Down-regulated	PAX6 up-regulated
SFPQ	Unchanged	Unchanged
PSPC1	Down-regulated	Up-regulated
NONO	Unchanged	Up-regulated

5. Literature

[1]   A. Christianson and C. P. Howson, "GLOBAL REPORT O N BIRT H DEFECTS," p. 18.

[2]   D. Cyranoski, "First CRISPR babies: six questions that remain," *Nature*, Nov. 2018, doi: 10.1038/d41586-018-07607-3.

[3]   K. K. Niakan, J. Han, R. A. Pedersen, C. Simon, and R. A. R. Pera, "Human pre-implantation embryo development," *Development*, vol. 139, no. 5, pp. 829–841, Mar. 2012, doi: 10.1242/dev.060426.

[4]   J. A. Thomson *et al.*, "Embryonic stem cell lines derived from human blastocysts," *Science*, vol. 282, no. 5391, pp. 1145–1147, Nov. 1998, doi: 10.1126/science.282.5391.1145.

[5]   K. N. Schulz and M. M. Harrison, "Mechanisms regulating zygotic genome activation," *Nat Rev Genet*, vol. 20, no. 4, pp. 221–234, Apr. 2019, doi: 10.1038/s41576-018-0087-x.

[6]   Hertig AT. Rock J. Adams EC. and Mulligan W.J. "On the preimplantation stages of the human ovum: a description of four normal and four abnormal specimens ranging from the second to the fifth day of development". (1954) Carnegie Instn. Wash. Publ. 603, Contrib. Embryol., 35: 199-220.

[7]   S.-M. Kim and J.-S. Kim, "A Review of Mechanisms of Implantation," *Dev Reprod*, vol. 21, no. 4, pp. 351–359, Dec. 2017, doi: 10.12717/DR.2017.21.4.351.

[8]   A. J. Wilcox, D. D. Baird, and C. R. Weinberg, "Time of Implantation of the Conceptus and Loss of Pregnancy," *New England Journal of Medicine*, vol. 340, no. 23, pp. 1796–1799, Jun. 1999, doi: 10.1056/NEJM199906103402304.

[9]   B. Bénazéraf and O. Pourquié, "Formation and Segmentation of the Vertebrate Body Axis," *Annual Review of Cell and Developmental Biology*, vol. 29, no. 1, pp. 1–26, 2013, doi: 10.1146/annurev-cellbio-101011-155703.

[10]   D. Purves *et al.*, "The Initial Formation of the Nervous System: Gastrulation and Neurulation," *Neuroscience. 2nd edition*, 2001.

[11]   M. Hill, "Early Human Development," *Clinical Obstetrics and Gynecology*, vol. 50, no. 1, pp. 2–9, Mar. 2007, doi: 10.1097/GRF.0b013e31802f119d.

[12]   I. Schneider and J. Ellenberg, "Mysteries in embryonic development: How can errors arise so frequently at the beginning of mammalian life?," *PLOS Biology*, vol. 17, no. 3, p. e3000173, Mar. 2019, doi: 10.1371/journal.pbio.3000173.

[13]   M. Meskus and I. de Miguel Beriain, "Embryo-like features of induced pluripotent stem cells defy legal and ethical boundaries," *Croat Med J*, vol. 54, no. 6, pp. 589–591, Dec. 2013, doi: 10.3325/cmj.2013.54.589.

[14]   I. Chambers and S. R. Tomlinson, "The transcriptional foundation of pluripotency," *Development*, vol. 136, no. 14, pp. 2311–2322, Jul. 2009, doi: 10.1242/dev.024398.

[15]   H. Niwa *et al.*, "Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation," *Cell*, vol. 123, no. 5, pp. 917–929, Dec. 2005, doi: 10.1016/j.cell.2005.08.040.

[16]   J.-L. Chew *et al.*, "Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells," *Mol. Cell. Biol.*, vol. 25, no. 14, pp. 6031–6046, Jul. 2005, doi: 10.1128/MCB.25.14.6031-6046.2005.

[17]   A. Rizzino, "Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels," *Stem Cells*, vol. 31, no. 6, pp. 1033–1039, Jun. 2013, doi: 10.1002/stem.1352.

[18]   X. Chen *et al.*, "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells," *Cell*, vol. 133, no. 6, pp. 1106–1117, Jun. 2008, doi: 10.1016/j.cell.2008.04.043.

[19]   K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell*, vol. 126, no. 4, pp. 663–676, Aug. 2006, doi: 10.1016/j.cell.2006.07.024.

[20] J. Yu *et al.*, "Induced pluripotent stem cell lines derived from human somatic cells," *Science*, vol. 318, no. 5858, pp. 1917–1920, Dec. 2007, doi: 10.1126/science.1151526.

[21] M. Stadtfeld and K. Hochedlinger, "Induced pluripotency: history, mechanisms, and applications." [Online]. Available: http://genesdev.cshlp.org. [Accessed: 20-Sep-2019].

[22] K. Onishi and P. W. Zandstra, "LIF signaling in stem cells and development," *Development*, vol. 142, no. 13, pp. 2230–2236, Jul. 2015, doi: 10.1242/dev.117598.

[23] L. Vallier *et al.*, "Activin/Nodal signalling maintains pluripotency by controlling Nanog expression," *Development*, vol. 136, no. 8, pp. 1339–1349, Apr. 2009, doi: 10.1242/dev.033951.

[24] D. M. Ornitz and N. Itoh, "The Fibroblast Growth Factor signaling pathway," *Wiley Interdiscip Rev Dev Biol*, vol. 4, no. 3, pp. 215–266, May 2015, doi: 10.1002/wdev.176.

[25] M. F. Pera, "In Search of Naivety," *Cell Stem Cell*, vol. 15, no. 5, pp. 543–545, Nov. 2014, doi: 10.1016/j.stem.2014.10.013.

[26] J. Nichols and A. Smith, "Naive and Primed Pluripotent States," *Cell Stem Cell*, vol. 4, no. 6, pp. 487–492, Jun. 2009, doi: 10.1016/j.stem.2009.05.015.

[27] S. Chandrasekaran *et al.*, "Comprehensive mapping of pluripotent stem cell metabolism using dynamic genome-scale network modeling," *Cell Rep*, vol. 21, no. 10, pp. 2965–2977, Dec. 2017, doi: 10.1016/j.celrep.2017.07.048.

[28] W. Gu *et al.*, "Glycolytic metabolism plays a functional role in regulating human pluripotent stem cell state," *Cell Stem Cell*, vol. 19, no. 4, pp. 476–490, Oct. 2016, doi: 10.1016/j.stem.2016.08.008.

[29] C. B. Ware, "Concise Review: Lessons from Naïve Human Pluripotent Cells," *Stem Cells*, vol. 35, no. 1, pp. 35–41, Jan. 2017, doi: 10.1002/stem.2507.

[30] L. E. Bates and J. C. Silva, "Reprogramming human cells to naïve pluripotency: how close are we?," *Current Opinion in Genetics & Development*, vol. 46, pp. 58–65, Oct. 2017, doi: 10.1016/j.gde.2017.06.009.

[31] K. A. D'Amour, A. D. Agulnick, S. Eliazer, O. G. Kelly, E. Kroon, and E. E. Baetge, "Efficient differentiation of human embryonic stem cells to definitive endoderm," *Nat. Biotechnol.*, vol. 23, no. 12, pp. 1534–1541, Dec. 2005, doi: 10.1038/nbt1163.

[32] K. C. Davidson *et al.*, "Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4," *Proc Natl Acad Sci U S A*, vol. 109, no. 12, pp. 4485–4490, Mar. 2012, doi: 10.1073/pnas.1118777109.

[33] P. Reinhardt *et al.*, "Derivation and Expansion Using Only Small Molecules of Human Neural Progenitors for Neurodegenerative Disease Modeling," *PLOS ONE*, vol. 8, no. 3, p. e59252, Mar. 2013, doi: 10.1371/journal.pone.0059252.

[34] Q. Qu *et al.*, "High-efficiency motor neuron differentiation from human pluripotent stem cells and the function of Islet-1," *Nature Communications*, vol. 5, p. 3449, Mar. 2014, doi: 10.1038/ncomms4449.

[35] A. Shaltouki, J. Peng, Q. Liu, M. S. Rao, and X. Zeng, "Efficient generation of astrocytes from human pluripotent stem cells in defined conditions," *Stem Cells*, vol. 31, no. 5, pp. 941–952, May 2013, doi: 10.1002/stem.1334.

[36] R. Morizane and J. V. Bonventre, "Generation of nephron progenitor cells and kidney organoids from human pluripotent stem cells," *Nat Protoc*, vol. 12, no. 1, pp. 195–207, Jan. 2017, doi: 10.1038/nprot.2016.170.

[37] A. Carpentier *et al.*, "Engrafted human stem cell-derived hepatocytes establish an infectious HCV murine model," *J. Clin. Invest.*, vol. 124, no. 11, pp. 4953–4964, Nov. 2014, doi: 10.1172/JCI75456.

[38] B. Schaefke, W. Sun, Y.-S. Li, L. Fang, and W. Chen, "The evolution of posttranscriptional regulation," *Wiley Interdisciplinary Reviews: RNA*, vol. 9, no. 5, p. e1485, 2018, doi: 10.1002/wrna.1485.

[39] S. Gerstberger, M. Hafner, M. Ascano, and T. Tuschl, "Evolutionary Conservation and Expression of Human RNA-Binding Proteins and Their Role in Human Genetic Disease," *Adv Exp Med Biol*, vol. 825, pp. 1–55, 2014, doi: 10.1007/978-1-4939-1221-6_1.

[40] C. Y. Cheong and T. Lufkin, "Alternative Splicing in Self-Renewal of Embryonic Stem Cells," *Stem Cells Int*, vol. 2011, Jun. 2011, doi: 10.4061/2011/560261.

[41] M. Gabut *et al.*, "An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming," *Cell*, vol. 147, no. 1, pp. 132–146, Sep. 2011, doi: 10.1016/j.cell.2011.08.023.

[42] X. Lu *et al.*, "SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells," *Nat. Cell Biol.*, vol. 15, no. 10, pp. 1141–1152, Oct. 2013, doi: 10.1038/ncb2839.

[43] H. Han *et al.*, "MBNL proteins repress ES-cell-specific alternative splicing and reprogramming," *Nature*, vol. 498, no. 7453, pp. 241–245, Jun. 2013, doi: 10.1038/nature12270.

[44] R. Elkon, A. P. Ugalde, and R. Agami, "Alternative cleavage and polyadenylation: extent, regulation and function," *Nature Reviews Genetics*, vol. 14, no. 7, pp. 496–506, Jul. 2013, doi: 10.1038/nrg3482.

[45] Y. Shi, "Alternative polyadenylation: new insights from global analyses," *RNA*, vol. 18, no. 12, pp. 2105–2117, Dec. 2012, doi: 10.1261/rna.035899.112.

[46] B. Lackford *et al.*, "Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal," *EMBO J.*, vol. 33, no. 8, pp. 878–889, Apr. 2014, doi: 10.1002/embj.201386537.

[47] M. Modic *et al.*, "Cross-Regulation between TDP-43 and Paraspeckles Promotes Pluripotency-Differentiation Transition," *Mol. Cell*, vol. 74, no. 5, pp. 951-965.e13, Jun. 2019, doi: 10.1016/j.molcel.2019.03.041.

[48] J. Ye and R. Blelloch, "Regulation of Pluripotency by RNA Binding Proteins," *Cell Stem Cell*, vol. 15, no. 3, pp. 271–280, Sep. 2014, doi: 10.1016/j.stem.2014.08.010.

[49] A. Fatica and I. Bozzoni, "Long non-coding RNAs: new players in cell differentiation and development," *Nature Reviews Genetics*, vol. 15, no. 1, pp. 7–21, Jan. 2014, doi: 10.1038/nrg3606.

[50] H. Hezroni, D. Koppstein, M. G. Schwartz, A. Avrutin, D. P. Bartel, and I. Ulitsky, "Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species," *Cell Rep*, vol. 11, no. 7, pp. 1110–1122, May 2015, doi: 10.1016/j.celrep.2015.04.023.

[51] R. B.-T. Perry and I. Ulitsky, "The functions of long noncoding RNAs in development and stem cells," *Development*, vol. 143, no. 21, pp. 3882–3894, Nov. 2016, doi: 10.1242/dev.140962.

[52] F. Kopp and J. T. Mendell, "Functional classification and experimental dissection of long noncoding RNAs," *Cell*, vol. 172, no. 3, pp. 393–407, Jan. 2018, doi: 10.1016/j.cell.2018.01.011.

[53] J. T. Lee and M. S. Bartolomei, "X-inactivation, imprinting, and long noncoding RNAs in health and disease," *Cell*, vol. 152, no. 6, pp. 1308–1323, Mar. 2013, doi: 10.1016/j.cell.2013.02.016.

[54] D. Tian, S. Sun, and J. T. Lee, "The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation," *Cell*, vol. 143, no. 3, pp. 390–403, Oct. 2010, doi: 10.1016/j.cell.2010.09.049.

[55] D. Mancini-Dinardo, S. J. S. Steele, J. M. Levorse, R. S. Ingram, and S. M. Tilghman, "Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes," *Genes Dev.*, vol. 20, no. 10, pp. 1268–1282, May 2006, doi: 10.1101/gad.1416906.

[56] M. Mallo and C. R. Alonso, "The regulation of Hox gene expression during animal development," *Development*, vol. 140, no. 19, pp. 3951–3963, Oct. 2013, doi: 10.1242/dev.068346.

[57] K. C. Wang *et al.*, "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression," *Nature*, vol. 472, no. 7341, pp. 120–124, Apr. 2011, doi: 10.1038/nature09819.

[58] M.-C. Tsai *et al.*, "Long noncoding RNA as modular scaffold of histone modification complexes," *Science*, vol. 329, no. 5992, pp. 689–693, Aug. 2010, doi: 10.1126/science.1192002.

[59] C. A. Klattenhoff *et al.*, "Braveheart, a Long Noncoding RNA Required for Cardiovascular Lineage Commitment," *Cell*, vol. 152, no. 3, pp. 570–583, Jan. 2013, doi: 10.1016/j.cell.2013.01.003.

[60] P. Grote *et al.*, "The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse," *Dev. Cell*, vol. 24, no. 2, pp. 206–214, Jan. 2013, doi: 10.1016/j.devcel.2012.12.012.

[61] A. N. Kallen *et al.*, "The imprinted H19 lncRNA antagonizes let-7 microRNAs," *Mol. Cell*, vol. 52, no. 1, pp. 101–112, Oct. 2013, doi: 10.1016/j.molcel.2013.08.027.

[62] M. Kretz *et al.*, "Control of somatic tissue differentiation by the long non-coding RNA TINCR," *Nature*, vol. 493, no. 7431, pp. 231–235, Jan. 2013, doi: 10.1038/nature11661.

[63] D. Bernard *et al.*, "A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression," *EMBO J*, vol. 29, no. 18, pp. 3082–3093, Sep. 2010, doi: 10.1038/emboj.2010.199.

[64] J. Feng, C. Bi, B. S. Clark, R. Mady, P. Shah, and J. D. Kohtz, "The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator," *Genes Dev.*, vol. 20, no. 11, pp. 1470–1484, Jun. 2006, doi: 10.1101/gad.1416106.

[65] M. Guttman *et al.*, "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals," *Nature*, vol. 458, no. 7235, pp. 223–227, Mar. 2009, doi: 10.1038/nature07672.

[66] M. Guttman *et al.*, "lincRNAs act in the circuitry controlling pluripotency and differentiation," *Nature*, vol. 477, no. 7364, pp. 295–300, Sep. 2011, doi: 10.1038/nature10398.

[67] J. Sheik Mohamed, P. M. Gaughwin, B. Lim, P. Robson, and L. Lipovich, "Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells," *RNA*, vol. 16, no. 2, pp. 324–337, Feb. 2010, doi: 10.1261/rna.1441510.

[68] N. Lin *et al.*, "An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment," *Mol Cell*, vol. 53, no. 6, pp. 1005–1019, Mar. 2014, doi: 10.1016/j.molcel.2014.01.021.

[69] A. K. Jain *et al.*, "LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56," *Molecular Cell*, vol. 64, no. 5, pp. 967–981, Dec. 2016, doi: 10.1016/j.molcel.2016.10.039.

[70] S.-Y. Ng, R. Johnson, and L. W. Stanton, "Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors," *EMBO J.*, vol. 31, no. 3, pp. 522–533, Feb. 2012, doi: 10.1038/emboj.2011.459.

[71] Y. Wang *et al.*, "Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal," *Developmental Cell*, vol. 25, no. 1, pp. 69–80, Apr. 2013, doi: 10.1016/j.devcel.2013.03.002.

[72] K. Daneshvar *et al.*, "DIGIT Is a Conserved Long Noncoding RNA that Regulates GSC Expression to Control Definitive Endoderm Differentiation of Embryonic Stem Cells," *Cell Reports*, vol. 17, no. 2, pp. 353–365, Oct. 2016, doi: 10.1016/j.celrep.2016.09.017.

[73] X. Xu, M. Guo, N. Zhang, and S. Ye, "Telomeric noncoding RNA promotes mouse embryonic stem cell self-renewal through inhibition of TCF3 activity," *American Journal of Physiology-Cell Physiology*, vol. 314, no. 6, pp. C712–C720, Mar. 2018, doi: 10.1152/ajpcell.00292.2017.

[74] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, "Biomolecular condensates: organizers of cellular biochemistry," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 5, pp. 285–298, May 2017, doi: 10.1038/nrm.2017.7.

[75] E. Gomes and J. Shorter, "The molecular language of membraneless organelles," *J. Biol. Chem.*, vol. 294, no. 18, pp. 7115–7127, May 2019, doi: 10.1074/jbc.TM118.001192.

[76] C. P. Brangwynne *et al.*, "Germline P granules are liquid droplets that localize by controlled dissolution/condensation," *Science*, vol. 324, no. 5935, pp. 1729–1732, Jun. 2009, doi: 10.1126/science.1172046.

[77] C. P. Brangwynne, T. J. Mitchison, and A. A. Hyman, "Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 11, pp. 4334–4339, Mar. 2011, doi: 10.1073/pnas.1017150108.

[78] A. Molliex *et al.*, "Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization," *Cell*, vol. 163, no. 1, pp. 123–133, Sep. 2015, doi: 10.1016/j.cell.2015.09.015.

[79] J. B. Woodruff, B. Ferreira Gomes, P. O. Widlund, J. Mahamid, A. Honigmann, and A. A. Hyman, "The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin," *Cell*, vol. 169, no. 6, pp. 1066-1077.e10, Jun. 2017, doi: 10.1016/j.cell.2017.05.028.

[80] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, "Phase separation drives heterochromatin domain formation," *Nature*, vol. 547, no. 7662, pp. 241–245, Jul. 2017, doi: 10.1038/nature22989.

[81] A. H. Fox, S. Nakagawa, T. Hirose, and C. S. Bond, "Paraspeckles: Where Long Noncoding RNA Meets Phase Separation," *Trends Biochem. Sci.*, vol. 43, no. 2, pp. 124–135, Feb. 2018, doi: 10.1016/j.tibs.2017.12.001.

[82] A. H. Fox *et al.*, "Paraspeckles: a novel nuclear domain," *Curr. Biol.*, vol. 12, no. 1, pp. 13–25, Jan. 2002, doi: 10.1016/s0960-9822(01)00632-7.

[83] T. Naganuma, S. Nakagawa, A. Tanigawa, Y. F. Sasaki, N. Goshima, and T. Hirose, "Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles," *EMBO J.*, vol. 31, no. 20, pp. 4020–4034, Oct. 2012, doi: 10.1038/emboj.2012.251.

[84] T. Hirose and S. Nakagawa, "Paraspeckles: possible nuclear hubs by the RNA for the RNA," *Biomol Concepts*, vol. 3, no. 5, pp. 415–428, Oct. 2012, doi: 10.1515/bmc-2012-0017.

[85] Y. T. F. Sasaki, T. Ideue, M. Sano, T. Mituyama, and T. Hirose, "MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 8, pp. 2525–2530, Feb. 2009, doi: 10.1073/pnas.0807899106.

[86] C. M. Clemson *et al.*, "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles," *Mol. Cell*, vol. 33, no. 6, pp. 717–726, Mar. 2009, doi: 10.1016/j.molcel.2009.01.026.

[87] H. Sunwoo, M. E. Dinger, J. E. Wilusz, P. P. Amaral, J. S. Mattick, and D. L. Spector, "MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles," *Genome Res.*, vol. 19, no. 3, pp. 347–359, Mar. 2009, doi: 10.1101/gr.087775.108.

[88] L.-L. Chen and G. G. Carmichael, "Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA," *Mol. Cell*, vol. 35, no. 4, pp. 467–478, Aug. 2009, doi: 10.1016/j.molcel.2009.06.027.

[89] B. Zhang *et al.*, "Identification and Characterization of a Class of MALAT1-like Genomic Loci," *Cell Rep*, vol. 19, no. 8, pp. 1723–1738, 23 2017, doi: 10.1016/j.celrep.2017.05.006.

[90] S. Souquere, G. Beauclair, F. Harper, A. Fox, and G. Pierron, "Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies," *Mol. Biol. Cell*, vol. 21, no. 22, pp. 4020–4027, Nov. 2010, doi: 10.1091/mbc.E10-08-0690.

[91] J. A. West *et al.*, "Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization," *J. Cell Biol.*, vol. 214, no. 7, pp. 817–830, 26 2016, doi: 10.1083/jcb.201601071.

[92] T. Trcek, M. Grosch, A. York, H. Shroff, T. Lionnet, and R. Lehmann, "*Drosophila* germ granules are structured and contain homotypic mRNA clusters," *Nature Communications*, vol. 6, p. 7962, Aug. 2015, doi: 10.1038/ncomms8962.

[93] J. R. Wheeler, T. Matheny, S. Jain, R. Abrisch, and R. Parker, "Distinct stages in stress granule assembly and disassembly," *eLife*, vol. 5, doi: 10.7554/eLife.18413.

[94] Y. W. Lam and L. Trinkle-Mulcahy, "New insights into nucleolar structure and function," *F1000Prime Rep*, vol. 7, Apr. 2015, doi: 10.12703/P7-48.

[95] R. Li, A. R. Harvey, S. I. Hodgetts, and A. H. Fox, "Functional dissection of NEAT1 using genome editing reveals substantial localization of the NEAT1_1 isoform outside paraspeckles," *RNA*, vol. 23, no. 6, pp. 872–881, Jun. 2017, doi: 10.1261/rna.059477.116.

[96] M. Shuaib *et al.*, "AGO1 in association with NEAT1 lncRNA contributes to nuclear and 3D chromatin architecture in human cells: Supplemental Data," Genomics, preprint, Jan. 2019.

[97] Y. Wang *et al.*, "Genome-wide screening of NEAT1 regulators reveals cross-regulation between paraspeckles and mitochondria," *Nat. Cell Biol.*, vol. 20, no. 10, pp. 1145–1158, 2018, doi: 10.1038/s41556-018-0204-2.

[98] I. Chillón and A. M. Pyle, "Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function," *Nucleic Acids Res*, vol. 44, no. 19, pp. 9462–9471, Nov. 2016, doi: 10.1093/nar/gkw599.

[99] K. V. Prasanth *et al.*, "Regulating gene expression through RNA nuclear retention," *Cell*, vol. 123, no. 2, pp. 249–263, Oct. 2005, doi: 10.1016/j.cell.2005.08.033.

[100] E. Eisenberg and E. Y. Levanon, "A-to-I RNA editing — immune protector and transcriptome diversifier," *Nat Rev Genet*, vol. 19, no. 8, pp. 473–490, Aug. 2018, doi: 10.1038/s41576-018-0006-1.

[101] M. M. Lamers, B. G. van den Hoogen, and B. L. Haagmans, "ADAR1: 'Editor-in-Chief' of Cytoplasmic Innate Immunity," *Front Immunol*, vol. 10, Jul. 2019, doi: 10.3389/fimmu.2019.01763.

[102] P. Deininger, "Alu elements: know the SINEs," *Genome Biology*, vol. 12, no. 12, p. 236, Dec. 2011, doi: 10.1186/gb-2011-12-12-236.

[103] L.-L. Chen, J. N. DeCerbo, and G. G. Carmichael, "Alu element-mediated gene silencing," *EMBO J.*, vol. 27, no. 12, pp. 1694–1705, Jun. 2008, doi: 10.1038/emboj.2008.94.

[104] S. Boeynaems *et al.*, "Protein Phase Separation: A New Phase in Cell Biology," *Trends Cell Biol*, vol. 28, no. 6, pp. 420–435, Jun. 2018, doi: 10.1016/j.tcb.2018.02.004.

[105] K. Imamura *et al.*, "Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli," *Mol. Cell*, vol. 53, no. 3, pp. 393–406, Feb. 2014, doi: 10.1016/j.molcel.2014.01.009.

[106] T. Hirose *et al.*, "NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies," *Mol. Biol. Cell*, vol. 25, no. 1, pp. 169–183, Jan. 2014, doi: 10.1091/mbc.E13-09-0558.

[107] L. Jiang *et al.*, "NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing," *Nat. Struct. Mol. Biol.*, vol. 24, no. 10, pp. 816–824, Oct. 2017, doi: 10.1038/nsmb.3455.

[108] M. Rutenberg-Schoenberg, A. N. Sexton, and M. D. Simon, "The Properties of Long Noncoding RNAs That Regulate Chromatin," *Annu. Rev. Genom. Hum. Genet.*, vol. 17, no. 1, pp. 69–94, Aug. 2016, doi: 10.1146/annurev-genom-090314-024939.

[109] J. A. West *et al.*, "The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites," *Mol. Cell*, vol. 55, no. 5, pp. 791–802, Sep. 2014, doi: 10.1016/j.molcel.2014.07.012.

[110] T. Kawaguchi *et al.*, "SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 14, pp. 4304–4309, Apr. 2015, doi: 10.1073/pnas.1423819112.

[111] C.-C. Kuo *et al.*, "Detection of RNA–DNA binding sites in long noncoding RNAs," *Nucleic Acids Res*, vol. 47, no. 6, pp. e32–e32, Apr. 2019, doi: 10.1093/nar/gkz037.

[112] S. Nakagawa *et al.*, "The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice," *Development*, vol. 141, no. 23, pp. 4618–4627, Dec. 2014, doi: 10.1242/dev.110544.

[113] A. Hupalowska, A. Jedrusik, M. Zhu, M. T. Bedford, D. M. Glover, and M. Zernicka-Goetz, "CARM1 and Paraspeckles Regulate Pre-implantation Mouse Embryo Development," *Cell*, vol. 175, no. 7, pp. 1902-1916.e13, Dec. 2018, doi: 10.1016/j.cell.2018.11.027.

[114] M.-E. Torres-Padilla, D.-E. Parfitt, T. Kouzarides, and M. Zernicka-Goetz, "Histone arginine methylation regulates pluripotency in the early mouse embryo," *Nature*, vol. 445, no. 7124, pp. 214–218, Jan. 2007, doi: 10.1038/nature05458.

[115] P. Dong *et al.*, "Long Non-coding RNA NEAT1: A Novel Target for Diagnosis and Therapy in Human Tumors," *Front Genet*, vol. 9, Oct. 2018, doi: 10.3389/fgene.2018.00471.

[116] X. Yu, Z. Li, H. Zheng, M. T. V. Chan, and W. K. K. Wu, "NEAT1: A novel cancer-related long non-coding RNA," *Cell Prolif.*, vol. 50, no. 2, Apr. 2017, doi: 10.1111/cpr.12329.

[117] E. Rheinbay *et al.*, "Recurrent and functional regulatory mutations in breast cancer," *Nature*, vol. 547, no. 7661, pp. 55–60, 06 2017, doi: 10.1038/nature22992.

[118] H. Choudhry *et al.*, "Tumor hypoxia induces nuclear paraspeckle formation through HIF-2α dependent transcriptional activation of NEAT1 leading to cancer cell survival," *Oncogene*, vol. 34, no. 34, pp. 4482–4490, Aug. 2015, doi: 10.1038/onc.2014.378.

[119] A. R. Barutcu *et al.*, "RUNX1 contributes to higher-order chromatin organization and gene regulation in breast cancer cells," *Biochim. Biophys. Acta*, vol. 1859, no. 11, pp. 1389–1397, 2016, doi: 10.1016/j.bbagrm.2016.08.003.

[120] Q. Chen *et al.*, "Long Noncoding RNA NEAT1, Regulated by the EGFR Pathway, Contributes to Glioblastoma Progression Through the WNT/β-Catenin Pathway by Scaffolding EZH2," *Clin. Cancer Res.*, vol. 24, no. 3, pp. 684–695, 01 2018, doi: 10.1158/1078-0432.CCR-17-0605.

[121] C. Adriaens *et al.*, "p53 induces formation of *NEAT1* lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity," *Nature Medicine*, vol. 22, no. 8, pp. 861–868, Aug. 2016, doi: 10.1038/nm.4135.

[122] C. Zhang *et al.*, "Long Noncoding RNA NEAT1 Promotes Growth and Metastasis of Cholangiocarcinoma Cells," *Oncol. Res.*, vol. 26, no. 6, pp. 879–888, Jul. 2018, doi: 10.3727/096504017X15024935181289.

[123] F. Prinz, A. Kapeller, M. Pichler, and C. Klec, "The Implications of the Long Non-Coding RNA NEAT1 in Non-Cancerous Diseases," *Int J Mol Sci*, vol. 20, no. 3, Feb. 2019, doi: 10.3390/ijms20030627.

[124] H. Ma *et al.*, "The Long Noncoding RNA NEAT1 Exerts Antihantaviral Effects by Acting as Positive Feedback for RIG-I Signaling," *Journal of Virology*, vol. 91, no. 9, pp. e02250-16, May 2017, doi: 10.1128/JVI.02250-16.

[125] Q. Zhang, C.-Y. Chen, V. S. R. K. Yedavalli, and K.-T. Jeang, "NEAT1 Long Noncoding RNA and Paraspeckle Bodies Modulate HIV-1 Posttranscriptional Expression," *mBio*, vol. 4, no. 1, pp. e00596-12, Mar. 2013, doi: 10.1128/mBio.00596-12.

[126] Z. Wang *et al.*, "NEAT1 modulates herpes simplex virus-1 replication by regulating viral gene transcription," *Cell. Mol. Life Sci.*, vol. 74, no. 6, pp. 1117–1131, Mar. 2017, doi: 10.1007/s00018-016-2398-4.

[127] Y. Nishimoto *et al.*, "The long non-coding RNA nuclear-enriched abundant transcript 1_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis," *Molecular Brain*, vol. 6, no. 1, p. 31, Jul. 2013, doi: 10.1186/1756-6606-6-31.

[128] T. A. Shelkovnikova *et al.*, "Protective paraspeckle hyper-assembly downstream of TDP-43 loss of function in amyotrophic lateral sclerosis," *Molecular Neurodegeneration*, vol. 13, no. 1, p. 30, Jun. 2018, doi: 10.1186/s13024-018-0263-7.

[129] H. An, N. G. Williams, and T. A. Shelkovnikova, "NEAT1 and paraspeckles in neurodegenerative diseases: A missing lnc found?," *Non-coding RNA Research*, vol. 3, no. 4, pp. 243–252, Dec. 2018, doi: 10.1016/j.ncrna.2018.11.003.

[130] "Meta-Analysis of Parkinson's Disease Transcriptome Data Using TRAM Software: Whole Substantia Nigra Tissue and Single Dopamine Neuron Differential ... - PubMed - NCBI." [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27611585. [Accessed: 26-Sep-2019].

[131] C. Cheng *et al.*, "The long non-coding RNA NEAT1 is elevated in polyglutamine repeat expansion diseases and protects from disease gene-dependent toxicities," *Hum. Mol. Genet.*, vol. 27, no. 24, pp. 4303–4314, 15 2018, doi: 10.1093/hmg/ddy331.

[132] M. Spreafico, B. Grillo, F. Rusconi, E. Battaglioli, and M. Venturin, "Multiple Layers of CDK5R1 Regulation in Alzheimer's Disease Implicate Long Non-Coding RNAs," *International Journal of Molecular Sciences*, vol. 19, no. 7, p. 2022, Jul. 2018, doi: 10.3390/ijms19072022.

[133] A. H. Fox, C. S. Bond, and A. I. Lamond, "P54nrb Forms a Heterodimer with PSP1 That Localizes to Paraspeckles in an RNA-dependent Manner," *Mol Biol Cell*, vol. 16, no. 11, pp. 5304–5315, Nov. 2005, doi: 10.1091/mbc.E05-06-0587.

[134] B. Dong, D. S. Horowitz, R. Kobayashi, and A. R. Krainer, "Purification and cDNA cloning of HeLa cell p54nrb, a nuclear protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and Drosophila NONA/BJ6," *Nucleic Acids Res.*, vol. 21, no. 17, pp. 4085–4092, Aug. 1993, doi: 10.1093/nar/21.17.4085.

[135] G. J. Knott, C. S. Bond, and A. H. Fox, "The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold," *Nucleic Acids Res.*, vol. 44, no. 9, pp. 3989–4004, 19 2016, doi: 10.1093/nar/gkw271.

[136] M. Lee *et al.*, "The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation," *Nucleic Acids Res.*, vol. 43, no. 7, pp. 3826–3840, Apr. 2015, doi: 10.1093/nar/gkv156.

[137] S. Li, Z. Li, F.-J. Shu, H. Xiong, A. C. Phillips, and W. S. Dynan, "Double-strand break repair deficiency in NONO knockout murine embryonic fibroblasts and compensation by spontaneous upregulation of the PSPC1 paralog," *Nucleic Acids Res.*, vol. 42, no. 15, pp. 9771–9780, Sep. 2014, doi: 10.1093/nar/gku650.

[138] D. Mircsof *et al.*, "Mutations in *NONO* lead to syndromic intellectual disability and inhibitory synaptic defects," *Nature Neuroscience*, vol. 18, no. 12, pp. 1731–1736, Dec. 2015, doi: 10.1038/nn.4169.

[139] X. Gao *et al.*, "Paraspeckle Protein 1 (PSPC1) Is Involved in the Cisplatin Induced DNA Damage Response—Role in G1/S Checkpoint," *PLOS ONE*, vol. 9, no. 5, p. e97174, May 2014, doi: 10.1371/journal.pone.0097174.

[140] S. P. Yadav *et al.*, "The transcription-splicing protein NonO/p54nrb and three NonO-interacting proteins bind to distal enhancer region and augment rhodopsin expression," *Hum Mol Genet*, vol. 23, no. 8, pp. 2132–2144, Apr. 2014, doi: 10.1093/hmg/ddt609.

[141] A. L. Amelio *et al.*, "A coactivator trap identifies NONO (p54nrb) as a component of the cAMP-signaling pathway," *PNAS*, vol. 104, no. 51, pp. 20314–20319, Dec. 2007, doi: 10.1073/pnas.0707999105.

[142] X. Dong, J. Sweet, J. R. G. Challis, T. Brown, and S. J. Lye, "Transcriptional activity of androgen receptor is modulated by two RNA splicing factors, PSF and p54nrb," *Mol. Cell. Biol.*, vol. 27, no. 13, pp. 4863–4875, Jul. 2007, doi: 10.1128/MCB.02144-06.

[143] X. Dong, C. Yu, O. Shynlova, J. R. G. Challis, P. S. Rennie, and S. J. Lye, "p54nrb Is a Transcriptional Corepressor of the Progesterone Receptor that Modulates Transcription of the Labor-Associated Gene, Connexin 43 (Gja1)," *Mol Endocrinol*, vol. 23, no. 8, pp. 1147–1160, Aug. 2009, doi: 10.1210/me.2008-0357.

[144] S. Kaneko, O. Rozenblatt-Rosen, M. Meyerson, and J. L. Manley, "The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination," *Genes Dev.*, vol. 21, no. 14, pp. 1779–1789, Jul. 2007, doi: 10.1101/gad.1565207.

[145] F. Heyd and K. W. Lynch, "PSF controls expression of histone variants and cellular viability in thymocytes," *Biochem. Biophys. Res. Commun.*, vol. 414, no. 4, pp. 743–749, Nov. 2011, doi: 10.1016/j.bbrc.2011.09.149.

[146] C. A. Yarosh, J. R. Iacona, C. S. Lutz, and K. W. Lynch, "PSF: nuclear busy-body or nuclear facilitator?," *Wiley Interdisc Rev RNA*, vol. 6, no. 4, pp. 351–367, Aug. 2015, doi: 10.1002/wrna.1280.

[147] F. Heyd and K. W. Lynch, "Phosphorylation-dependent regulation of PSF by GSK3 controls CD45 alternative splicing," *Mol. Cell*, vol. 40, no. 1, pp. 126–137, Oct. 2010, doi: 10.1016/j.molcel.2010.09.013.

[148] P. Ray, A. Kar, K. Fushimi, N. Havlioglu, X. Chen, and J. Y. Wu, "PSF Suppresses Tau Exon 10 Inclusion by Interacting with a Stem-Loop Structure Downstream of Exon 10," *J Mol Neurosci*, vol. 45, no. 3, pp. 453–466, Nov. 2011, doi: 10.1007/s12031-011-9634-z.

[149] K. K. Kim, Y. C. Kim, R. S. Adelstein, and S. Kawamoto, "Fox-3 and PSF interact to activate neural cell-specific alternative splicing," *Nucleic Acids Res.*, vol. 39, no. 8, pp. 3064–3078, Apr. 2011, doi: 10.1093/nar/gkq1221.

[150] S. Cho *et al.*, "PSF contacts exon 7 of SMN2 pre-mRNA to promote exon 7 inclusion," *Biochim. Biophys. Acta*, vol. 1839, no. 6, pp. 517–525, Jun. 2014, doi: 10.1016/j.bbagrm.2014.03.003.

[151] H. Izumi, A. McCloskey, K. Shinmyozu, and M. Ohno, "p54nrb/NonO and PSF promote U snRNA nuclear export by accelerating its export complex assembly," *Nucleic Acids Res*, vol. 42, no. 6, pp. 3998–4007, Apr. 2014, doi: 10.1093/nar/gkt1365.

[152] Y. Kanai, N. Dohmae, and N. Hirokawa, "Kinesin Transports RNA: Isolation and Characterization of an RNA-Transporting Granule," *Neuron*, vol. 43, no. 4, pp. 513–525, Aug. 2004, doi: 10.1016/j.neuron.2004.07.022.

[153] K. E. Cosker, S. J. Fenstermacher, M. F. Pazyra-Murphy, H. L. Elliott, and R. A. Segal, "The RNA-binding protein SFPQ orchestrates an RNA regulon to promote axon viability," *Nat. Neurosci.*, vol. 19, no. 5, pp. 690–696, 2016, doi: 10.1038/nn.4280.

[154] S. Crnko, B. C. D. Pré, J. P. G. Sluijter, and L. W. V. Laake, "Circadian rhythms and the molecular clock in cardiovascular biology and disease," *Nat Rev Cardiol*, vol. 16, no. 7, pp. 437–447, Jul. 2019, doi: 10.1038/s41569-019-0167-4.

[155] U. Albrecht and G. Eichele, "The mammalian circadian clock," *Current Opinion in Genetics & Development*, vol. 13, no. 3, pp. 271–277, Jun. 2003, doi: 10.1016/S0959-437X(03)00055-8.

[156] S. A. Brown *et al.*, "PERIOD1-Associated Proteins Modulate the Negative Limb of the Mammalian Circadian Oscillator," *Science*, vol. 308, no. 5722, pp. 693–696, Apr. 2005, doi: 10.1126/science.1107373.

[157] H. A. Duong, M. S. Robles, D. Knutti, and C. J. Weitz, "A Molecular Mechanism for Circadian Clock Negative Feedback," *Science*, vol. 332, no. 6036, pp. 1436–1439, Jun. 2011, doi: 10.1126/science.1196766.

[158] C. L. Bladen, D. Udayakumar, Y. Takeda, and W. S. Dynan, "Identification of the polypyrimidine tract binding protein-associated splicing factor.p54(nrb) complex as a candidate DNA double-strand break rejoining factor," *J. Biol. Chem.*, vol. 280, no. 7, pp. 5205–5210, Feb. 2005, doi: 10.1074/jbc.M412758200.

[159] Y. Morozumi, Y. Takizawa, M. Takaku, and H. Kurumizaka, "Human PSF binds to RAD51 and modulates its homologous-pairing and strand-exchange activities," *Nucleic Acids Res.*, vol. 37, no. 13, pp. 4296–4307, Jul. 2009, doi: 10.1093/nar/gkp298.

[160] M. Lek *et al.*, "Analysis of protein-coding genetic variation in 60,706 humans," *Nature*, vol. 536, no. 7616, pp. 285–291, 18 2016, doi: 10.1038/nature19057.

[161] Z. Zhu *et al.*, "p54(nrb)/NONO regulates lipid metabolism and breast cancer growth through SREBP-1A," *Oncogene*, vol. 35, no. 11, pp. 1399–1410, Mar. 2016, doi: 10.1038/onc.2015.197.

[162] M. J. Larriba *et al.*, "Novel snail1 target proteins in human colon cancer identified by proteomic analysis," *PLoS ONE*, vol. 5, no. 4, p. e10221, Apr. 2010, doi: 10.1371/journal.pone.0010221.

[163] I. Cristobo, M. J. Larriba, V. de los Ríos, F. García, A. Muñoz, and J. I. Casal, "Proteomic analysis of 1α,25-dihydroxyvitamin D3 action on human colon cancer cells reveals a link to splicing regulation," *J Proteomics*, vol. 75, no. 2, pp. 384–397, Dec. 2011, doi: 10.1016/j.jprot.2011.08.003.

[164] V. S. Greco-Stewart, C. S.-L. Thibault, and M. Pelchat, "Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis delta virus RNA," *Virology*, vol. 356, no. 1–2, pp. 35–44, Dec. 2006, doi: 10.1016/j.virol.2006.06.040.

[165] S. Landeras-Bueno, N. Jorba, M. Pérez-Cidoncha, and J. Ortín, "The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription," *PLoS Pathog.*, vol. 7, no. 11, p. e1002397, Nov. 2011, doi: 10.1371/journal.ppat.1002397.

[166] A. Kula, L. Gharu, and A. Marcello, "HIV-1 pre-mRNA commitment to Rev mediated export through PSF and Matrin 3," *Virology*, vol. 435, no. 2, pp. 329–340, Jan. 2013, doi: 10.1016/j.virol.2012.10.032.

[167] C. St Gelais, J. Roger, and L. Wu, "Non-POU Domain-Containing Octamer-Binding Protein Negatively Regulates HIV-1 Infection in CD4(+) T Cells," *AIDS Res. Hum. Retroviruses*, vol. 31, no. 8, pp. 806–816, Aug. 2015, doi: 10.1089/AID.2014.0313.

[168] R. Olmer, J. Dahlmann, S. Merkert, S. Baus, G. Göhring, and U. Martin, "Generation of a NKX2.1 knock-in reporter cell line from human induced pluripotent stem cells (MHHi006-A-2)," *Stem Cell Research*, vol. 39, p. 101492, Aug. 2019, doi: 10.1016/j.scr.2019.101492.

[169] C. Kunze *et al.*, "Synthetic AAV/CRISPR vectors for blocking HIV-1 expression in persistently infected astrocytes," *Glia*, vol. 66, no. 2, pp. 413–427, 2018, doi: 10.1002/glia.23254.

[170] C. Krendl *et al.*, "GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 45, pp. E9579–E9588, 07 2017, doi: 10.1073/pnas.1708341114.

[171] F. Matheus *et al.*, "Pathological ASXL1 Mutations and Protein Variants Impair Neural Crest Development," *Stem Cell Reports*, vol. 12, no. 5, pp. 861–868, May 2019, doi: 10.1016/j.stemcr.2019.03.006.

[172] X. Lian *et al.*, "Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions," *Nat Protoc*, vol. 8, no. 1, pp. 162–175, Jan. 2013, doi: 10.1038/nprot.2012.150.

[173] M. Kajiwara *et al.*, "Donor-dependent variations in hepatic differentiation from human-induced pluripotent stem cells," *PNAS*, vol. 109, no. 31, pp. 12538–12543, Jul. 2012, doi: 10.1073/pnas.1209979109.

[174] S. Konishi *et al.*, "Directed Induction of Functional Multi-ciliated Cells in Proximal Airway Epithelial Spheroids from Human Pluripotent Stem Cells," *Stem Cell Reports*, vol. 6, no. 1, pp. 18–25, Jan. 2016, doi: 10.1016/j.stemcr.2015.11.010.

[175] Y. Shi, P. Kirwan, and F. J. Livesey, "Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks," *Nat Protoc*, vol. 7, no. 10, pp. 1836–1846, Oct. 2012, doi: 10.1038/nprot.2012.116.

[176] L. A. Gonçalves, A. M. Vigário, and C. Penha-Gonçalves, "Improved isolation of murine hepatocytes for in vitro malaria liver stage studies," *Malaria Journal*, vol. 6, no. 1, p. 169, Dec. 2007, doi: 10.1186/1475-2875-6-169.

[177] T. Trcek, T. Lionnet, H. Shroff, and R. Lehmann, "mRNA quantification using single-molecule FISH in Drosophila embryos," *Nat Protoc*, vol. 12, no. 7, pp. 1326–1348, Jul. 2017, doi: 10.1038/nprot.2017.030.

[178] S. Klemm *et al.*, "Transcriptional profiling of cells sorted by RNA abundance," *Nature Methods*, vol. 11, no. 5, pp. 549–551, May 2014, doi: 10.1038/nmeth.2910.

[179] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the CRISPR-Cas9 system," *Nat Protoc*, vol. 8, no. 11, pp. 2281–2308, Nov. 2013, doi: 10.1038/nprot.2013.143.

[180] V. Heurtier *et al.*, "The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells," *Nat Commun*, vol. 10, no. 1, pp. 1–15, Mar. 2019, doi: 10.1038/s41467-019-09041-z.

[181] E. Afgan *et al.*, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W537–W544, 02 2018, doi: 10.1093/nar/gky379.

[182] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions," *Genome Biology*, vol. 14, no. 4, p. R36, Apr. 2013, doi: 10.1186/gb-2013-14-4-r36.

[183] C. Trapnell *et al.*, "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nat. Biotechnol.*, vol. 28, no. 5, pp. 511–515, May 2010, doi: 10.1038/nbt.1621.

[184] M.-R. Rafiee, C. Girardot, G. Sigismondo, and J. Krijgsveld, "Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins," *Molecular Cell*, vol. 64, no. 3, pp. 624–635, Nov. 2016, doi: 10.1016/j.molcel.2016.09.019.

[185] C. S. Hughes, S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz, and J. Krijgsveld, "Ultrasensitive proteome analysis using paramagnetic bead technology," *Mol. Syst. Biol.*, vol. 10, p. 757, Oct. 2014, doi: 10.15252/msb.20145625.

[186] Q. Sun, Q. Hao, and K. V. Prasanth, "Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression," *Trends in Genetics*, vol. 34, no. 2, pp. 142–157, Feb. 2018, doi: 10.1016/j.tig.2017.11.005.

[187] T. L Ramos *et al.*, "MSC surface markers (CD44, CD73, and CD90) can identify human MSC-derived extracellular vesicles by conventional flow cytometry," *Cell Commun. Signal*, vol. 14, p. 2, Jan. 2016, doi: 10.1186/s12964-015-0124-8.

[188] M. E. Bronner and N. M. LeDouarin, "Development and evolution of the neural crest: an overview," *Dev. Biol.*, vol. 366, no. 1, pp. 2–9, Jun. 2012, doi: 10.1016/j.ydbio.2011.12.042.

[189] R. Bajpai *et al.*, "CHD7 cooperates with PBAF to control multipotent neural crest formation," *Nature*, vol. 463, no. 7283, pp. 958–962, Feb. 2010, doi: 10.1038/nature08733.

[190] A. H. Fox and A. I. Lamond, "Paraspeckles," *Cold Spring Harb Perspect Biol*, vol. 2, no. 7, Jul. 2010, doi: 10.1101/cshperspect.a000687.

[191] A. H. Fox and A. I. Lamond, "Paraspeckles," *Cold Spring Harb Perspect Biol*, vol. 2, no. 7, p. a000687, Jul. 2010, doi: 10.1101/cshperspect.a000687.

[192] J. R. Tollervey *et al.*, "Characterizing the RNA targets and position-dependent splicing regulation by TDP-43," *Nat. Neurosci.*, vol. 14, no. 4, pp. 452–458, Apr. 2011, doi: 10.1038/nn.2778.

[193] C. A. Gifford *et al.*, "Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells," *Cell*, vol. 153, no. 5, pp. 1149–1163, May 2013, doi: 10.1016/j.cell.2013.04.037.

[194] H. Chen, X. Liu, and D. J. Patel, "DNA bending and unwinding associated with actinomycin D antibiotics bound to partially overlapping sites on DNA," *J. Mol. Biol.*, vol. 258, no. 3, pp. 457–479, May 1996, doi: 10.1006/jmbi.1996.0262.

[195] S. Kamitori and F. Takusagawa, "Crystal structure of the 2:1 complex between d(GAAGCTTC) and the anticancer drug actinomycin D," *J. Mol. Biol.*, vol. 225, no. 2, pp. 445–456, May 1992, doi: 10.1016/0022-2836(92)90931-9.

[196] F. Brueckner and P. Cramer, "Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation," *Nature Structural & Molecular Biology*, vol. 15, no. 8, pp. 811–818, Aug. 2008, doi: 10.1038/nsmb.1458.

[197] M. Sastry, R. Fiala, and D. Patel, "Solution Structure of Mithramycin Dimers Bound to Partially Overlapping Sites on DNA," *Journal of Molecular Biology*, vol. 251, no. 5, pp. 674–689, Aug. 1995, doi: 10.1006/jmbi.1995.0464.

[198] M. Sriram, G. A. van der Marel, H. L. Roelen, J. H. van Boom, and A. H. Wang, "Conformation of B-DNA containing O6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[e6G]AATTCGCG complexed with Hoechst 33258 or Hoechst 33342," *EMBO J.*, vol. 11, no. 1, pp. 225–232, Jan. 1992.

[199] M. B. Clark *et al.*, "Genome-wide analysis of long noncoding RNA stability," *Genome Res*, vol. 22, no. 5, pp. 885–898, May 2012, doi: 10.1101/gr.131037.111.

[200] S. J. Robles and G. R. Adami, "Agents that cause DNA double strand breaks lead to p16INK4a enrichment and the premature senescence of normal fibroblasts," *Oncogene*, vol. 16, no. 9, pp. 1113–1123, Mar. 1998, doi: 10.1038/sj.onc.1201862.

[201] L. J. Kuo and L.-X. Yang, "Gamma-H2AX - a novel biomarker for DNA double-strand breaks," *In Vivo*, vol. 22, no. 3, pp. 305–309, Jun. 2008.

[202] M. J. Wagner, J. A. Livingston, S. R. Patel, and R. S. Benjamin, "Chemotherapy for Bone Sarcoma in Adults," *J Oncol Pract*, vol. 12, no. 3, pp. 208–216, Mar. 2016, doi: 10.1200/JOP.2015.009944.

[203] R. A. Stanton, K. M. Gernert, J. H. Nettles, and R. Aneja, "Drugs That Target Dynamic Microtubules: A New Molecular Perspective," *Med Res Rev*, vol. 31, no. 3, pp. 443–481, May 2011, doi: 10.1002/med.20242.

[204] A. G. Patel and S. H. Kaufmann, "How does doxorubicin work?," *eLife*, vol. 1, Dec. 2012, doi: 10.7554/eLife.00387.

[205] J. L. Nitiss, "Targeting DNA topoisomerase II in cancer chemotherapy," *Nat Rev Cancer*, vol. 9, no. 5, pp. 338–350, May 2009, doi: 10.1038/nrc2607.

[206] R. Chen, M. J. Keating, V. Gandhi, and W. Plunkett, "Transcription inhibition by flavopiridol: mechanism of chronic lymphocytic leukemia cell death," *Blood*, vol. 106, no. 7, pp. 2513–2519, Oct. 2005, doi: 10.1182/blood-2005-04-1678.

[207] K. C. Bible *et al.*, "Flavopiridol binds to duplex DNA," *Cancer Res.*, vol. 60, no. 9, pp. 2419–2428, May 2000.

[208] C.-C. Wu *et al.*, "Structural basis of type II topoisomerase inhibition by the anticancer drug etoposide," *Science*, vol. 333, no. 6041, pp. 459–462, Jul. 2011, doi: 10.1126/science.1204117.

[209] V. Turinetto *et al.*, "High basal γH2AX levels sustain self-renewal of mouse embryonic and induced pluripotent stem cells," *Stem Cells*, vol. 30, no. 7, pp. 1414–1423, Jul. 2012, doi: 10.1002/stem.1133.

[210] A. Bajc Česnik *et al.*, "Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies," *J. Cell. Sci.*, vol. 132, no. 5, Mar. 2019, doi: 10.1242/jcs.224303.

[211] T. Yamazaki *et al.*, "Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation," *Mol. Cell*, vol. 70, no. 6, pp. 1038-1053.e7, 21 2018, doi: 10.1016/j.molcel.2018.05.019.

[212] D. E. Buchholz, B. O. Pennington, R. H. Croze, C. R. Hinman, P. J. Coffey, and D. O. Clegg, "Rapid and efficient directed differentiation of human pluripotent stem cells into retinal pigmented epithelium," *Stem Cells Transl Med*, vol. 2, no. 5, pp. 384–393, May 2013, doi: 10.5966/sctm.2012-0163.

[213] S. M. Chambers *et al.*, "Combined small molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors," *Nat Biotechnol*, vol. 30, no. 7, pp. 715–720, Jul. 2012, doi: 10.1038/nbt.2249.

[214] P. Perez-Pinera *et al.*, "RNA-guided gene activation by CRISPR-Cas9-based transcription factors," *Nat. Methods*, vol. 10, no. 10, pp. 973–976, Oct. 2013, doi: 10.1038/nmeth.2600.

[215] M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, and R. D. Vale, "A protein-tagging system for signal amplification in gene expression and fluorescence imaging," *Cell*, vol. 159, no. 3, pp. 635–646, Oct. 2014, doi: 10.1016/j.cell.2014.09.039.

[216] D. James, A. J. Levine, D. Besser, and A. Hemmati-Brivanlou, "TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells," *Development*, vol. 132, no. 6, pp. 1273–1282, Mar. 2005, doi: 10.1242/dev.01706.

[217] A. J. Levine and A. H. Brivanlou, "GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos," *Development*, vol. 133, no. 2, pp. 209–216, Jan. 2006, doi: 10.1242/dev.02192.

[218] D. L. Bentley, "Coupling mRNA processing with transcription in time and space," *Nat Rev Genet*, vol. 15, no. 3, pp. 163–175, Mar. 2014, doi: 10.1038/nrg3662.

[219] M. Heidemann, C. Hintermair, K. Voß, and D. Eick, "Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription," *Biochim. Biophys. Acta*, vol. 1829, no. 1, pp. 55–62, Jan. 2013, doi: 10.1016/j.bbagrm.2012.08.013.

[220] A. M. Eastham *et al.*, "Epithelial-mesenchymal transition events during human embryonic stem cell differentiation," *Cancer Res.*, vol. 67, no. 23, pp. 11254–11262, Dec. 2007, doi: 10.1158/0008-5472.CAN-07-2253.

[221] A. Moussaieff *et al.*, "Glycolysis-Mediated Changes in Acetyl-CoA and Histone Acetylation Control the Early Differentiation of Embryonic Stem Cells," *Cell Metabolism*, vol. 21, no. 3, pp. 392–402, Mar. 2015, doi: 10.1016/j.cmet.2015.02.002.

[222] C. Ma *et al.*, "Nono, a bivalent domain factor, regulates Erk signaling and mouse embryonic stem cell pluripotency," *Cell Rep*, vol. 17, no. 4, pp. 997–1007, Oct. 2016, doi: 10.1016/j.celrep.2016.09.078.

[223] J. Wang *et al.*, "RNA-binding protein PSPC1 promotes the differentiation-dependent nuclear export of adipocyte RNAs," *J. Clin. Invest.*, vol. 127, no. 3, pp. 987–1004, Mar. 2017, doi: 10.1172/JCI89484.

[224] S. Chan, E.-A. Choi, and Y. Shi, "Pre-mRNA 3'-end processing complex assembly and function," *Wiley Interdiscip Rev RNA*, vol. 2, no. 3, pp. 321–335, 2011, doi: 10.1002/wrna.54.

[225] V. H. Meller, S. S. Joshi, and N. Deshpande, "Modulation of Chromatin by Noncoding RNA," *Annu. Rev. Genet.*, vol. 49, pp. 673–695, 2015, doi: 10.1146/annurev-genet-112414-055205.

[226] R. Chaudhary *et al.*, "Prosurvival long noncoding RNA PINCR regulates a subset of p53 targets in human colorectal cancer cells by binding to Matrin 3," *eLife*, vol. 6, p. e23244, Jun. 2017, doi: 10.7554/eLife.23244.

[227] L. Chen *et al.*, "Regulatory network analysis of LINC00472, a long noncoding RNA downregulated by DNA hypermethylation in colorectal cancer," *Clin. Genet.*, vol. 93, no. 6, pp. 1189–1198, 2018, doi: 10.1111/cge.13245.

[228] S. W. Cho *et al.*, "Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element," *Cell*, vol. 173, no. 6, pp. 1398-1412.e22, 31 2018, doi: 10.1016/j.cell.2018.03.068.

[229] K. N. Smith, A. M. Singh, and S. Dalton, "Myc represses primitive endoderm differentiation in pluripotent stem cells," *Cell Stem Cell*, vol. 7, no. 3, pp. 343–354, Sep. 2010, doi: 10.1016/j.stem.2010.06.023.

[230] M. Nakagawa *et al.*, "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts," *Nat. Biotechnol.*, vol. 26, no. 1, pp. 101–106, Jan. 2008, doi: 10.1038/nbt1374.

[231] A. Al-Rugeebah, M. Alanazi, and N. R. Parine, "MEG3: an Oncogenic Long Non-coding RNA in Different Cancers," *Pathol. Oncol. Res.*, vol. 25, no. 3, pp. 859–874, Jul. 2019, doi: 10.1007/s12253-019-00614-3.

[232] W. Zhuang, P. Chen, X. Ge, J. Fu, and B. Li, "Upregulation of lncRNA MEG3 Promotes osteogenic differentiation of mesenchymal Stem Cells from Multiple Myeloma By Targeting BMP4 Transcription," *Blood*, vol. 124, no. 21, pp. 3435–3435, Dec. 2014, doi: 10.1182/blood.V124.21.3435.3435.

[233] R. N. Wang *et al.*, "Bone Morphogenetic Protein (BMP) signaling in development and human diseases," *Genes Dis*, vol. 1, no. 1, pp. 87–105, Sep. 2014, doi: 10.1016/j.gendis.2014.07.005.

[234] H.-L. Cao, Z.-J. Liu, P.-L. Huang, Y.-L. Yue, and J.-N. Xi, "lncRNA-RMRP promotes proliferation, migration and invasion of bladder cancer via miR-206," *Eur Rev Med Pharmacol Sci*, vol. 23, no. 3, pp. 1012–1021, Feb. 2019, doi: 10.26355/eurrev_201902_16988.

[235] R. Pardal, A. V. Molofsky, S. He, and S. J. Morrison, "Stem cell self-renewal and cancer cell proliferation are regulated by common networks that balance the activation of proto-oncogenes and tumor suppressors," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 70, pp. 177–185, 2005, doi: 10.1101/sqb.2005.70.057.

[236] M. Kalwa *et al.*, "The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation," *Nucleic Acids Res*, vol. 44, no. 22, pp. 10631–10643, Dec. 2016, doi: 10.1093/nar/gkw802.

[237] V. Pachnis, A. Belayew, and S. M. Tilghman, "Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 81, no. 17, pp. 5523–5527, Sep. 1984, doi: 10.1073/pnas.81.17.5523.

[238] R. Goshen *et al.*, "The expression of the H-19 and IGF-2 genes during human embryogenesis and placental development," *Mol. Reprod. Dev.*, vol. 34, no. 4, pp. 374–379, Apr. 1993, doi: 10.1002/mrd.1080340405.

[239] H. Yoshimura, Y. Matsuda, M. Yamamoto, S. Kamiya, and T. Ishiwata, "Expression and role of long non-coding RNA H19 in carcinogenesis," *Front Biosci (Landmark Ed)*, vol. 23, pp. 614–625, Jan. 2018, doi: 10.2741/4608.

[240] J. E. Wright and R. Ciosk, "RNA-based regulation of pluripotency," *Trends Genet.*, vol. 29, no. 2, pp. 99–107, Feb. 2013, doi: 10.1016/j.tig.2012.10.007.

[241] I. Heo, C. Joo, J. Cho, M. Ha, J. Han, and V. N. Kim, "Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA," *Mol. Cell*, vol. 32, no. 2, pp. 276–284, Oct. 2008, doi: 10.1016/j.molcel.2008.09.014.

[242] N. Komarovsky Gulman, L. Armon, T. Shalit, and A. Urbach, "Heterochronic regulation of lung development via the Lin28-Let-7 pathway," *FASEB J.*, vol. 33, no. 11, pp. 12008–12018, Nov. 2019, doi: 10.1096/fj.201802702R.

[243] M. N. Cabili *et al.*, "Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution," *Genome Biology*, vol. 16, no. 1, p. 20, Jan. 2015, doi: 10.1186/s13059-015-0586-4.

[244] T. Chujo, T. Yamazaki, and T. Hirose, "Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies," *Biochim. Biophys. Acta*, vol. 1859, no. 1, pp. 139–146, Jan. 2016, doi: 10.1016/j.bbagrm.2015.05.007.

[245] T. Hirose, T. Yamazaki, and S. Nakagawa, "Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles," *Wiley Interdisciplinary Reviews: RNA*, vol. 10, no. 6, p. e1545, 2019, doi: 10.1002/wrna.1545.

[246] W. Zakrzewski, M. Dobrzyński, M. Szymonowicz, and Z. Rybak, "Stem cells: past, present, and future," *Stem Cell Res Ther*, vol. 10, Feb. 2019, doi: 10.1186/s13287-019-1165-5.

[247] S. Nakagawa, T. Naganuma, G. Shioi, and T. Hirose, "Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice," *J. Cell Biol.*, vol. 193, no. 1, pp. 31–39, Apr. 2011, doi: 10.1083/jcb.201011110.

[248] M. S. Kukharsky *et al.*, "Long non-coding RNA Neat1 regulates adaptive behavioural response to stress in mice," Neuroscience, preprint, Sep. 2019.

[249] P. Katsel *et al.*, "The expression of long noncoding RNA NEAT1 is reduced in schizophrenia and modulates oligodendrocytes transcription," *npj Schizophr*, vol. 5, no. 1, pp. 1–9, Jan. 2019, doi: 10.1038/s41537-019-0071-2.

[250] S. A. Liddelow and B. A. Barres, "Reactive Astrocytes: Production, Function, and Therapeutic Potential," *Immunity*, vol. 46, no. 6, pp. 957–967, 20 2017, doi: 10.1016/j.immuni.2017.06.006.

[251] P. Zhang, L. Cao, R. Zhou, X. Yang, and M. Wu, "The lncRNA Neat1 promotes activation of inflammasomes in macrophages," *Nat Commun*, vol. 10, no. 1, pp. 1–17, Apr. 2019, doi: 10.1038/s41467-019-09482-6.

[252] M. Wang, Q. Yuan, and L. Xie, "Mesenchymal Stem Cell-Based Immunomodulation: Properties and Clinical Application," *Stem Cells Int*, vol. 2018, Jun. 2018, doi: 10.1155/2018/3057624.

[253] F. M. Wensveen, S. Valentić, M. Šestan, T. T. Wensveen, and B. Polić, "Interactions between adipose tissue and the immune system in health and malnutrition," *Semin. Immunol.*, vol. 27, no. 5, pp. 322–333, Sep. 2015, doi: 10.1016/j.smim.2015.10.006.

[254] C. E. Metzger and S. A. Narayanan, "The Role of Osteocytes in Inflammatory Bone Loss," *Front Endocrinol (Lausanne)*, vol. 10, May 2019, doi: 10.3389/fendo.2019.00285.

[255] R. N. Mukherjee, P. Chen, and D. L. Levy, "Recent advances in understanding nuclear size and shape," *Nucleus*, vol. 7, no. 2, pp. 167–186, Apr. 2016, doi: 10.1080/19491034.2016.1162933.

[256] E. Pena, M. T. Berciano, R. Fernandez, J. L. Ojeda, and M. Lafarga, "Neuronal body size correlates with the number of nucleoli and Cajal bodies, and with the organization of the splicing machinery in rat trigeminal ganglion neurons," *J. Comp. Neurol.*, vol. 430, no. 2, pp. 250–263, Feb. 2001.

[257] S. Marguerat and J. Bähler, "Coordinating genome expression with cell size," *Trends in Genetics*, vol. 28, no. 11, pp. 560–565, Nov. 2012, doi: 10.1016/j.tig.2012.07.003.

[258] D. Chakravarty *et al.*, "The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer," *Nat Commun*, vol. 5, p. 5383, Nov. 2014, doi: 10.1038/ncomms6383.

[259] R. A. Röber, K. Weber, and M. Osborn, "Differential timing of nuclear lamin A/C expression in the various organs of the mouse embryo and the young animal: a developmental study," *Development*, vol. 105, no. 2, pp. 365–378, Feb. 1989.

[260] R. Stick and P. Hausen, "Changes in the nuclear lamina composition during early development of Xenopus laevis," *Cell*, vol. 41, no. 1, pp. 191–200, May 1985, doi: 10.1016/0092-8674(85)90073-x.

[261] Y. Gruenbaum and O. Medalia, "Lamins: the structure and protein complexes," *Current Opinion in Cell Biology*, vol. 32, pp. 7–12, Feb. 2015, doi: 10.1016/j.ceb.2014.09.009.

[262] P. Jevtić and D. L. Levy, "Mechanisms of nuclear size regulation in model systems and cancer," *Adv. Exp. Med. Biol.*, vol. 773, pp. 537–569, 2014, doi: 10.1007/978-1-4899-8032-8_25.

[263] T. A. Dittmer, N. J. Stacey, K. Sugimoto-Shirasu, and E. J. Richards, "LITTLE NUCLEI Genes Affecting Nuclear Morphology in Arabidopsis thaliana," *The Plant Cell*, vol. 19, no. 9, pp. 2793–2803, Sep. 2007, doi: 10.1105/tpc.107.053231.

[264] D. L. Levy and R. Heald, "Nuclear size is regulated by importin α and Ntf2 in Xenopus," *Cell*, vol. 143, no. 2, pp. 288–298, Oct. 2010, doi: 10.1016/j.cell.2010.09.012.

[265] Y. S. Mao, H. Sunwoo, B. Zhang, and D. L. Spector, "Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs," *Nat. Cell Biol.*, vol. 13, no. 1, pp. 95–101, Jan. 2011, doi: 10.1038/ncb2140.

[266] S. P. Shevtsov and M. Dundr, "Nucleation of nuclear bodies by RNA," *Nat. Cell Biol.*, vol. 13, no. 2, pp. 167–173, Feb. 2011, doi: 10.1038/ncb2157.

[267] A. Jain and R. D. Vale, "RNA phase transitions in repeat expansion disorders," *Nature*, vol. 546, no. 7657, pp. 243–247, 08 2017, doi: 10.1038/nature22386.

[268] Y. Lin, B. F. Schmidt, M. P. Bruchez, and C. J. McManus, "Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture," *Nucleic Acids Res.*, vol. 46, no. 7, pp. 3742–3752, 20 2018, doi: 10.1093/nar/gky046.

[269] Y. Li, J. Syed, and H. Sugiyama, "RNA-DNA Triplex Formation by Long Noncoding RNAs," *Cell Chemical Biology*, vol. 23, no. 11, pp. 1325–1333, Nov. 2016, doi: 10.1016/j.chembiol.2016.09.011.

[270] M. Aznab and M. Hematti, "Evaluation of clinical process in osteosarcoma patients treated with chemotherapy including cisplatin, adriamycin, ifosfamide, and etoposide and determination of the treatment sequels in a long-term 11-year follow-up," *J Cancer Res Ther*, vol. 13, no. 2, pp. 291–296, Jun. 2017, doi: 10.4103/0973-1482.199447.

[271] S. Goossens, B. Janssens, G. Vanpoucke, R. De Rycke, J. van Hengel, and F. van Roy, "Truncated isoform of mouse alphaT-catenin is testis-restricted in expression and function," *FASEB J.*, vol. 21, no. 3, pp. 647–655, Mar. 2007, doi: 10.1096/fj.06-6066com.

[272] C. Barbosa, I. Peixeiro, and L. Romão, "Gene Expression Regulation by Upstream Open Reading Frames and Human Disease," *PLOS Genetics*, vol. 9, no. 8, p. e1003529, Aug. 2013, doi: 10.1371/journal.pgen.1003529.

[273] E. de Klerk and P. A. C. 't Hoen, "Alternative mRNA transcription, processing, and translation: insights from RNA sequencing," *Trends in Genetics*, vol. 31, no. 3, pp. 128–139, Mar. 2015, doi: 10.1016/j.tig.2015.01.001.

[274] FANTOM Consortium and the RIKEN PMI and CLST (DGT) *et al.*, "A promoter-level mammalian expression atlas," *Nature*, vol. 507, no. 7493, pp. 462–470, Mar. 2014, doi: 10.1038/nature13182.

[275] G. S. Eakin and A.-K. Hadjantonakis, "Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos," *Nat Protoc*, vol. 1, no. 3, pp. 1145–1153, 2006, doi: 10.1038/nprot.2006.173.

[276] B.-Y. Liao and J. Zhang, "Null mutations in human and mouse orthologs frequently result in different phenotypes," *PNAS*, vol. 105, no. 19, pp. 6987–6992, May 2008, doi: 10.1073/pnas.0800387105.

[277] H. Darr, Y. Mayshar, and N. Benvenisty, "Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features," *Development*, vol. 133, no. 6, pp. 1193–1201, Mar. 2006, doi: 10.1242/dev.02286.

[278] A. Saunders, F. Faiola, and J. Wang, "Concise Review: Pursuing Self-Renewal and Pluripotency with the Stem Cell Factor Nanog," *STEM CELLS*, vol. 31, no. 7, pp. 1227–1236, 2013, doi: 10.1002/stem.1384.

[279] X. Zhang *et al.*, "Pax6 is a human neuroectoderm cell fate determinant," *Cell Stem Cell*, vol. 7, no. 1, pp. 90–100, Jul. 2010, doi: 10.1016/j.stem.2010.04.017.

[280] G. Huang, S. Ye, X. Zhou, D. Liu, and Q.-L. Ying, "Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network," *Cell Mol Life Sci*, vol. 72, no. 9, pp. 1741–1757, May 2015, doi: 10.1007/s00018-015-1833-2.

[281] C. Adriaens *et al.*, "The lncRNA NEAT1_1 is seemingly dispensable for normal tissue homeostasis and cancer cell growth.," *RNA*, p. rna.071456.119, Sep. 2019, doi: 10.1261/rna.071456.119.

[282] M. Isobe *et al.*, "Forced isoform switching of Neat1_1 to Neat1_2 leads to the loss of Neat1_1 and the hyperformation of paraspeckles but does not affect the development and growth of mice," *RNA*, Dec. 2019, doi: 10.1261/rna.072587.119.

[283] G. Pineda *et al.*, "Proteomics studies of the interactome of RNA polymerase II C-terminal repeated domain," *BMC Research Notes*, vol. 8, no. 1, p. 616, Oct. 2015, doi: 10.1186/s13104-015-1569-y.

[284] A. Möller *et al.*, "Proteomic Analysis of Mitotic RNA Polymerase II Reveals Novel Interactors and Association With Proteins Dysfunctional in Disease," *Mol Cell Proteomics*, vol. 11, no. 6, Jun. 2012, doi: 10.1074/mcp.M111.011767.

[285] T. Naganuma and T. Hirose, "Paraspeckle formation during the biogenesis of long non-coding RNAs," *RNA Biol*, vol. 10, no. 3, pp. 456–461, Mar. 2013, doi: 10.4161/rna.23547.

[286] L.-L. Chen and L. Yang, "ALUternative Regulation for Gene Expression," *Trends in Cell Biology*, vol. 27, no. 7, pp. 480–490, Jul. 2017, doi: 10.1016/j.tcb.2017.01.002.

[287] K. Bahar Halpern *et al.*, "Nuclear Retention of mRNA in Mammalian Tissues," *Cell Reports*, vol. 13, no. 12, pp. 2653–2662, Dec. 2015, doi: 10.1016/j.celrep.2015.11.036.

[288] D. R. Cooper, G. Carter, P. Li, R. Patel, J. E. Watson, and N. A. Patel, "Long Non-Coding RNA NEAT1 Associates with SRp40 to Temporally Regulate PPARγ2 Splicing during Adipogenesis in 3T3-L1 Cells," *Genes (Basel)*, vol. 5, no. 4, pp. 1050–1063, Nov. 2014, doi: 10.3390/genes5041050.

[289] L. A. Lowery, J. Rubin, and H. Sive, "Whitesnake/sfpq is required for cell survival and neuronal development in the zebrafish," *Dev. Dyn.*, vol. 236, no. 5, pp. 1347–1357, May 2007, doi: 10.1002/dvdy.21132.

[290] J. G. Patton, S. A. Mayer, P. Tempst, and B. Nadal-Ginard, "Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing," *Genes Dev.*, vol. 5, no. 7, pp. 1237–1251, Jul. 1991, doi: 10.1101/gad.5.7.1237.

[291] T. Hall-Pogar, S. Liang, L. K. Hague, and C. S. Lutz, "Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR," *RNA*, vol. 13, no. 7, pp. 1103–1115, Jul. 2007, doi: 10.1261/rna.577707.

[292] C. Mora Gallardo *et al.*, "Dido3-dependent SFPQ recruitment maintains efficiency in mammalian alternative splicing," *Nucleic Acids Res*, vol. 47, no. 10, pp. 5381–5394, Jun. 2019, doi: 10.1093/nar/gkz235.

[293] A. Takeuchi *et al.*, "Loss of Sfpq Causes Long-Gene Transcriptopathy in the Brain," *Cell Reports*, vol. 23, no. 5, pp. 1326–1341, May 2018, doi: 10.1016/j.celrep.2018.03.141.

[294] S.-T. Yang, A. J. B. Kreutzberger, J. Lee, V. Kiessling, and L. K. Tamm, "The Role of Cholesterol in Membrane Fusion," *Chem Phys Lipids*, vol. 199, pp. 136–143, Sep. 2016, doi: 10.1016/j.chemphyslip.2016.05.003.

[295] J.-Y. Moon, M. H. Choi, and J. Kim, "Metabolic profiling of cholesterol and sex steroid hormones to monitor urological diseases," *Endocr Relat Cancer*, vol. 23, no. 10, pp. R455–R467, Oct. 2016, doi: 10.1530/ERC-16-0285.

[296] J. Cabral-Teixeira, A. Martinez-Fernandez, W. Cai, A. Terzic, M. Mercola, and E. Willems, "Cholesterol-derived glucocorticoids control early fate specification in embryonic stem cells," *Stem Cell Res*, vol. 15, no. 1, pp. 88–95, Jul. 2015, doi: 10.1016/j.scr.2015.04.010.

[297] K. Saito *et al.*, "Ablation of cholesterol biosynthesis in neural stem cells increases their VEGF expression and angiogenesis but causes neuron apoptosis," *PNAS*, vol. 106, no. 20, pp. 8350–8355, May 2009, doi: 10.1073/pnas.0903541106.

[298] N. M. F. S. A. Cerqueira *et al.*, "Cholesterol Biosynthesis: A Mechanistic Overview," *Biochemistry*, vol. 55, no. 39, pp. 5483–5506, Oct. 2016, doi: 10.1021/acs.biochem.6b00342.

[299] D. Guallar *et al.*, "RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells," *Nat. Genet.*, vol. 50, no. 3, pp. 443–451, 2018, doi: 10.1038/s41588-018-0060-9.

[300] P. J. Tesar *et al.*, "New cell lines from mouse epiblast share defining features with human embryonic stem cells," *Nature*, vol. 448, no. 7150, pp. 196–199, Jul. 2007, doi: 10.1038/nature05972.

[301] M. Simunovic *et al.*, "A 3D model of a human epiblast reveals BMP4-driven symmetry breaking," *Nat. Cell Biol.*, vol. 21, no. 7, pp. 900–910, 2019, doi: 10.1038/s41556-019-0349-7.

[302] S. Alberti and D. Dormann, "Liquid–Liquid Phase Separation in Disease," *Annual Review of Genetics*, vol. 53, no. 1, pp. 171–194, 2019, doi: 10.1146/annurev-genet-112618-043527.

[303] N. Sentürk Cetin, C.-C. Kuo, T. Ribarska, R. Li, I. G. Costa, and I. Grummt, "Isolation and genome-wide characterization of cellular DNA:RNA triplex structures," *Nucleic Acids Res*, vol. 47, no. 5, pp. 2306–2321, Mar. 2019, doi: 10.1093/nar/gky1305.

6. Appendix

6.1 Supplementary Table 1. Chemicals, reagents and solutions routinely used in this study.

Reagent	cat. #	Supplier
16% Formaldehyde (w/v), Methanol-free	10321714	Thermo Fisher Scientific
2-Mercaptoethanol	M3148	Sigma
Albumin from bovine serum (BSA)	A9647-10G	Sigma
Ampicillin sodium salt	A-166	Sigma
Ammonium sulfate solution (3.2M)	sc-291897	Santa Cruz
Biozym LE Agarose	840004	Biozym
Boric acid, electrophoresis grade	15166.02	Serva Electrophoresis
DNA Gel loading dye, 6x	R0611	Thermo Fisher Scientific
Dynabeads(R) Protein A for Immunoprecipitation	10001D	Life Technologies
EDTA Dinatriumsalz Dihydrat >99%	X986.1	Carl Roth
Ethanol, 99.8%	9065.2	Carl Roth
Deionized Formamide	4610	Merck Millipore
Ultra pure glycerol	15514011	Thermo Fisher Scientific
Glycine	23391.02	Serva Electrophoresis
HEPES, 1M Buffer Solution	15630122	Life Technologies
Isopropanol	6752.2	Carl Roth
Lithium chloride (LiCl)	62480-500G-F	Sigma
Lithium Acetate dihydrate, 98%	15157442	Thermo Fisher Scientific
Methanol	45631.02	Serva Electrophoresis
Magenesium chloride (MgCl$_2$)	KK36	Carl Roth
Na-deoxycholate	D6750-10G	Sigma
NaHCO3	S5761	Sigma
Tergitol type NP-40 70% solution	NP40S	Sigma
Nuclease-free water (H$_2$O)	AM9932	Life Technologies
Powder Milk, blotting grade	T145.1	Carl Roth
Protease Inhibitor Cocktail Set III, EDTA-Free	539134	Merk Millipore
SDS Solution, 20 %	20768.02	Serva Electrophoresis
Sodium Chloride (NaCl)	P029.2	Carl Roth
Sodium Citrate	S4641	Sigma
SYBR® Safe DNA Gel Stain	5001208	Life Technologies
TE buffer, pH 7.4, RNAse free	93302	Sigma
TRIS PUFFERAN®	5429.3	Carl Roth
Triton™ X-100	X100-500ML	Sigma-Aldrich
Tween(R)-20	P9416	Sigma-Aldrich

6.2 Supplementary Table 2. Kits routinely used in this study.

Kit	cat. #	Supplier
GeneJET Plasmid Miniprep Kit	K0502	Fermentas
P3 Primary Cell 4D-Nucleofector® X Kit	V4XP-3024	Lonza
PureLink HiPure Plasmid Filter Maxiprep Kit	K210017	Life Technologies
QIAprep Spin Miniprep Kit	27104	Qiagen
QIAquick PCR Purification Kit	28104	Qiagen
QIAquick Gel Extraction Kit	28704	Qiagen
RNeasy MinElute cleanup kit	74204	Qiagen
RNeasy Mini Kit	4104	Qiagen

6.3 Supplementary Table 3: List of primary antibodies.

Targeted protein	company	Clone	Catalogue number
SFPQ	Thermo Fisher Scientific	B92	MA1-25325
NONO	Abcam	polyclonal	ab70335
PSPC1	Thermo Fisher Scientific	1L4	SAB4200503
OCT4	CST	C30A3	2840
NANOG	CST	D73G4	4903
SOX2	CST	D6D9	3579
TRA1-60	Abcam	2A6	ab16288
SSEA5	Hybridoma supernatant	8.00E+11	---
FOXG1	Abcam	polyclonal	ab18259
PAX6	Biolegend	polyclonal	901301
CD24-PE	Miltenyi	32D12	130-095-953
CD73-PE	BD Biosciences	AD2	550257
CD90-PC5	Beckman Coulter	Thy-1/310	IM3703
CXCR4-PE	LifeTechnologies	S3.5	MHCXCR404
CD117-APC	LifeTechnologies	104D2	CD11705
EPCAM-APC	LifeTechnologies	G8.8	17-5791-82
SIX2	Santa Cruz	H-4	sc-377193
PAX2	Santa Cruz	G-3	sc-377181
CDH5	Santa Cruz	F-8	sc-9989
WT1	Santa Cruz	H-1	sc-393498
SOX1	R&D Systems	polyclonal	AF3369
NESTIN	R&D Systems	polyclonal	MAB1259
PAX6	Merck Millipore	polyclonal	ab2237
HNF4A	Sigma Aldrich	3C6	SAB1412164
ALB	Abcam	polyclonal	ab106582
AFP	Sigma Aldrich	1G7	WH0000174M1
GFAP	CST	D1F4Q	12389
SOX9	CST	D8G8H	82630
MNX1	Merck Millipore	polyclonal	ABN174
TUBB3	CST	D71G9	5558
CHAT	Abcam	polyclonal	Ab18736
NFH	CST	RMdO 20	2836
ISL1	Abcam	polyclonal	Ab20670
MAP2	Sigma Aldrich	HM-2	M4403
TBR1	Abcam	polyclonal	ab31940
VIM	Santa Cruz Biotechnology	V-9	sc-6260
HSP47	Santa Cruz Biotechnology	G-12	sc-5293
KRT14	Santa Cruz Biotechnology	LL001	sc-53253
IVL	Santa Cruz Biotechnology	A-5	sc-398952
CAS9	CST	7A9-3A3	14697
Phospho RNA-PolymeraseII (S2)	Bethyl Laboratories	polyclonal	A304-407A
Phospho H2A.X (Ser139)	CST	20E3	9718
IgG rabbit	GeneTex	polyclonal	GTX35035
IgG1 kappa mouse	eBioscience	P3.6.2.8.1	14-4714-82

6.4 Supplementary Table 4: Sequences of custom-designed smFISH oligonucleotides.

Mouse NEAT1 5`end	Human NEAT1 5`end
cttctggaagctgaaagggt	caagttgaagattagccctc
tccagacaactaggccaaag	agcccttggtctggaaaaaa
aggtatccgtatcatgatgt	aagttcagttccacaagacc
caacagtcacactgttgtct	caggccgagcgaaaattaca
tctgtttgccaagagatctg	ctgtcaaacatgctaggtgc
ggtggacatgtcttttcatg	aagcgttggtcaatgttgtc
cttacacagtgtctttctgg	gtggagtgagctcacaagaa
acccttcttatacacaggaa	cttaccagatgaccaggtaa
tggctagaaaccctacaagg	ttaccaacaataccgactcc
aacatacccaccagagacaa	cggtccatgaagcatttttg
aggagagatgcatgggagat	tcgccatgaggaacactata
acaaaagcagctcctcagat	atctgcaggcatcaattgag
taattattgtcctctgcgtc	agcaaggcctggaaacagaa
tagccttcaaactacctgta	catctgctgtggactttta
acacagaagacagttaccgc	ttcatgggctctggaacaag
caagacagtgtgtgatcctt	gatgcagcatctgaaaacct
ttacagctcagtgttaaggc	aaactagtatgaccggaggc
gttaactcagcatcttgtct	ttgaagcaaggttccaagca
ctctaaggaaacatccctgt	tgttctacagcttagggatc
tgaaaccatcagtgtgacgc	tacaaggcatcaatctgcgt
tatcctgacattcaggtagg	caaacaggtgggtaggtgag
agacctcttaatcagctcta	cttctccgagaaacgcacaa
atagctgtgtactctgtagg	ccaagttatttcatcaggct
agactggatgtcttcagagt	tctaatatatccccagtcta
aagggcagaagcagagcaag	cacaacacaatgacaccctt
cattcttctgtagttacctc	caaactagacctgccatttc
gtgatttctattctcactcc	ctcctagtaatctgcaatgc
aacgccattcaaacctttca	aaagagcactaccggtgtac
ccttcaaccaacaaccacaa	tcctcttactagaatgccaa
tctcagtgttagtagctagg	ctaagcaacttctcacttcc
	taacacttcttcagtcttcc
	cctttggttctcggaaaact
	tgtgagatggcatcacacac
	ccaggaggaagctggtaaag
	ctctgaaacaggctgtcttg
	tcacttgataacacccacac
	cagcgaaggatgctgatctg
	atcaaccacctaagttgcta
	gtggtcccttaaatacgtta
	agaagagcccatctaatctc
	gatgtgtttctaaggcacga
	ggtcttgttttccaaactga
	catgtagtaaaggcacctcg
	ccattggtattactttacca
	ctctaaatcccaacgacagt
	atttcacaacagcatacccg
	ccagtactttcaaccatcta
	agttcttaccatacagagca

6.5 Supplementary Table 5: List of primers used for RT-qPCR.

Name	Primer sequence	Name	Primer sequence
GAPDH_Fwd	GCTCATTTCCTGGTATGACAACG	AFP_Fwd	GCTTACACAAAGAAAGCCC
GAPDH_Rev	GAGATTCAGTGTGGTGGGGG	AFP_Rev	TAATAATGTCAGCCGCTCC
OCT4_Fwd	CAATTTGCCAAGCTCCTGAAG	HNF4A_Fwd	TACTCCTGCAGATTTAGCCG
OCT4_Rev	AAAGCGGCAGATGGTCGTT	HNF4A_Rev	GCATTTCTTGAGCCTGCAG
NANOG_Fwd	CCTTCCTCCATGGATCTGCTT	ALB_Fwd	GCCAAGACATATGAAACCAC
NANOG_Rev	CTTGACCGGGACCTTGTCTTC	ALB_Rev	TTCATCGAACACTTTGGCA
SOX2_Fwd	CCTCCGGGACATGATCAGCATGTA	ASGR1_Fwd	ACGTTCAGCAACTTCACAG
SOX2_Rev	GCAGTGTGCCGTTAATGGCCGTG	ASGR1_Rev	TTTCTTCCCACATTGCCTC
T_Fwd	CAACCTCACTGACGGTGAAAAA	CYP3A4_Fwd	TTGTCCTACCATAAGGGCT
T_Rev	ACAAATTCTGGTGTGCCAAAGTT	CYP3A4_Rev	GATCTGTGATAGCCAGCAC
MESP1_Fwd	CTGCCTGAGGAGCCCAAGT	DHCR7_Fwd	CTCATCAACCTGTCCTTCG
MESP1_Rev	GCAGTCTGCCAAGGAACCA	DHCR7_Rev	CAATCACGTAGATGGCCTG
MIXL1_Fwd	CCGAGTCCAGGATCCAGGTA	MVK_Fwd	GTACATGGCAAGGTAGCAC
MIXL1_Rev	CTCTGACGCCGAGACTTGG	MVK_Rev	CACTTTCCCATTGCTGTGG
FZD4_Fwd	TACAACGTGACCAAGATGC	MSMO1_Fwd	CTGCATAGACTCTTACACCAC
FZD4_Rev	AAAGGAAGAACTGCAGCTG	MSMO1_Rev	CCATTCCAAATGGAGCCTG
TWIST_Fwd	GAGCTGGACTCCAAGATGG	MVD_Fwd	TGCTCATCCTTGTGGTGAG
TWIST_Rev	TTAAGAAATCTAGGTCTCCGGC	MVD_Rev	TGTCCTTCATGGTCAGCTG
SLUG_Fwd	CACATTAGAACTCACACGGG	SQLE_Fwd	CTTAGAAGCCACTGACAATTCTC
SLUG_Rev	CAAATGCTCTGTTGCAGTG	SQLE_Rev	GAAGAACACCTCGTTTCTTCAC
SNAIL_Fwd	TCTTTCCTCGTCAGGAAGC	DHCR24_Fwd	GTGAAACACTTTGAAGCCAG
SNAIL_Rev	AGGTAAACTCTGGATTAGAGTCC	DHCR24_Rev	ATACAGCATCTGGAAGCCA
PPARG_Fwd	TTCCATTCACAAGAACAGATCC	EBP_Fwd	TACGAAGACCTGCTTGGAG
PPARG_Rev	CTTTGATTGCACTTTGGTACTC	EBP_Rev	TTGTCACCCAGGATGTATCG
GATA6_Fwd	GACTTGCTCTGGTAATAGCA	SCAP_Fwd	ATCTTAGCCTGCTGCTACC
GATA6_Rev	CTGTAGGTTGTGTTGTGGG	SCAP_Rev	CTTGTTTGCGGTCAGAGTC
SIM2_Fwd	CTTATCCCAGGTGGAGCTC	PCSK9_Fwd	AAGTGTGACAGTCATGGCA
SIM2_Rev	CGAAGAAAGAACGACCTCTC	PCSK9_Rev	AAACTCCAGGCCTATGAGG
FOXP1_Fwd	TGCTCAAGGCATGATTCCA	CYP51A1_Fwd	TACTAGATGCTACATACAAGGATGG
FOXP1_Rev	CCTGTGGTTTCTTCTGCAG	CYP51A1_Rev	CTGCCAAGAGTAATCCAATAAGC
SOX11_Fwd	ACGCAGGAAGATCATGGAG	INSIG1_Fwd	CACGCCAGTGCTAAATTGG
SOX11_Rev	CAGCCTCTTGGAGATCTCG	INSIG1_Rev	CAAATGTCCACCAAAGGCC
CXCR4_Fwd	GAGCCCTCAGATTTGACCTGTC	HMGCR_Fwd	GGGAATTGTCACTTATGGCAG
CXCR4_Rev	CACCGCATCTGGAGAACCA	HMGCR_Rev	AATTGATCTTCGACCTGTTGTG
SOX17_Fwd	GCCCATTTCCTCGGTGTAGTT	HMGCS1_Fwd	CTAGCACAGTACTCACCTCAG
SOX17_Rev	GGCGCAGCAGAATCCAGA	HMGCS1_Rev	GAGAGTACAGAGTGGCAGC
FOXA2_Fwd	CCACGACTTGCCCAGCAT	NPC2_Fwd	CCTTCACCAGCAATATTCAGTC
FOXA2_Rev	GGGAGCGGTGAAGATGGA	NPC2_Rev	AATGGGAAAGGGAACTGGG
NKX2.1_Fwd	CTTCCCCGCCATCTCCCGCTTC	FDPS_Fwd	GGTAGTAGCATTCCGGGAG
NKX2.1_Rev	GCCGACAGGTACTTCTGTTGCTTG	FDPS_Rev	GGAAGAAAGCTTGCAGCAG
PAX6_Fwd	GCGGAGTTATGATACCTACACC	PMVK_Fwd	AGTATGCTCAGGAGCATGG
PAX6_Rev	GAAATGAGTCCTGTTGAAGTGG	PMVK_Rev	GATCATGTCCTTCCGAAAGG
SOX1_Fwd	GAGAACCCCAAGATGCACAA	LSS_Fwd	TCTTCACAAGAAAGGTGGTG
SOX1_Rev	CCTCGGACATGACCTTCCA	LSS_Rev	TCCCAGCTGTAAACATTCAG
ASCL1_Fwd	TTCACCAACTGGTTCTGAG	FDFT1_Fwd	AAACAAACATCATCCGTGAC
ASCL1_Rev	TAAAGATGCAGGTTGTGCG	FDFT1_Rev	CATACCTGCTCCAAACCTC
FOXG1_Fwd	GCTGGACATGGGAGATAGG	TUBB3_Fwd	TCAGCGTCTACTACAACGAGGC
FOXG1_Rev	GTTGATGCTGAACGAGGAC	TUBB3_Rev	GCCTGAAGAGATGTCCAAAGGC
PVT1_Fwd	GGATTTCCTTGCGGAAAGG	MNX1_Fwd	TCATGCTCACCGAGACCCA
PVT1_Rev	GACAGCTATGGTCTGGAGG	MNX1_Rev	TGGGTCACAAGTGCAAAGGTA
KCNQ1OT_Fwd	CTTAAACAGCAACCTACACCA	CHAT_Fwd	CGTAGGCACCTGTAGCTGTTT

KCNQ1OT1_Rev	CATTCATCCATTCTACCACCT	CHAT_Rev	AAAGAGGGTCTATCCTGGGCT
PINT_Fwd	CAGAATAAACCACTGAACAGGA	S100B_Fwd	GAAGAAATCCGAACTGAAGGAGC
PINT_Rev	AAGAGGTAGCTCATCTGCG	S100B_Rev	TCCTGGAAGTCACATTCGCCGT
MALAT1_Fwd	TCATCAGTAGTAAGAATCTCAGGG	SLC1A2_Fwd	CAGCTTAATCACAGGGTTGTC
MALAT1_Rev	GATTATATGTCATACCTCCATTGGG	SLC1A2_Rev	GACATGAATACACCATGGCTC
TERC_Fwd	TTTCTCGCTGACTTTCAGC	SLC1A3_Fwd	CTGTCATTGTGGGTACAATCC
TERC_Rev	CTAGAATGAACGGTGGAAGG	SLC1A3_Rev	GAAAGGAGAAGTACTTGACTTCC
MEG3_Fwd	CATCTACACCTCACGAGGG	TBR1_Fwd	ATGGGCAGATGGTGGTTTTA
MEG3_Rev	ATCCTTTGCCATCCTGGTC	TBR1_Rev	GACGGCGATGAACTGAGTCT
LINC00472_Fwd	TTTCTCGACTCGTCGTCAG	TBR2_Fwd	CACCGCCACCAAACTGAGAT
LINC00472_Rev	GGAGTACCTGAAATCCGCA	TBR2_Rev	CGAACACATTGTAGTGGGCAG
TUG1_Fwd	GAAGACCTGAGTTTCTGTCCA	NEAT1_Fwd	GTGGCTGTTGGAGTCGGTAT
TUG1_Rev	CAAGGAGTCTGCTATCATAATTCAC	NEAT1_Rev	TAACAAACCACGGTCCATGA
KLRA1P_Fwd	AGAATTTCCTGCCGTTGATGT	NEAT1_2_Fwd	GTCTTTCCATCCACTCACGTCTATTT
KLRA1P_Rev	CTGATGATAGTCACAGTGTTGGT	NEAT1_2_Rev	GTACTCTGTGATGGGGTAGTCAGTCAG
PINCR_Fwd	ATGAGGAAAGCTCCTATTCCA	DCX_Fwd	GCCAGGGAGAACAAGGACTTT
PINCR_Rev	ATCTCCTAGGTATACTTCAAGGAC	DCX_Rev	CACCCCACTGCGGATGA
MANTIS_Fwd	AACTCCTGCTCCAAACTCACTC	SOX9_Fwd	AGGAAGCTCGCGGACCAGTAC
MANTIS_Rev	CCAGAGACTTTCCATTCTGATG	SOX9_Rev	GGTGGTCCTTCTTGTGCTGCAC
LncPRESS1_Fwd	CAGTAATTCTCCAGCAACAG	ALDH_Fwd	CAGAGGCCATTCACAACTG
LncPRESS1_Rev	TGGCAGGTAATCATCTCATAT	ALDH_Rev	ATGTCAGTTTCTGTTCACAGG
HAND2-AS_Fwd	CTAGCCTGTTTGAAGGCAC	OTX2_Fwd	CCAGACATCTTCATGCGAG
HAND2-AS_Rev	CTGCGAAAGTGAAGATCCC	OTX2_Rev	TCGATTCTTAAACCATACCTGC
HOTAIR_Fwd	ATCAGAAAGGTCCTGCTCC	OLIG2_Fwd	ATGCACGACCTCAACATCGCCA
HOTAIR_Rev	GTCTGTAACTCTGGGCTCC	OLIG2_Rev	ACCAGTCGCTTCATCTCCTCCA
RMRP_Fwd	CTGAGGACTCTGTTCCTCC	NODAL_Fwd	GCATACATCCAGAGTCTGCT
RMRP_Rev	ATGTCTACGTGCGTATGCA	NODAL_Rev	CACATACAGCATGCTCAGC
PANDA_Fwd	TCTCAAACCTCGACCTCAG	GDF3_Fwd	GAGACTTATGCTACGTAAAGGA
PANDA_Rev	CTGTAATCTCAGCACTTTGGG	GDF3_Rev	GGTAAAGAAAGAAACCTTGGTC
H19_Fwd	CTTGGAAATGAATATGCTGCAC	NOG_Fwd	AGCACTATCTCCACATCCG
H19_Rev	TTCCTCTAGCTTCACCTTCC	NOG_Rev	GATAGGGTCTGGGTGTTCG
TARID_Fwd	GCAACAACTAGATGCTGCT	BMP4_Fwd	CCACCACGAAGAACATCTG
TARID_Rev	TATTGCACTTCTGTGCTTCAG	BMP4_Rev	ATGCTGCTGAGGTTAAAGAG
SLERT_Fwd	TTAGTCAGCTCAGGCCCAGT	SFPQ_Fwd	CACATGAAGTGGATAGATACTTCTC
SLERT_Rev	AAGTGCTCCACCAACTCCAG	SFPQ_Rev	GTTGTCAGTCTGCTTGTGG
FIRRE_Fwd	AGTAGAAATGGGAAGACTTGG	NONO_Fwd	TGAGATGGAGAAGCAGCAG
FIRRE_Rev	CTTAGTGATCCATGCCCTC	NONO_Rev	CTCATCAAATCCTGTCTCATTAGC
ANRASSF1_Fwd	GGCAATTAGAACGCTCCTTG	PSPC1_Fwd	TCATCCGCTTGGAATCCAG
ANRASSF1_Rev	CTGTGCTAGGCGATAGAGATCC	PSPC1_Rev	CGTAGAGGTCTGCTCTTGAG
PARTICLE_Fwd	GGCTCAGTGGGAAACAAAGG	KLF4_Fwd	GGGAGAAGACACTGCGTCA
PARTICLE_Rev	ATGTGGTCACTGAGTCTGGG	KLF4_Rev	GGAAGCACTGGGGGAAGT
APTR_Fwd	AATTGCCGGGAATCAAGTC	KLF2_Fwd	CATCTGAAGGCGCATCTG
APTR_Rev	TACCTGGTGAAGCCTTGTC	KLF2_Rev	CGTGTGCTTTCGGTAGTGG
CER_Fwd	CAGGACAGTGCCCTTCAGCCA	DNMT3L_Fwd	TTCTGGATGTTCGTGGACAA
CER_Rev	ACAGTGAGAGCAGGAGGTATGG	DNMT3L_Rev	ACATCTGGGATGGTGACTGG
EOMES_Fwd	ACAGGAGATTTCATTCGGG	LBP9_Fwd	GCTCTTCAACGCCATCAAA
EOMES_Rev	TTGTAAGACTATCATCTGGGTG	LBP9_Rev	CAGGGGCACTCGATTCTG
NKX2-5_Fwd	TAAACCTGGAACAGCAGCA	PAX2_Fwd	CTCTGCTCTTTGTCCAGCCTC
NKX2-5_Rev	TAGGCACGTGGATAGAAGG	PAX2_Rev	CCTCACAGGTTCCCTTTCTCT
ISL1_Fwd	CAGTATTTGGACGAGAGCTG	SIX2_Fwd	GCACAACCCCTACCCTTCAC
ISL1_Rev	CCCGTACAACCTGATATAATCTC	SIX2_Rev	AGGTCTACTTACTCGTACCTTTCC
TBX6_Fwd	CGTGTGAAGAGGAAACTGCG	PAX8_Fwd	AAGGTGGTGGAGAAGATTGGG
TBX6_Rev	GACTACACTCACCTCCGCTC	PAX8_Rev	AGGCTGCTTTCTCTCTTACCTA
LHX1_Fwd	CCTCGCTCTCTGTAAGCCACT	WTI_Fwd	GGAATAGTGCGTGGCTATCTT

LHX1_Rev	CCTCCCCTTGATTTACTCCCAG	WTI_Rev	AAACAGTAGGGACCTGGCTT
SALL1_Fwd	TTGCTCTCTCTGCACCATACC	OSR1_Fwd	CCTCTGGCCTCACCATCTTTC
SALL1_Rev	TCTCCCCCGTCAACCATGT	OSR1_Rev	TTTCCTGAACCCATGCTCCAA
HOXD11_Fwd	TTCCTGGGCCGTTGTAAAGT	NNAT_Fwd	TGCTGCATTTACTGGGTAGGA
HOXD11_Rev	ACTGGGAAAAGGCTCTCGAC	NNAT1_Rev	CACCGTGTATGCCAGCTTC
GSC_Fwd	GAGGAGAAAGTGGAGGTCTG		
GSC_Rev	CTCCGACTCCTCTGATGAG		

6.6 Supplementary Table 6: Guide RNA sequences and genomic location.

gRNA name	gRNA sequence	Genomic position (Hg19)
NONO$^{-/-}$_up	AGGGGAGAAAATGCGCGCGT	ChrX: 70503075
NONO$^{-/-}$_down	GACCGCCGGAAACGAGACGA	ChrX: 70503477
PSPC1-/-_up	TGCGTGTACGTCTTCTCGCC	Chr13: 20356659
PSPC1-/-_down	AGTGTCAGGCGCCCGCGCGA	Chr13: 20356503
SFPQ$^{-/-}$_up	CTGTGGTCAAGGGGCGGTCG	Chr1: 35658652
SFPQ$^{-/-}$_down	CGAGGAGAAGATCTCGGACT	Chr1: 35657824
NEAT1Δ1150_up	CGAAAGTCACGCGCGCCTCC	Chr11: 65189762
NEAT1_Δ1150_down	CCAGACCTGGACGCTCCACC	Chr11: 65190905
NEAT1Δ2700_up	ACATTTCGCCTGCGTCTGTG	Chr11: 65188930
NEAT1Δ2700_down	CTGCAGGCATCAATTGAGGC	Chr11: 65191548
NEAT1$^{-/-}$_up	GGGGCGGCGCTTTAGAGTTG	Chr11: 65186374
NEAT1$^{-/-}$_down	CTTTGGGGAATTTAGTGCGT	Chr11: 65192559
NEAT1ΔTH_up	TCCCTTGTAAAGGCATAGCC	Chr11: 65212842
NEAT1ΔTH_down	CTGCTCACTCTTTCACAGAT	Chr11: 65212999
NEAT1ΔpA_up	ATGCAAACAATTACTGTCGT	Chr11: 65193724
NEAT1ΔpA _down	TGTTGAGAGTTGGTAATCAT	Chr11: 65194218
NEAT1^{YFP}_cut	GGTCCAGCCGGAGTTAGCGA	Chr11: 65190160
NEAT1^{STOP}_cut	CATCTGAAAACCTTTACCCC	Chr11: 65191779
NEAT1 gRNA#1 (SunTag)	ATACACTGGGGTCCTTGCGT	Chr11: 65190090
NEAT1 gRNA#2 (SunTag)	CTGGGAGACCATGCACCGCC	Chr11: 65190119
NEAT1 gRNA#3 (SunTag)	AGAGACTCCCGGGCGGTGCA	Chr11: 65190130
NEAT1 gRNA#4 (SunTag)	GCACCGCCCGGGAGTCTCTC	Chr11: 65190131
NEAT1 gRNA#5 (SunTag)	TTTGGGAGGCGAATGCCATG	Chr11: 65190015

6.7 Supplementary Table 7: Sequence and genomic location of primers used for screening of genomic deletion.

PCR primer name	Primer sequence	Genomic position (Hg19)
NONO[-/-]_det_F	CCAGCAACAGGAGAAGCATC	ChrX: 70502876
NONO[-/-]_det_R	GCCTCTTCCTTCGCTGATTG	ChrX: 70503648
PSPC1[-/-]_det_F	CAGTGATACGTCTGGTCCGA	Chr13: 20356050
PSPC1[-/-]_det_R	GCAAGTGCGCATTGAGAAAA	Chr13: 20356876
SFPQ[-/-]_det_F	GCCTGCGCTTTTATGGAACTT	Chr1: 35657477
SFPQ[-/-]_det_R	AGGAATGATCAGAGGTTCGCA	Chr1: 35658880
NEAT1Δ1150_det_F	CAGGAGTTCACCAGGTTTGC	Chr11: 65189212
NEAT1Δ1150_det_R	AATACCGACTCCAACAGCCA	Chr11: 65191265
NEAT1Δ2700_det_F	GAATCTTCCCCTGGCAGAGAAACAG	Chr11: 65188748
NEAT1Δ2700_det_R	CTGCTGGCATTCATGGGCTCTGGAAC	Chr11: 65191848
NEAT1[-/-]_det_F	ACCAGCCCACATTAGGTCAA	Chr11: 65185808
NEAT1[-/-]_det_R	CCCACACCCCAAACAAAACA	Chr11: 65192890
NEAT1ΔTH_det_F	CTCGTGAAGGCAGAGGGAG	Chr11: 65212646
NEAT1ΔTH _det_R	CCCAATGCTACCCCTCTAGG	Chr11: 65213152
NEAT1ΔpA_det_F	TGAGCCAAGACTAGAGGGGA	Chr11: 65193434
NEAT1ΔpA_det_R	CCTTGCTGCTCCCTTTGAAA	Chr11: 65194789

6.8 Supplementary Table 8: List of antisense oligonucleotides.

ASO name	ASO sequence	Genomic position (Hg19)
NEAT1_ASO	[mC]*[mC]*[mC]*[mU]*[mC]*T*A*G*T*C*T*T* G*G*C*[mU]*[mC]*[mA]*[mU]*[mU]	Chr11: 65193437
MALAT1_ASO	[mG]*[mG]*[mC]*[mA]*[mT]*A*T*G*C*A*G*A* T*A*A*[mT]*[mG]*[mT]*[mT]*[mC]	Chr11: 65270276
scrambled	[mG]*[mT]*[mT]*[mA]*[mG*T*G*A*T*A*C*G* A*T*G*[mA]*[mT]*[mA]*[mA]*[mA]	---

Asterix = phosphothioate-modified backbone, mN = 2`O-methoxyribonucleotides

7. List of Abbreviations

ASOs	antisense oligonucleotides
bp	base pairs
BSA	bovine serum albumine
CTD	C-terminal domain
dsDNA	double-stranded DNA
EMT	epithelial-to-mesenchymal transition
ESCs	embryonic stem cells
EtOH	ethanol
ICM	inner cell mass
iPSCs	induced pluripotent stem cells
kb:	kilo base pairs
LncRNA	long non-coding RNA
MSCs	mesenchymal stem cells
NEAT1	Nuclear Paraspeckle Assembly Transcript 1
NSCs	neural stem cells
ON	over night
polyA site	polyadenylation site
PCR	polymerase chain reaction
PSCs	pluripotent stem cells
RA	retinoic acid
RBP	RNA binding protein
RNA PolII	RNA polymerase II
RT	room temperature
smFISH	single molecule fluorescence *in situ* hybridization
SEM	standard error of the mean
TSS	transcription start site
WT	wildtype
XCI	X-chromosome inactivation

8. List of Figures

Publications based on this thesis (prepared or published)

- **Nucleus size and DNA accessibility are linked to the regulation of paraspeckle formation in cellular differentiation**. **Markus Grosch**, Sebastian Ittermann, Ejona Rusha, Tobias Greisle, Chaido Ori, Dong-Jiunn Jeffery Truong, Anna Pertek, Adam C. O´Neill, Gil Westmeyer, Micha Drukker. *BMC Biology*, 2020. **Outlined in sections 3.1-3.4**.

- **DBHS proteins regulate pluripotency and differentiation of human pluripotent stem cells**. **Markus Grosch**, Sebastian Ittermann, Ruhita Gupta, Stefanie Hauck, Micha Drukker. *In preparation*. **Outlined in section 3.6.**

- **Nuclear RNA foci from *C9ORF72* expansion mutation form paraspeckle-like bodies**. Ana Bajc Česnik, Simona Darovic, Sonja Prpar Mihevc, Maja Štalekar, Mirjana Malnar, Helena Motaln, Youn-Bok Lee, Julija Mazej, Jure Pohleven, **Markus Grosch**, Miha Modic, Marko Fonovič, Boris Turk, Micha Drukker, Christopher E. Shaw, Boris Rogelj. *Journal of Cell Science*, 2019. **Outlined in section 3.4.1.**

- **Cross-Regulation between TDP-43 and paraspeckles promotes pluripotency-differentiation Transition.** Miha Modic, **Markus Grosch**, Gregor Rot, Silvia Schirge, Tjasa Lepko, Tomohiro Yamazaki, Flora C.Y.Lee, EjonaRusha, DmitryShaposhnikov, Michael Palo, JulianeMerl-Pham, Davide Cacchiarelli, Boris Rogelj, Stefanie M.Hauck, Christian von Mering, Alexander Meissner, Heiko Lickert, Tetsuro Hirose, Jernej Ule, Micha Drukker. *Molecular Cell*, 2019. **Outlined in section 3.2.5.**

- **Neat1 in hematopoietic stem cells.** Noam Fallik, Yael Bar-Lavan, Yariv Greenshpan, Oron Goldstein, **Markus Grosch**, Micha Drukker, and Roi Gazit. *Oncotarget*. 2016. **Not described here.**

Acknowledgements

This work was carried out from October 2015 to December 2019 at the Institute of Stem Cell Research of the Helmholtz Center Munich under supervision of Dr. Micha Drukker and Prof. Dr. Michael Kiebler from the department of cell biology of the Ludwig-Maximilians-Universität in Munich.

I would like to sincerely thank my PhD mentor Dr. Micha Drukker for providing guidance, ideas and most importantly the great scientific freedom that allowed me to pursue this project in many directions. He also taught me the basis of good scientific writing and figure design, a skill that is going to be tremendously useful for my future career. Moreover, I wish to thank Prof. Dr. Michael Kiebler who supported this work in every possible way and to whom I could always turn for guidance and to get an unbiased, fresh perspective. This work was not done in isolation but with the help of many students who I would like to acknowledge by name: Sebastian Ittermann, Tobias Greisle, Paola D`Albora, Soumya Vastrad, Isabel Weisheit, Andrea Neuner, Ruhita Gupta, Hoda Kooshapour and Alexandra de la Porte. Even though not all their work could be included here, their help was vital for the exploration of many avenues that relate to paraspeckle biology. A special thank you goes to Ejona Rusha, who helped with great passion with the establishment of differentiation routes that were used to generate the paraspeckle cell atlas. Lastly and most importantly, I thank with all my heart Julia Garbaciok, who accompanied me from the beginning of my PhD thesis until the end, and in doing so had to endure many difficult times but her relentless support made it possible for me to finish this project with my sanity intact. Danke für alles!

Eidesstattliche Versicherung

Grosch, Markus

Ich erkläre hiermit an Eides statt, dass die vorliegende Dissertation mit dem Thema

The function of paraspeckle components in pluripotency maintenance and differentiation

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 2.6.2020

Ort, Datum

Markus Grosch

Unterschrift Doktorandin/Doktorand