Supplemental Information

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, Ejona Rusha, Dmitry Shaposhnikov, Michael Palo, Juliane Merl-Pham, Davide Cacchiarelli, Boris Rogelj, Stefanie M. Hauck, Christian von Mering, Alexander Meissner, Heiko Lickert, Tetsuro Hirose, Jernej Ule, and Micha Drukker

Figure S1. Methods related to Figure 1 - paraspeckle counting and subcellular localization of *NEAT1*

(**A**) Maximum projection photomicrographs of spontaneously differentiating hESCs demonstrating the analysis of *NEAT1* foci. Two single molecule (sm)FISH probe sets targeting an overlapping region in *NEAT1_1* (using Q570 and Q670 dyes) confirmed the specificity of *NEAT1_1* probes (left). *NEAT1_1* and *NEAT1_2* probes, which target non-overlapping regions of the transcript, exhibited co-localization in differentiated cells, in accordance with the definition of paraspeckles as foci of full length *NEAT1* (right). A spot detection algorithm was used to detect the foci using arbitrary settings according to (Trcek et al., 2017). Identical analysis was conducted for mESCs, differentiated progeny, and all modified cell lines in this study.

(**B,C**) Representative maximum projection photomicrographs of *NEAT1_1, _2, Neat1_1, 2* foci used for paraspeckle counting in human (**B**) and mouse ESCs (**C**) as depicted in **Fig. 1A** and **Fig. 1B**. Samples include undifferentiated, spontaneously differentiated (KSR media), BMP4- and CHIR99021-treated hESCs (left), and undifferentiated and 3 d spontaneously differentiated mESCs untreated or treated by doxycycline to induce TDP-43-eGFP expression (scale bars = 10 μm. Red: *NEAT1_1, _2, Neat1_1, _2* probes; blue: DAPI - nuclear stain).

(**D**) RT-qPCR analysis utilizing *NEAT1_1* and *NEAT1_2* assays demonstrating that the expression of full length *NEAT1* corresponds to the amount of paraspeckles (**Fig. 1A**). Samples include undifferentiated, spontaneously differentiated, BMP4- and RA-treated hESCs (n=3 independent replicates, error bars=SD).

(**E**) A Western blot analysis of predominant proteins in the cytoplasm (ACTB), nucleus (OCT4), nuclear lamina (LMNB1) and chromatin (ser5 phosphorylated RNA PolII CTD) following subcellular fractionation of undifferentiated hESCs (as outlined in Methods section).

(**F-H**) The enrichment of transcripts in subcellular fractionation-RNA-Seq of hESCs confirmed the locations of *NEAT1* isoforms, *_1* and *_2* in the nucleoplasm and chromatin fractions respectively. The nucleoplasmic enrichment of the snoRNA *SCARNA10* and of mature mRNAs *(RPS27)*, as well as the enrichment of intronic reads in the chromatin fraction, validates the separation of subnuclear components. These results are in line with previous subcellular fractionation-RNA-Seq results of mouse macrophages (Bhatt et al, 2012, Cell), demonstrating the enrichment of *Neat1* isoforms, *_1* and *_2* in the nucleoplasm and chromatin fractions respectively (**I,J**).

Figure S2. Methods related to Figure 2 - the genetic editing of *NEAT1* and *Neat1* and to global analysis of RBP-mRNA occupancy

(**A**) An illustration of a generic form of human/mouse *NEAT1* where the endogenous pA site (*NEAT1ΔpA*) and the triple helix region (*Neat1ΔTH*) were deleted using CRISPR-Cas9 (gRNA oligonucleotides and primers are outlined in Methods section).

(**B**) RT-qPCR analysis of *NEAT1* isoforms in *NEAT1ΔpA* and *Neat1ΔpA* ESC lines. Note the higher mean expression of *NEAT1_2* and *Neat1_2* in *ΔpA* clones compared to parental lines (n=2 independent replicates/clone, SD=SEM, two sided t-test; P value ** < 0.01). The smaller increase of Neat1_1 reflects a higher level of the isoform in undifferentiated mESCs compared to hESCs.

(**C**) A representative Western blotting of a conditional TDP-43 KO mESCs line (named cTdp-43 KO) harboring floxed alleles of *Tdp-43* and a CAG-ErCreEr cassette (Ling et al., 2015), and following tamoxifen-induced deletion of the gene (3 d). Beta Actin was used as loading control.

(**D,E**) SDS-PAGE analysis related to the preparation of peptides for LC-MS (**Fig. 2A**), demonstrating the migration of RBPs crosslinked to RNAs from *NEAT1ΔpA* and parental WT hESC samples. Proteins treated by oligo d(T) magnetic beads remained in gel pockets or migrated following treatment by RNase I, and whole cell lysate samples (input control) showed migration of numerous proteins.

(**F**) RT-qPCR analysis of *NEAT1_2* of samples treated by oligo d(T) magnetic beads, captured and supernatant samples derived from *NEAT1ΔpA* and parental WT hESCs (n=4, error bars = SD, two sided t-test; P value *** < 0.001). This analysis confirmed the efficient retention of *NEAT1_2* in the oligo d(T) supernatant and is refractory to oligo d(T) mRNA-RBP capture.

(**G**) Confirmatory Western blot analysis related to **Fig. 2B,C**, showing that mRNA occupancy of paraspeckle RBPs, NONO, SFPQ and of TDP-43 decreased in *NEAT1ΔpA* hESC clones compared to parental WT cells. The specific enrichment of RBPs is shown by comparing input (left) where Histone H3 signal is clearly visible, to oligo(dT) treated samples (right) where the Histone H3 signal is absent. The RBP CPEB1, which does not localize to paraspeckles, was used as a loading control.

Figure S3. Results related to Figure 4 - identification of pA sites in ESCs

(**A**) A scatter plot displaying transcripts detected by RNA-Seq, samples include undifferentiated hESCs and mesoderm progenitors generated by CHIR99021 treatment (3 d). Differentially expressed genes with an adjusted p-value of < 1e-20 (Fisher's exact test, false discovery rate 1e-15) were labeled red (n=2 biological replicates per condition). *TARDBP* (gene encoding TDP-43*), canonical* pluripotency and mesoderm genes are highlighted.

(**B**) A scatter-plot displaying degrees of lengthening and shortening of transcripts caused by changes in the location of pA sites upon the differentiation of hESCs to mesoderm progenitors, samples as in **A**. Locations of pA sites were determined by 3' mRNA-Seq (QuantSeq), and pA sites with statistically significant changes (adjusted p-value <0.001, Fisher's exact test, explained below) were colored red (n=4 biological replicates per condition).

(**C-E**) Analyses conducted using a workflow for detecting 3' UTR lengthened or shortened transcripts (direction distal or proximal, respectively) in accordance with (Rot et al., 2017). (**C,D**) The positions and the number of pA sites per gene detected in hESCs and mesoderm progenitors (samples corresponding to **B**) versus sites identified in HEK293 cells (Derti et al, 2012). (**E**) The nucleotide composition around pA sites that were uniquely identified in hESCs and mesoderm progenitors (samples as in **B**), or pA sites that overlapped with an existing dataset (Derti et al., 2012). The similar patterns support the *bona fide* classification of novel pA sites in this study.

(**F**) Representative outlines showing the frequencies of proximal and distal pA sites in transcripts encoding pluripotency factors comparing cTdp-43 KO mESCs treated or untreated by tamoxifen and primitive streak-like progenitors.

Figure S4. Results supporting the Figure 4 - role of TDP-43 in position-dependent regulation of APA in h/mESCs

(**A**) The highest ranked multivalent RNA motifs that were enriched around the pA sites exhibiting significant changes upon differentiation of hESCs or TDP-43 KD (as defined in **Fig. 4A**), or following cTdp-43 KO in mESCs (**Fig. 4B,C**). Fisher exact test and h-index (the motif coverage threshold for cluster formation) of multivalent RNA motifs were calculated as previously described (Rot et al., 2017).

(**B-D**) The positions of UGU/GUG motif enrichment plotted around enhanced and repressed pA sites, red and blue respectively, corresponding to samples as defined above.

Figure S5. Methods and results related to Figures 4 and 5 - iCLIP, TDP-43 RNA maps, and regulation of *Neat1* pA site by TDP-43

(**A**) Representative SDS-PAGE autoradiographs of ^{32}P-labelled RNAs from m/hESCs treated by UV-C for crosslinking of RNAs and RBPs, and purified using an anti-TDP-43 antibody. Treatment by high [(+++)] concentration of RNase I indicated the presence of RNAs bound to TDP-43 by location of the radioactive signal. An asterisk marks the position in the autoradiograph corresponding to the size of TDP-43 monomer.

(**B,C**) Z-scores of pentamer occurrences surrounding (−30 nt to +30 nt) all TDP-43 crosslinked sites in undifferentiated hESCs (**B**) and mESCs (**C**) as determined by iCLIP. The sequences of the two highest enriched pentamers are shown. The Pearson's correlation coefficient between the two replicates in hESCs was R=0.91 (**B**), and these pentamers were also most enriched in mESCs (**C**) by iCLIP performed using *iTDP-43-eGFP* mESCs and antibodies targeting the endogenous TDP-43 or GFP (fused to TDP-43).

(**D**) A pie chart depicting the regional distribution of TDP-43 binding sites in mRNAs identified by TDP-43 iCLIP in mESCs (samples as in **Fig. 5D,E**).

(**E-G**) "RNA maps" related to **Figs. 4** and **5** showing the relationship between the positions of TDP-43 crosslinking (**E**) or GU-rich motifs of TDP-43 bound genes (**F,G**) to TDP-43 regulated pA sites. Positive and negative correlations are shown in red and blue respectively. (**E, F**) Plots that are based on 89 and 119 pA sites that were enriched or depleted respectively in undifferentiated hESCs compared to cells treated by TDP-43 shRNAs. The black lines represent 893 pA sites that did not show a significant change, and thereby were used as a control. Pairs exhibiting highest proximal-distal site fold changes are shown (adjusted p value <0.05). (**G**) Same analysis as in (**F**), comparing undifferentiated untreated cTdp-43 KO mESCs with cells following tamoxifen treatment. Plots are based on 316 and 127 pA sites that were enriched or depleted respectively, or 1616 pA sites that did not show a significant change, represented by black lines.

(**H**) Related to **Fig. 5C:** A Western blot of TDP-43 and Histone H3, samples: undifferentiated hESCs and following 3 d of differentiation towards mesoderm (CHIR99021) and trophoblast (BMP4) progenitors respectively.

(**I**) A heat map showing the relative abundance of a selected panel of genes during reprogramming of hiF-T cells, including of pluripotency *LIN28A, LIN28B, NANOG, ZFP42* (*REX1*), epithelial *CDH1* (E-Cadherin), and mesenchymal *CDH2* (N-Cadherin) markers compared to *NEAT1* and *TDP-43* (RNA-Seq dataset from (Cacchiarelli et al., 2015)).

 (**J**) Related to **Fig. 5F,G**: qPCR analysis of *Neat1_2* transcripts in *Neat1Δ100nt* and *Neat1* WT *iTDP-43-eGFP* mESCs, treated by doxycycline or untreated (n=4 and 6 independent replicates for *Neat1ΔTDP-43* and *Neat1* WT respectively, error bars = SD, two sided t-test; P value ** < 0.01). Analysis shows that *Neat1* isoform switch depends on TDP-43 binding to the

UG-rich conserved region in *Neat1_1*.

(**K**) Quantification of gated positive cells according to IgG control (red line), samples of WT and *Neat1ΔTDP-43* mESCs spontaneously differentiated (2 d) and immunostained for NANOG (error bars = SD, two sided t-test, biological replicates n=3; P value ** < 0.001).

Figure S6. Results related to Figure 6 - the developmental competence of *Neat1ΔTH* mESCs

(**A**) The number of paraspeckles analyzed in 3d spontaneously differentiating WT parental and *Neat1ΔTH* mESC clones. Number of analyzed cells and statistical analysis as in **Fig 1B**. (**B-F**) *In vivo* analysis of the developmental potency of mESCs by aggregation with 4n embryos, which together give rise to chimeric embryos (**Fig. 6D**). Embryos consititutively expressing membrane tdTomato were used to distinguish the extraembryonic tissues (visceral endoderm and extra-embryonic ectoderm) which were derived from 4n cells, and embryonic cells that were exclusively derived from mESCs. (**B,C**) Chimeras produced with *Neat1ΔTH* mESCs, and immunostained using BRACHYURY (**B**) and FOXA2 (**C**) antibodies, respectively (5 out of 7 embryos shown). Note the abnormal anatomical characteristics around the node and axial mesoderm during gastrulation in 4n mT ↔ *Neat1ΔTH* mESC chimeras. (**D,E**) Chimeras produced using *Neat1ΔpA* (**D**) and the WT parental mESCs (**E**), and immunostained as above (2 out of 7 embryos shown in **D**). * displayed also in main figure **6E,F**. (Blue: DAPI – nuclear stain; Scale bars = 100 μm). A=anterior, P=posterior, p=proximal, d=distal.

(**F**) qPCR and RNA-seq analysis of pluripotency *SOX2*, *OCT4*, *LEFTY1*, *GDF3* early mesoderm *MESP1*, *T*, *MIXL1*, *EOMES*, *MSX2*, *MYOD*, endoderm *SOX17*, *FOXD3*, *CXCR4*, *NODAL*, *CDH2*, *CER1*, *GSC* markers in *NEAT1ΔpA* versus parental WT hESCs propagated in pluripotency conditions (Error bars: +SD; Mann-Whitney U-test, P value *** < .001, ** < .01, * < 0.05; n = 2 per clone and 3 of WT hESCs).

Fig. S1

A Neat1ΔpA

NEAT1/Neat1 — pA

tripple helix deletion

Neat1 3'end

B Relative expression of Neat1/NEAT1

hESC mESC

ΔpA WT

_1/2 _2 _1/2 _2

NEAT1 Neat1

** ** * **

D

Ladder
WT hESC sample 1
WT hESC sample 2
hESC NEAT1ΔpA clone 1
hESC NEAT1ΔpA clone 2
WT hESC sample 1
WT hESC sample 2
hESC NEAT1ΔpA clone 1
hESC NEAT1ΔpA clone2

RNAse I

input
(whole cell lysates)

oligo (dT) captured mRNA-RBP occupancy
(+RNAse)

E

WT hESC sample 1
WT hESC sample 2
hESC NEAT1ΔpA clone 1
hESC NEAT1ΔpA clone 2
Ladder

(-RNAse)

C iTDP-43 KO mESCs

TDP43
ACTb

kDa
50

30

+ tamoxi - tamoxi Marker

F fold enrichment of NEAT1_2 in the mRNA-RBP experiment of NEAT1ΔpA undifferentiated hESCs

***

oligo d(T) capture | oligo d(T) supernatant

G

WT hESC sample 1
WT hESC sample 2
hESC NEAT1ΔpA clone 1
hESC NEAT1ΔpA clone 2
WT hESC sample 1
WT hESC sample 2
hESC NEAT1ΔpA clone 1
hESC NEAT1ΔpA clone 2

TDP-43
NONO
SFPQ
CPEB1
H3

input
(whole cell lysates)

oligo (dT) captured
mRNA-RBP occupancy

Fig. S2

A

mesoderm progenitors FPKM values (log₁₀) vs hESC FPKM values (log₁₀)

Labeled genes: T, MIXL1, CDX2, FGF8, EOMES, POU5F1, ZFP42, TARDBP, SOX2, NANOG

Fisher exact test, p value; <1e-20 / 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

B

mesoderm progenitors log₂ (prox/distal pA site) vs hESC log₂ (proximal/distal pA site)

RNA shortening upon differentiation

RNA lengthening upon differentiation

Fisher exact test, p value; <0.00005 / Lorem ipsum dolor sit amet, 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

C

Distribution of 14.434 query pA-sites around 15.123 HEK293 pA sites

76.7% pA-sites from a published atlas are covered

Distance of identified pA sites from published atlas of pA sites[nt]

D

number of pA-sites vs Distance from pA-site [nt]

strong / weak / pA-less

E

nucleotide composition around 3,368 newly identified pA sites

A / U / C / G

nucleotide percentage vs Distance from pA-site [nt]

nucleotide composition around 11,056 pA-sites that overlap with known pA-sites (Derti et al., 2012)

A / U / C / G

nucleotide percentage vs Distance from pA-site [nt]

F

Tracks (absolute read counts):
undifferentiated *cTdp-43 KO* mESC
undifferentiated *cTdp-43 KO* mESC
undifferentiated *cTdp-43 WT* mESC
undifferentiated *cTdp-43 WT* mESC
PS-like progenitors (72h diff)
PS-like progenitors (72h diff)

Tbx3 — 119682685 to 119684720

Tcf7l1 — 72626400 to 7262720

Stat3 — 100889383 to 100886311

Zfp281 — 136628250 to 136630000

Nr5a2 — 136844300 to 136842600

Fig. S3

A

The most enriched multivalent motifs around the regulated APA events:

undiff hESC vs mesoderm progenitors		
Enriched sequence motif	valency	fisher
GUG	9	0.042
UCAG	5	0.042
CUCA	5	0.076
UUUU	10	0.080

undiff hESC vs *TDP-43* KD		
Enriched sequence motif	valency	fisher
UGU	16	0.0002
AUAU	11	0.0007
CCCG	4	0.015
GUGU	9	0.019

undiff mESC vs iTDP-43 KO		
Enriched sequence motif	valency	fisher
GUG	17	0.003
GUGU	13	0.008
UGUG	15	0.010
UAAC	5	0.014

B

UGU/GUG motif enrichment around significantly changed pA sites (undiff hESC vs mesoderm prog.)

C

UGU/GUG motif enrichment around significantly changed pA sites (undiff hESC vs *TDP-43* KD)

D

GUGU motif enrichment around significantly changed pA sites (undiff mESC vs iTDP-43 KO)

Fig. S4

A
mESC / hESC

B
kmer z-score, TDP-43 iCLIP hESC, rep1 vs kmer z-score, TDP-43 iCLIP hESC, rep2
UGUGU
GUGUG

C
kmer z-score, TDP-43-GFP iCLIP mESC vs kmer z-score, endogenous TDP-43 iCLIP mESC
UGUGU
GUGUG

D
iCLIP binding sites distribution
- 3UTR, protein coding
- 5UTR, protein coding
- protein coding
- intergenic region
- intron
- lincRNA
- other ncRNA
- telomeres

7.4% 3UTR
21.3% inter
46.7% intron
5.5% lincRNA

H
TDP-43
H3
undiff hESC | undiff hESC | 3 d CHIR | 3 d BMP4

E — pA-sites enhanced by TDP-43 / pA-sites silenced by TDP-43
iCLIP crosslinking around significantly changed pA sites (undiff hESC vs *TDP-43 KD*)
TDP-43 binding sites (iCLIP)

F
UGU/GUG density around significantly changed pA sites (undiff hESC vs *TDP-43 KD*)
Percent of polyA sites with GU-rich motifs

G
UGU/GUG density around significantly changed pA sites (undiff mESC vs *iTdp-43 KO*)
Percent of polyA sites with GU-rich motifs
Distance from polyA site (nt)

I
dox days: *TDP-43*, *NEAT1*, *CDH2*, *LIN28B*, *CDH1*, *NANOG*, *LIN28A*, *ZFP42*
0, 2, 5, 8, 10, 14, 18, 24, hIPS
FPKM (log₂): 8 6 4 2

J
iTDP-43-eGFP Neat1Δ100nt | *iTDP-43-eGFP WT*
Neat1_2 expression relative to *Gapdh*
-dox | +dox
**

K
NANOG fold enrichment WT cells vs Neat1Δ100nt cells
WT | *Neat1 Δ100nt*
**

Fig. S5

A

paraspeckles/diff mESC count

Neat1_1/2
Neat1_2

non modified parental mESC · Neat1ΔTH clone 1 · Neat1ΔTH clone 2

*** ***

B *Neat1ΔTH* individual embryos

membrane tomato (4n) · BRACHYURY

Neat1ΔTH Embryo 1

Neat1ΔTH Embryo 2

Embryo 3

C *Neat1ΔTH* individual embryos

membrane tomato (4n) · FOXA2

Neat1ΔTH Embryo 4

Neat1ΔTH Embryo 5

D *Neat1ΔpA* individual embryos

membrane tomato (4n) · FOXA2

LHF

Neat1ΔpA Embryo1

EHF

Neat1ΔpA Embryo 2

E parental IDG.3.2 mESC line - individual embryos

membrane tomato (4n) · BRACHYURY

EHF

WT Embryo1

membrane tomato (4n) · FOXA2

WT Embryo2

F

Expression relative to undifferentiated parental hESC

NEAT1ΔpA vs parental hESC line

RT-qPCR *NEAT1ΔpA* clone 1
RT-qPCR *NEAT1ΔpA* clone 2
RNA-seq *NEAT1ΔpA* clone 1+2

SOX2 OCT4 LEFTY1 MESP1 T MIXL1 SOX17 EOMES GDF3 FOXD3

pluripotency · mesoderm · endoderm

*** * * *** * * *** ** n.d. n.d. n.d.

Fig. S6