

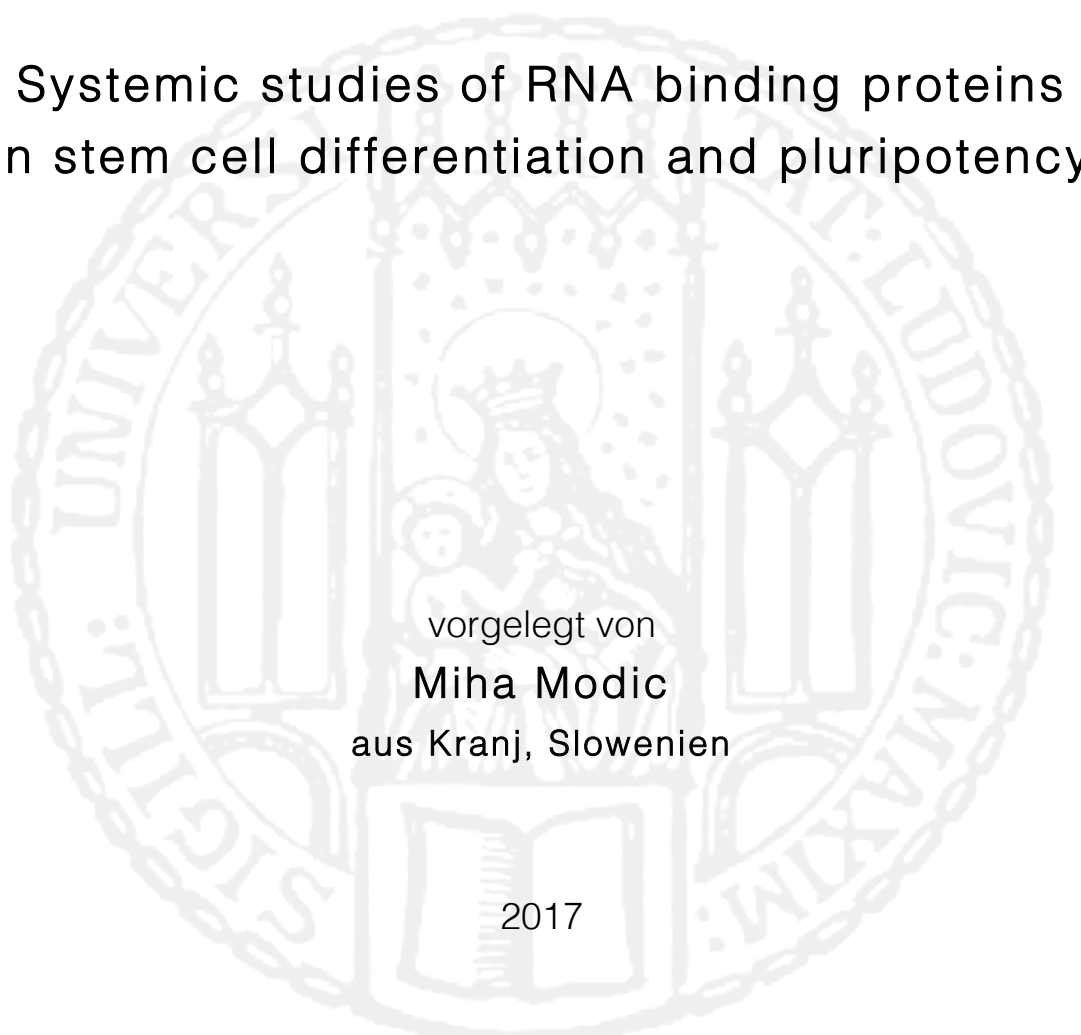
Aus dem **Helmholtz Zentrum München**
Deutsches Forschungszentrum für Gesundheit und Umwelt
Institut für Stammzellforschung

Direktorin: Prof. Dr. Magdalena Götz

Dissertation

Zum Erwerb
des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät
der **Ludwig-Maximilians-Universität München**

Systemic studies of RNA binding proteins
in stem cell differentiation and pluripotency



vorgelegt von
Miha Modic
aus Kranj, Slowenien

2017

Gedruckt mit Genehmigung der
Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Betreuerin: Prof. Dr. Magdalena Götz

Zweitgutachter: Prof. Dr. Michael Kiebler

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

Tag der mündlichen Prüfung: 13.10.2017

Eidesstattliche Versicherung:

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

“Systemic studies of RNA binding proteins in stem cell differentiation and pluripotency”

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, 11.5.2017

Unterschrift Miha Modic

Part of my doctoral thesis (chapter 3.2.) is as revised manuscript submitted to peer-reviewed journal:

Miha Modic, Gregor Rot, Markus Grosch, Dmitry Shaposhnikov, Tjasa Lepko, Ejona Rusha, Davide Cacchiarelli, Boris Rogelj, Stefanie Hauck, Alexander Meissner, Jernej Ule*, Micha Drukker*: Cross-regulation between TDP-43 and paraspeckles promotes pluripotency-differentiation transition – manuscript under evaluation (*Nature*)

Part of my thesis (chapter 3.1.) is under preparation for a publication in a peer-reviewed journal.

Miha Modic, Valter Bergant, Christopher Mulholland, Kaloyan Tsanov, Juliane Merl-Pham, Tajda Klobučar, Silvia Engert, Joel Ryan, Heiko Lickert, Stefanie Hauck, Derk ten Berge, Sebastian Bultmann, George Q. Daley, Micha Drukker: (tentative title) Developmental ontogenetic RBPome uncovers dynamic LIN28A translocation upon embryonic stem cell priming.

During my PhD studies I collaborated with others in the following scientific projects that are either published or in the final stage of preparation for publication in peer-reviewed journals:

Christopher R Sibley, Warren Emmett, Lorea Blazquez ,Ana Faro ,Nejc Haberman , Michael Briese, Daniah Tratzuni, Mina Ryten, John Hardy, **Miha Modic**, Tomaž Curk, Steve W Wilson, Vincent Plagnol, Jernej Ule: Recursive Splicing in Long Vertebrate Genes, *Nature* 521, 371–375 (2015)

Gregor Rot, Zhen Wang, Ina Huppertz, **Miha Modic**, Tina Lence, Martina Hallegger, Nejc Haberman, Tomaz Curk, Christian von Mering, Jernej Ule: High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43, *Cell Reports* 19(5), 1056–1067 (2017)

Miha Modic, Jernej Ule, Christopher R. Sibley: CLIPing the brain: Studies of protein-RNA interactions important for neurodegenerative disorders, *Mol Cell Neurosci.* 56, 429-35 (2013)

Katrin Strecker, Bettina Pitter, **Miha Modic**, Vincenzo Caprese, Denise Oroczo, Dieter Edbauer, Stefan Bonn, Christian Haass, Ulrich Pohl, Eloi Montanez, Bettina Schmid: Loss of TDP-43 leads to angiogenic defects by increasing fibronectin and integrin α 4/ β 1, submitted manuscript to journal *Cell*

Tomaz Bratkovic, **Miha Modic**, German Ortega, Micha Drukker, Boris Rogelj: Neuronal Differentiation Induces SNORD115 Expression and is Accompanied by Site-Specific RNA Editing of Serotonin Receptor 2c; submitted manuscript to *Nucleic Acid Research*

Alex Neagu et al.: The embryonic rosette displays a pluripotent state characterized by facultative pericentromeric heterochromatin, manuscript in preparation

Simona Darovic, Maja Štalekar, Youn-bok Lee, Jure Pohleven, Markus Grosch, **Miha Modic**, Marko Fonovič, Boris Turk, Micha Drukker, Christopher E Shaw, Boris Rogelj: Intracellular (GGGGCC) n RNA foci induce formation of paraspeckle-like structures, manuscript in final preparation

TABLE OF CONTENTS

Table of contents	VI
Abbreviations	IX
Abstract.....	XI
Zusammenfassung	13
1 Introduction	15
1.1 Pluripotent stem cell model	15
1.1.1 Early embryonic development	15
1.1.2 Pluripotent stem cells phase transitions.....	16
1.1.3 Comparison of naïve and primed pluripotency.....	18
1.1.4 Novel intermediate pluripotency rosette-like state in vitro	20
1.2 Regulatory networks safeguarding pluripotent stage.....	20
1.2.1 Pluripotency circuitry.....	20
1.2.2 Exit of pluripotency and early differentiation of pluripotent stem cells	21
1.3 Posttranscriptional control of pluripotency and its dissolution	23
1.3.1 RNA binding proteins in stem cell fate decisions	24
1.3.2 Pluripotency regulation by alternative splicing	25
1.3.3 Pluripotency regulation by alternative polyadenylation	26
1.3.4 miRNA regulation of pluripotency and early stem cell differentiation	27
1.3.5 Post-transcriptional regulation by noncoding RNAs	28
2 Aims of the PhD Thesis	31
3 Results.....	34
3.1 Developmental ontogenetic RBPome during embryonic stem cell priming and early differentiation.....	34
3.1.1 Recapturing early embryonic development in vitro	34
3.1.2 Identification and isolation of naïve-primed pluripotency transitioning states.....	34
3.1.3 Isolation of early primitive streak-like progenitors.....	37
3.1.4 Alternative polyadenylation changes accompanying early PSC differentiation	40
3.1.5 Analysis of the RBPome in PSC states and early differentiation.	42
3.1.6 Ontogenetic dynamic RBPome.....	48
3.1.7 Signaling-induced RBP translocation	51

3.2	Cross-regulation between TDP-43 and paraspeckles promotes pluripotency- differentiation transition	55
3.2.1	Alternative polyadenylation landscape in early PSC differentiation	55
3.2.2	TDP-43 regulated developmental alternative polyadenylation ...	59
3.2.3	Alternative polyadenylation of pluripotency circuitry mRNAs	64
3.2.4	TDP-43 governs formation of paraspeckles upon exit of pluripotency	68
3.2.5	Composition of paraspeckles	74
3.2.6	Feedback cross-regulation between TDP-43 and paraspeckles governs PSC differentiation and reprogramming	76
4	Discussion	82
4.1	Developmental ontogenetic RBPome uncovers dynamic RNA binding proteins.....	82
4.2	Signaling induced changes in RBP dynamicity and function.....	84
4.3	Identifying novel subnuclear roles of dynamic RBPs that act as stem cell differentiators	85
4.4	dissecting the post-transcriptional mechanism of lineage restriction by dynamic RBPs	88
4.5	Cross-regulation between TDP-43 and paraspeckles promotes pluripotency- differentiation transition by formation of attractor for pluripotency dissolution	89
4.6	Discovering the process of liquid phase transition in cell fate transitions	90
4.7	Role of TDP-43/NEAT1 axis in disease.....	91
5	Materials and methods.....	93
5.1	Cell culture.....	93
5.2	Human fibroblast reprogramming	94
5.3	RNA extraction and (q)RT-PCR assays	94
5.4	Single molecule FISH and paraspeckle quantification.....	95
5.5	QuantSeq T-fill 3'mRNAseq	96
5.6	Total RNA sequencing.....	96
5.7	Nascent RNA-seq	96
5.8	Western blot.....	97
5.9	Immunofluorescence	97
5.10	shRNA and siRNA knockdown	98
5.11	Generation of inducible TDP-43 over expression line	99
5.12	Flow cytometry analysis.....	99
5.13	iCLIP analysis	101

5.14	RNA-seq differential expression	101
5.15	apaExpress analysis platform.....	102
5.16	APA analysis	102
5.17	Identifying differentially polyadenylated genes	103
5.18	Visualizing position-dependent polyA site regulation with RNA-maps 103	
5.19	RBPome (RNA-RBP occupancy assay).....	104
5.20	Mass spectrometry	104
5.21	FASP digest	104
5.22	Label-free analysis of mRNA-interactome	105
5.23	MS identification of dynamic RNA binding proteins	105
6	References	110
7	Acknowledgment.....	119

ABBREVIATIONS

2i/LIF – naive pluripotency culturing conditions CHIR99021, PD0325091, LIF
3'UTR - 3' untranslated region
5'UTR - 5' untranslated region
AP - alkaline phosphatase
APA - alternative polyadenylation
CHIR – CHIR99021 is a Gsk3 β inhibitor
dsRNAs - double-stranded RNA
DTT - dithiothreitol
EpiSCs - epiblast stem cells
epiSCs - epiblast stem cells
ERK - extracellular signal-regulated kinase
ESC - embryonic stem cells
FACS - fluorescence-activated cell sorting
FASP - filter-aided sample preparation
FGF - fibroblast growth factor
GSK3b - glycogen synthase kinase 3 beta
hiPSC - human induced pluripotent stem cells
hnRNP - heterogeneous nuclear ribonucleoproteins
hPSCs - human pluripotent stem cells
iCLIP - individual-nucleotide resolution Cross-Linking and
ImmunoPrecipitation
ICM - inner cell mass
iPSC - induced pluripotent stem cells
iTDPKO - tamoxifen inducible TDP-43 knockout mESCs
IWP2 - WNT pathway inhibitor
KD - knockdown
KO - knockout
LC-MS) - liquid chromatography–mass spectrometry
LIF - leukemia inhibitory factor
limesSCs - novel intermediate pluripotency stage (*limes* [Latin]: boundary)
lncRNA - long non-coding RNA
mESCs - mouse embryonic stem cells
miR - microRNA
miRNA - microRNA
mRNA - messenger RNA

nhPSCs - naïve human pluripotent stem cells
nmPSCs - naïve mouse pluripotent stem cells
nPSCs - naïve PSCs
PAS - polyA site
PD - PD0325091 is MEK inhibitor
pre-mRNA - precursor mRNA
PS - primitive streak
PSCs - pluripotent stem cells
RBP - RNA-binding protein
RBP(ome) - mRNA bound proteome
RNA - Ribonucleic acid
RNP - Ribonucleoprotein
smFISH - single molecule fluorescent in situ hybridization
wt - wildtype

ABSTRACT

What mechanisms govern and maintain cell states during the process of differentiation is a pivotal question in science. What factors govern the commitment of developmental progenitors from pluripotent stem cells is a representative example of this question. Studies of transcriptional, signaling and chromatin regulation have been highly instrumental for elucidating mechanisms pluripotency maintenance. Nevertheless, current knowledge falls short in explaining the exit from pluripotency and its coupling to lineage commitment. It is unclear how pluripotency and differentiation become stabilized in a mutually exclusive manner.

Here, I deepen our knowledge concerning post-transcriptional mechanisms in pluripotency-differentiation transition. For this purpose I first characterize by quantitative mass spectrometry the changes that occur in the mRNA bound proteome (RBPome) and identify extensive dynamic rearrangements of the RBPome during early embryonic development, from naive to primed stem cell state and to purified primitive streak progenitors (Chapter I). In parallel I identified developmental post-transcriptional processing landscape and show that the dynamic mRNA binding of the RNA-binding protein TDP-43 is critical in pluripotent stem cells (PSCs) for the choice between self-renewal and differentiation/ pluripotency breakdown (Chapter II). In detail, I discovered that TDP-43 directly regulates an evolutionary conserved switch in alternative polyadenylation (APA) of hundreds of transcripts during early differentiation of mouse and human PSCs. Functional analysis revealed that TDP-43 integrates into pluripotency circuitry by repressing the production of lengthened transcripts of the pluripotency factor SOX2, which is targeted for degradation by miR-21. Furthermore, in pluripotent stem cells TDP-43 also promotes self-renewal by repressing the formation of paraspeckles, membraneless nuclear compartments found only in differentiated cells, by enhancing production of short isoform of the lncRNA *NEAT1*. Conversely, reduction of TDP-43 during differentiation triggers a short-to-long isoform switch of *NEAT1*, which polymerizes paraspeckles that in turn recruit TDP-43 and relocalise it away from its other RNA targets. Consistent with this cross-regulation, TDP-43 inhibits differentiation and improves somatic cell reprogramming, while paraspeckles promote early differentiation.

These findings reveal how the exit of pluripotency is regulated by a complex posttranscriptional network, which is functionally independent from lineage choices. Apart from its role in the exit of pluripotency, this cross-regulation between paraspeckles and TDP-43 has implications in cancer and neurodegeneration.

ZUSAMMENFASSUNG

Durch welche Mechanismen Zellen beim Differenzierungsprozess aus einem pluripotenten in einen spezialisierteren Zustand gelangen und wie die Zellzustände stabilisiert werden, gehört zu den Kernfragen der Stammzellforschung. Obwohl heute viele Faktoren bekannt sind, die die Differenzierung von pluripotenten Stammzellen in bestimmte Vorläuferzellen steuern, konnten die Studien, die sich der Transkription-, Signaltransduktion- und Chromatin-Regulierung widmen, Mechanismen zum Erhalt von Pluripotenz nicht völlig aufzeigen. Demgegenüber steht unser Verständnis davon, wie der konkrete Übergang aus der Pluripotenz in einen bestimmten Differenzierungsweg gelenkt wird, erst am Anfang. Wie gelingt es der Zelle aus einem stabilisierten Zustand der Pluripotenz einen wiederum stabilen Differenzierungsweg anzutreten, und wie interagieren dabei die Regulationsmechanismen der sich gegenseitig ausschließenden Zustände?

In meiner Studie analysiere ich die posttranskriptionellen Abläufe beim Übergang zwischen Pluripotenz und Differenzierung. Erst stelle ich Daten von mittels quantitativer massenspektrometrischer Verfahren charakterisierter mRNA gebundener Proteine (RBPome - mRNA bound proteome) vor in *'naïve'*, *'limes'* und *'primed'* pluripotenten Stammzellen und in aufgereinigten Vorläuferzellen des Primitivstreifens, Zelltypen der frühen Differenzierung. Diese Ergebnisse zeigen umfangreiche dynamische Umstrukturierungen des RBPomes auf. In parallel dazu durchgeführten Arbeiten habe ich posttranskriptionelle Muster aufgefunden, die zellentwicklungsabhängig auftreten. Ich kann zeigen, dass eine dynamische mRNA-Bindung des RNA-bindenden Proteins TDP-43 in pluripotenten Stammzellen besteht und für die Bestimmung zwischen den Selbsterneuerung einerseits oder Differenzierung / bzw. Wegfall der Pluripotenz andererseits kritisch ist. Im Einzelnen konnte ich neue Befunde liefern, die belegen, dass TDP-43 einen evolutionär konservierten Schalter reguliert. Dieser steuert die alternative Polyadenylierung (APA) von Hunderten von Transkripten in der frühen Phasen der Differenzierung von murinen und menschlichen PSCs.

Die funktionellen Analysen ergaben, dass man sich TDP-43 in einem Schaltkreis denken kann, der die Pluripotenz steuert. TDP-43 unterdrückt die Produktion von verlängerten Transkripten des Pluripotenzfaktors SOX2, die

sonst vermittels der Bindung von miR-21 zur Degradierung bestimmt wären. Des Weiteren wirkt TDP-43 auf den Erhalt der Fähigkeit zur Selbsterneuerung, indem es in pluripotenten Stammzellen die Bildung von *Paraspeckles* hemmt, - membranlose Strukturen im Zellkern, die nur in differenzierten Zellen gefunden werden. Dies geschieht über die Induktion einer erhöhten Produktion der kurzen Isoform der nicht-kodierenden RNA *NEAT1*. Auf der anderen Seite löst eine verminderte Expression von TDP-43, wie sie bei der Differenzierung vorgefunden wird, einen Isoform-Wechsel bei *NEAT1* aus – hin zur langen Isoform. Dadurch polymerisieren *Paraspeckles*, und sequestrieren TDP-43, wodurch die Wirkung von TDP-43 auf andere Transkripte vermindert wird. Dementsprechend hemmt TDP-43 die Differenzierung und wirkt begünstigend auf Reprogrammierung somatischer Zellen, während das Vorkommen von *Paraspeckles* frühe Phasen der Differenzierung begünstigt.

Die Erkenntnisse aus meinen Arbeiten weisen auf einen neuen postranskriptionellen Mechanismus hin, der den Ausgang der Pluripotenz funktionell unabhängig von der Entwicklungslinie reguliert, in die die Differenzierung der Zelle einmündet. Der beschriebene Mechanismus hat nicht nur Bedeutung für ein besseres Verständnis von Pluripotenz in der Stammzellforschung; vielmehr eröffnet die Erkenntnis über eine gegenseitige Regulierung von *Paraspeckles* und TDP-43 auch neue Ansätze in der Erforschung von Krebs und neurodegenerativen Erkrankungen

1 INTRODUCTION

1.1 PLURIPOTENT STEM CELL MODEL

Pluripotent stem cells (PSCs) are remarkable because they represent a basal stage of development that can be kept poised by self-renewal or differentiate upon stimuli. This enables detailed investigation of the fundamental molecular processes that govern embryogenesis. PSCs can be derived from inner cell mass (ICM) of preimplantation (blastocyst-stage) mammalian embryos or from somatic cells by forced reversion of developmental program of somatic cells, producing cell lines named embryonic and induced pluripotent stem cells (ESC/iPSC), respectively. Self-renewal is the capacity of cell to proliferate in reprogramming state and pluripotency presents a potential of single cell to differentiate to all three cell lineages of the developing embryo and adult organism.

1.1.1 Early embryonic development

Early embryogenesis is highly dynamic process and includes a cascade of transitory pluripotent cell stages along the path from blastocyst to egg cylinder development and many can be captured at different time in early embryogenesis. Historically, two forms of pluripotency have been maintained thus far *in vitro*, termed naive and primed state (Brons et al., 2007; Tesar et al., 2007; Ying et al., 2008). The most primitive state is mouse naive PSCs (nPSCs), which resemble *in vitro* “frozen in time” version of ~32 pluripotent cells embedded within ICM of pre-implantation embryo (E3.5, Fig. 1). These pluripotent cells are *in vivo* shielded by trophectoderm cells, together composing blastocyst (Rossant, 2008). Under proper culture conditions with addition of selected inhibitors, ICM derived cells are *in vitro* “locked” in the reprogramming state (Nichols and Smith, 2009; Ying et al., 2008). One day later in development (E4.5) upon fertilization when the blastocyst matures, the mouse embryo reaches the uterine wall and implants into uterus (Dey 2004, Bedzhov 2014). At this stage this early state of naive pluripotency evolves into the primed pluripotent state of the egg cylinder epiblast resembling rosette formation of pluripotent cells centered in monolayer surrounding the cavity, the precursor of the egg cylinder. At this

stage peri-implantation rosette becomes responsive to the differentiation-inducing signals of the gastrula (Bedzhov and Zernicka-Goetz, 2014).

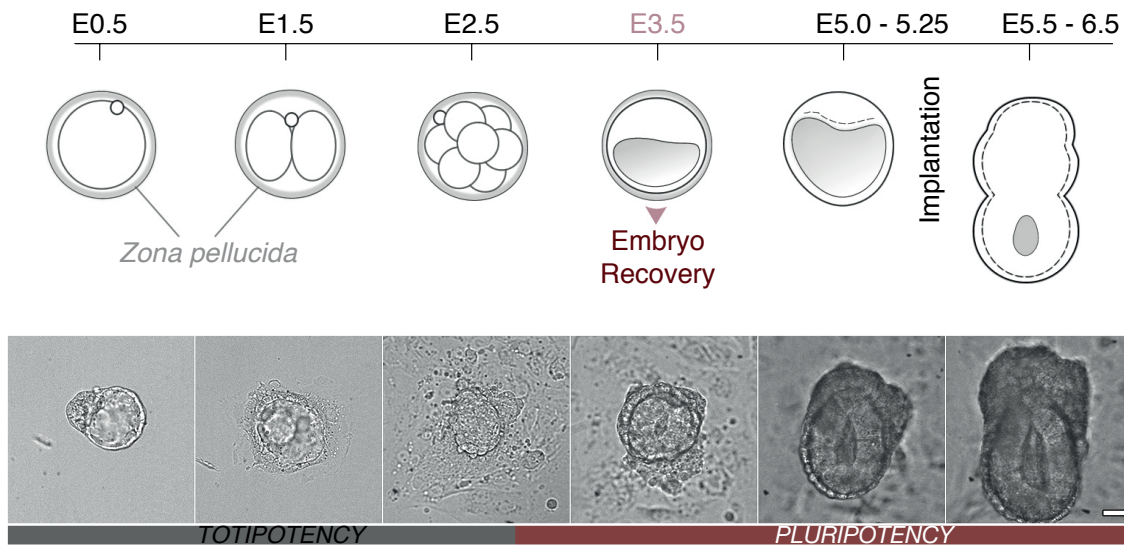


Fig 1. Overview of the early embryonic development.

Following implantation, the embryo is composed of polarized cup shaped epiblast (pluripotent derivative of ICM) and the trophoboderm-derived extra-embryonic ectoderm. This more committed “post-implantation” pluripotency state has been coined “primed” and retains the potential to develop into the three germ layers, but lack the reprogramming capacity and are unable to integrate into a preimplantation blastocyst. In developing embryo (E5.5), epiblast stem cells are restricted sub-population that is shielded from lineage specification (driven by Erk pathway activated by FGF) and retain upregulated Nanog (Lanner and Rossant, 2010; Nichols and Smith, 2009).

1.1.2 Pluripotent stem cells phase transitions

In vitro, primed pluripotent cell lines can be derived either from mouse epiblast stage embryos (Brons et al., 2007; Tesar et al., 2007) or by committing nPSCs to epiblast stem cells (EpiSCs) by treating them with FGF and Activin A (Joo et al., 2014; Kurek et al., 2015) (Fig. 2). Thus, nPSCs and pPSCs are the in vitro counterpart of ICM, rosette stage, and the epiblast, respectively.

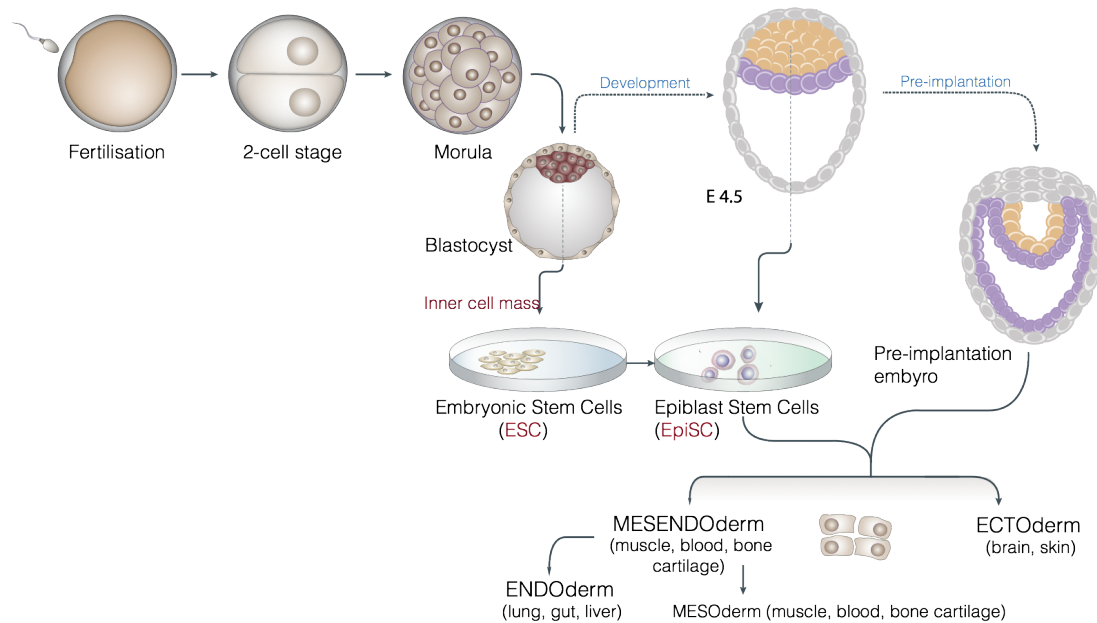


Figure 2: *In vitro* generation of naïve and primed PSCs and their differentiation to germ layer precursors.

Importantly human PSCs are thought to be equivalent of mouse primed PSCs, sharing distinct epigenetic landscape and increased DNA methylation, X chromosome activation and prevalent use of the proximal enhancer element to maintain OCT4 expression (Choi et al., 2016; Gafni et al., 2013; Theunissen et al., 2014). It seems that derivation of naïve hPSCs (nhPSCs) or conversion of hPSCs to an earlier naïve-like stage is possible, but currently there exist a dispute over their classification (Dodsworth et al., 2015). Improving the derivation and classification of hnPSCs is predicted to advance fundamental understanding of human development and aspects of cell differentiation, passaging, and low gene editing efficiency, as well as prospects for developing regenerative therapies using hPSC progeny. Understanding growth conditions of derived pluripotent lines is pivotal to model developmental progression as well as the utility of stem cells for research and regenerative. To date five publications reported conversion of hESCs to naïve state without the use of transgenes (Gafni et al., 2013; Huang et al., 2014; Theunissen et al., 2014; Wang et al., 2014). All studies rely on a protocol that closely relates to mouse naïve stem cell culturing condition; combination hLIF addition with simultaneous dual inhibition of MEK and GSK3b (Dodsworth et al., 2015). In addition human naïve PSCs require administration of competence factor FGF to block differentiation, providing evidence of distinct signaling requirements to safeguard pluripotency in naïve mouse and human ESCs.

Property	Naïve or Ground state	Limes	Primed state or xEpiSC
Embryonic tissue	Early epiblast/blastocyst	Embryonic rosette	Egg cylinder
Chimeras	Yes	Yes	No
Naïve markers	Rex1, Nanog, Klf4, Nr0b1	Rex1, Nanog, Klf4	Absent
Specification markers	absent	Otx2	Fgf5, Otx2, Oct6
XX status	XaXa	XaXa	XaXi
Response to 2i	Self renewal	Reversible state	Cell death
Response to Fgf5/Erk	Heterogeneous differentiation	Rapid differentiation	Self-renewal/priming

Table 1: Comparison of Naïve, Limes and Primed Pluripotent States

1.1.3 Comparison of naïve and primed pluripotency

While both, naïve and primed pluripotency cell stages are defined as pluripotent, several key features have been found to differ in naïve from primed PSCs, and have become criteria for designating the respective states (Table 1). nPSCs share some of epigenetic features of blastocyst with presence of two active X chromosomes in female cells, high chromatin condensation and lack of bivalent domains. These are activating and repressing epigenetic modifications at the same genomic position and are signature of primed ESCs with “alternative pluripotency configurations” (Weinberger et al., 2016).

However, differences do exist between naïve and primed state with relation to the core pluripotency network. Primed PSCs maintain expression of Sox2 and Oct4 pluripotency transcription factors, but already lose “naïve factors” (Esrrb, Nanog, Klf2/4). In addition, activation amplitudes by distinct signaling pathways prompt the cells towards stem cell stage progression. FGF/MAPK-ERK signaling is particularly potent in promoting naïve to primed state transition by inducing differentiation genes and inhibiting Nanog, and thus this pathway is kept inhibited for maintaining the naïve state by MEK

inhibitors (Lanner and Rossant, 2010). Similarly, the Nodal/Activin-TGF β pathway (through Smad2/3) drives naive to primed transition and subsequently promotes expression of differentiation genes. Conversely, Wnt/ β -catenin pathway is held constitutively active in naïve ESCs by inhibiting GSK3 β , in order to suppress of the pluripotency inhibitor Tcf3 (Cole et al., 2008), while in the primed state this pathway promotes differentiation in the direction of the primitive streak (Berge et al., 2011; Kim et al., 2013; Kurek et al., 2015).

Historically, ES cells were derived on mitotically inactivated fibroblasts, in the presence of serum and LIF (Martin, 1981), which activates JAK-stat3 pathway (Smith et al., 1988), thereby enhancing expression of naïve factor Klf4 (Hackett and Surani, 2014). In conjunction, these pathway inhibitors can substitute serum for propagation of naïve PSCs, a recipe commonly known as 2i/LIF (Ying et al., 2008). On the hand, *in vitro* conversion of naive to primed state pluripotency involves exclusion of 2i (MEK and GSK3 β) inhibitors, and addition of FGF together with Nodal/Activin (with or without LIF) (Guo et al., 2009). This is therefore taken as an indication that distinct signaling inputs safeguard pluripotency in nmPSCs versus hpPSCs (Fig. 3).

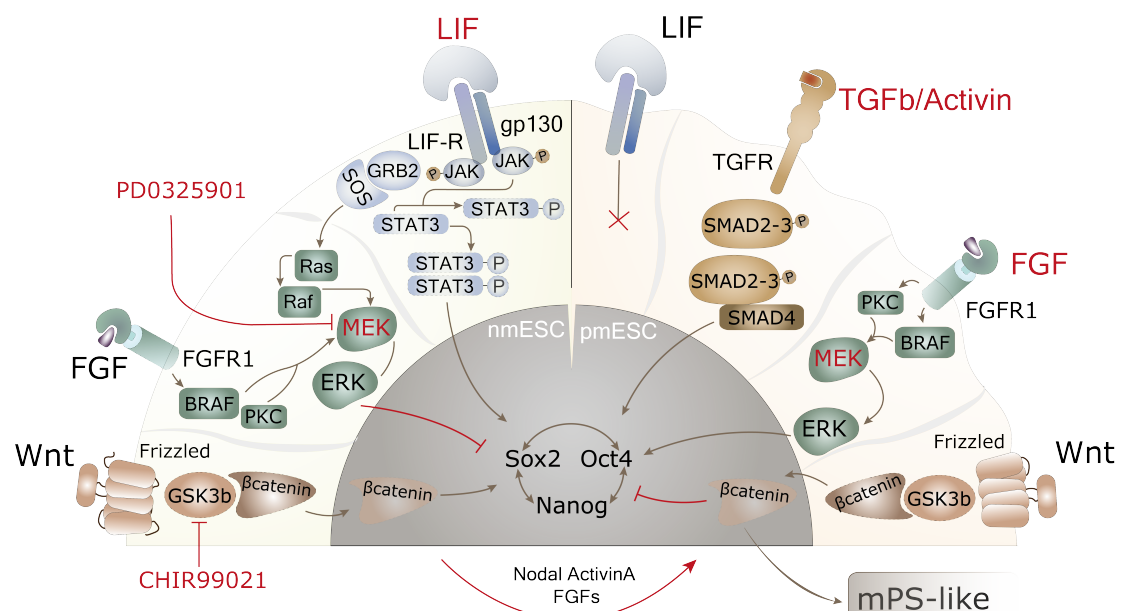


Fig3: Signaling pathways mediating mouse naive and primed state self-renewal and differentiation.

1.1.4 Novel intermediate pluripotency rosette-like state in vitro

Single-cell transcriptome sequencing is an emerging technology that has been applied study the continuum process of intermediate pluripotent states transitioning ranging between naïve and primed extremes (Hough et al., 2014; Kolodziejczyk et al., 2015; Petropoulos et al., 2016). Accompanied with revision of transition of early embryonic development from symmetric hollowed sphere (ICM) to a polarized cup, Bedzhov and colleagues identified that embryos at peri-implantation stage form rosette-like structure (E4.75–E5.0) organized around small luminal space (Bedzhov and Zernicka-Goetz, 2014; Bedzhov et al., 2014). It is worth mentioning that collaborative study managed to establish rosette-like pluripotent cells in vitro and indefinitely maintain by inhibition of Wnt/ β -catenin signaling in conjunction to active MEK signaling (Nagau et al., manuscript in preparation; Table 1).

Taken together, interplay of Wnt and MEK signals dictate distinct pluripotency stages that are accompanied with extensive epigenetic (Galonska et al., 2015) and post-transcriptional changes (Results; chapter I and II). We used this model to investigate how signaling pathways feed into remodeling of RBP functions during the transitions from naïve over peri-implantation to primed pluripotent stem cells functionally contribute to self-renewal and commitment to early primitive streak progenitors, and how changes of signaling pathways functionally regulate the post-transcriptome during pluripotency cell state transitions.

1.2 REGULATORY NETWORKS SAFEGUARDING PLURIPOTENT STAGE

1.2.1 Pluripotency circuitry

Understanding the mechanisms underlying self-renewal and pluripotency of stem cells has been the focus of numerous studies in recent years (Gonzales and Ng, 2016; De Los Angeles et al., 2015), particularly from the perspective of the transcriptional regulation of the pluripotency network governed by trinity of nuclear regulators, Oct4, Sox2, and Nanog (Ding et al., 2009; Masui et al., 2007; Mitsui et al., 2003; Pei, 2009; Vallier et al., 2009). Overexpression of later overcomes the LIF-STAT pathway requirement to safeguard pluripotency (Mitsui et al., 2003). Despite Nanog being essential for life and establishment of pluripotency in vivo, it is

dispensable during *in vitro* somatic cell reprogramming to iPS cells (Festuccia et al., 2012; Smith et al., 2016) and can be replaced by forced expression of Esrrb (Festuccia et al., 2012; Martello et al., 2012a, 2012b), its coactivator Ncoa3 (Percharde et al., 2012) or myriad of functional downstream factors that can replace core pluripotency network elements in maintenance (Schmidt and Plath, 2012) or establishment of pluripotency (Smith et al., 2016). Nanog-, Oct4-, or Sox2 deficient ES cells lose pluripotency and differentiate into extraembryonic lineage (Masui et al., 2007; Mitsui et al., 2003; Nichols et al., 1998), suggesting that the pluripotency trinity is major regulator of the self-renewing state. Depletion of some factors that can compensate core pluripotency factors results in failed colony forming capability (Martello et al., 2012a; Yeo and Ng, 2013). Current understanding of molecular, transcriptional and functional properties of naïve ESCs makes them the best available approach to model the developmental “ground state” of pluripotency, very close to their *in vivo* counterparts (Hackett and Surani, 2014).

1.2.2 Exit of pluripotency and early differentiation of pluripotent stem cells

Stem cell fate is specified by a mutual interplay between self-renewal specific differentiation signals. To initiate differentiation, stem cells have to be forced out of self-renewal. Unearthing the molecular mechanisms that dissolve the program of self-renewal in pluripotent stem cells is crucial for understanding the fundamental transition from pluripotency to germ layer precursors (Fig. 2) that have the potential to give rise to functional terminally differentiated cells of our body for regenerative purposes. The transcriptional and epigenetic control of the pluripotent state has been elaborated in detail, and significant headway has also been made in understanding germ layer differentiation. Namely, every pluripotent stem cells can be induced into endoderm, ectoderm and mesoderm, recapitulating the early events in embryogenesis. The process of differentiation can be conceptually divided into two processes: exit from pluripotency and germ layer differentiation/lineage specification (Gonzales et al. 2015). Current knowledge does not explain well the process of exit from pluripotency, and how the initial non-pluripotent stage is stabilized before epigenetic modifications generate the differentiation tracks.

In this regard, high-throughput screening has over the recent decade identified over hundred of genes, primarily transcription factors, involved in the process of differentiation. This has been highly instrumental to decipher various functional pluripotency networks substituting OCT4-NANOG-SOX2 cluster (Ding et al., 2009; Ivanova et al., 2006; Kagey et al., 2010) however, whether and how these factors initiate differentiation has largely remained unclarified. The process of differentiation can be conceptually subdivided into two major steps (Gonzales et al., 2015) composed of the process of exit from pluripotency and germ layer differentiation/lineage specification. Current knowledge does not explain well the process of exit from pluripotency, and how the initial non-pluripotent stage is stabilized before epigenetic tracks are set by modifications of the epigenome. Initial insights were unraveled by high-throughput screening approach, identifying dual specificity phosphatases in regulating exit of pluripotency via ERK signaling leading to nuclear clearance of TFE3 transcription factor (Betschinger et al., 2013) (Fig. 4).



Figure 4: Positive (blue) and negative (orange) regulators of naive pluripotency. Modified from Kalkan and Smith, 2014

1.3 POSTTRANSCRIPTIONAL CONTROL OF PLURIPOTENCY AND ITS DISSOLUTION

In the background, major improvements of deep sequencing techniques and proteomics enabled studies hinting the existence of another complex layer of gene expression regulation on RNA level that can modulate pluripotency and lineage commitment. Our understanding of eukaryotic gene expression has been reshaped by considering regulation on the RNA level (posttranscriptional gene regulation) as a central point in gene expression (Licatalosi et al., 2009) roles in a wide variety of biological processes (Gerstberger et al., 2013; Mitchell and Parker, 2014). In particular, system-wide studies showed that posttranscriptional regulation is critical for early development. Already in the earliest developmental stage, the zygotic genome is transcriptionally silent and development is guided by exclusively by regulation by maternally produced mRNAs and proteins that are loaded into the oocyte (Lasko, 2011). Post-transcriptional gene regulation by RBPs is the fastest and very effective way to adapt gene expression and change cellular responses in comparison to transcriptional regulation or signaling.

The biological importance of RBPs is underscored by the fact that many human diseases result from RBP malfunction (Castello et al., 2013). RBPs determine RNA processing of pre-mRNA to mature mRNA that involves attaching a guanosine cap to the 5' end of the nascent transcript, removal of introns and cleavage of transcripts at the 3' end followed by subsequent addition of multiple adenosine residues during polyadenylation. New experimental approaches based on high-throughput sequencing are challenging our insights into regulatory principles controlling gene expression and thereby cell fate with an effort to understand the functions of RBPs. Single RBP can regulate hundreds on mRNAs and in turn whole developmental pathways can be modulated.

Numerous studies have recently clarified that gene expression in eukaryotes involves tight regulation on the RNA level ensuring that correct complement of mRNAs and protein quantities are produced in cells (Presnyak et al., 2015). Throughout the RNA maturation from transcription to translation, RNA is namely associated with a myriad of RNA-binding proteins. These RBPs play a critical role in many aspects of gene expression and different RBPs regulate the unique fate of related transcripts in a cooperative and combinatorial fashion. This staggering variety of posttranscriptional

regulation modes mediated by combination of RBPs is thought to determine the cohort of posttranscriptional modifications of RNAs (Fu et al., 2015), thereby dictating cell fate. RBPs even permit reversal of cell-fate decisions, allowing replenishment of stem cells by cells that have already committed to differentiate (Wright and Ciosk, 2013).

1.3.1 RNA binding proteins in stem cell fate decisions

It is thus likely that another mode of regulation plays a central role in regulating the process of exit from pluripotency. It was already demonstrated that post-transcriptional regulation accounts for nearly 75% of the changes in protein levels after differentiation induced by knockdown of the transcription factor Nanog (Ye and Blelloch, 2014). Furthermore, recent screening approaches (Leeb et al., 2014; Lu et al., 2013; You et al., 2015) identified RNA-binding proteins (RBPs) as modifiers of pluripotency (Ye and Blelloch, 2014) indicating that RBPs are important for PSC differentiation. This became apparent in the context of PSCs and differentiation primarily through proteomics analysis of OCT4, SOX2, and NANOG interaction partners (Mallanna et al., 2010) (Fig. 5). This pluripotency interactome revealed that one third pluripotency trinity interaction partners are candidates or validated RBPs, and past studies have already confirmed that majority of proteins that interact with pluripotency trinity are essential regulators of self renewal. To study the posttranscriptional control of pluripotency dissolution, the studies from the Smith lab have been highly instrumental, discovering RBP PUM1 that negatively regulate mRNA isoforms that are specific to nPSCs when the cells begin to differentiate (Leeb et al., 2014). Also some other proteins (ESRP1, TRIM71 (Chang et al., 2012)) were hinted to fine-tune expression level of pluripotency factors by binding to their mRNAs and very likely control their turnover by 5'UTR binding (Fagoonee et al., 2013). This indicates posttranscriptional regulation of mRNAs is fundamental for rapid dissolution of the naïve apparatus of nPSCs.

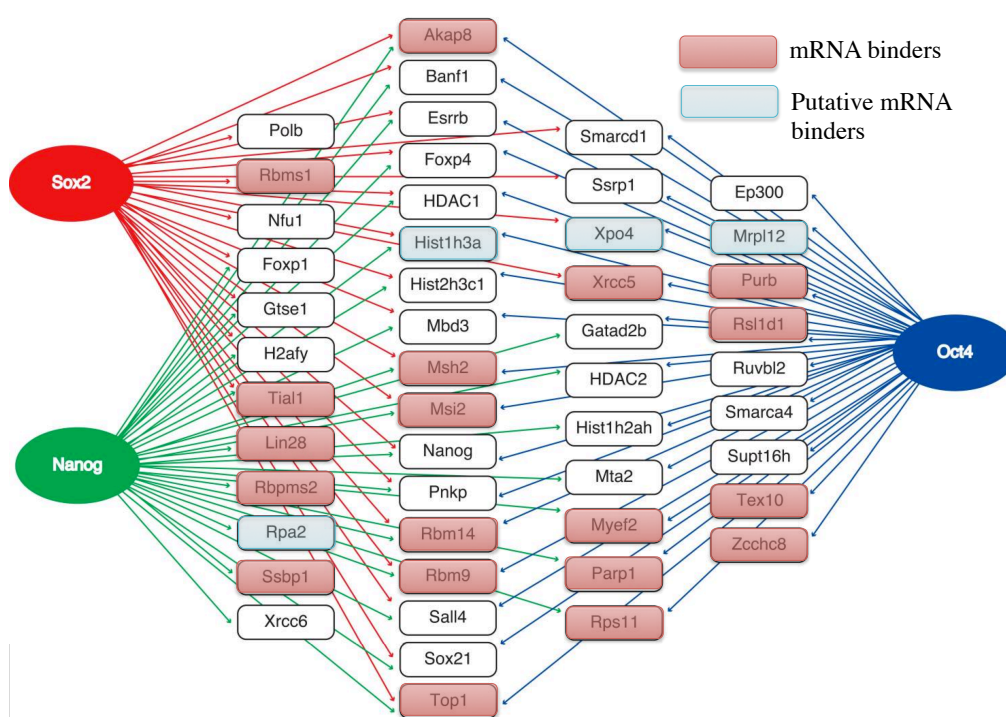


Figure 5: RNA binding factors that physically interact with core pluripotency transcription factors (OCT4, SOX2, and NANOG) are labeled in orange and putative RBPs are labeled in blue. Modified upon Mallanna et al., 2010.

1.3.2 Pluripotency regulation by alternative splicing

Alternative splicing results from variant recognition of the splicing signals, which can lead to skipping of alternative exons. Since most alternative exons are located within the coding region, alternative splicing is widely used mechanism for increasing coding diversity. Alternative pre-mRNA processing plays a central role in defining cell state and tissue specificity, since most human genes are processed in a cell type and tissue-specific manner (Licatalosi and Darnell, 2010). Recently numerous findings are unearthing distinct post-transcriptional landscape required for maintenance of pluripotent nature of PSCs. f.i. fundamental differences in patterns of alternative exon usage were observed between ES cells and differentiated cell types and tissues (Han et al., 2013). Also important pluripotency factors (Nanog, Sall4, Tcf3) encode various isoforms that are not all essential for maintaining pluripotency (Ye and Blelloch, 2014). Already skipping of stem cell specific exon of Foxp1 is sufficient to drive stem cell differentiation (Gabut et al., 2011). Blencowe and Ng labs showed that MBNL and SON proteins in general repress the generation of a cassette of gene

isoforms characteristic of human PSCs (Lu et al., 2013) and in detail repress ES-cell-switch exon of *Foxp1* (Han et al., 2013) (Fig. 6).

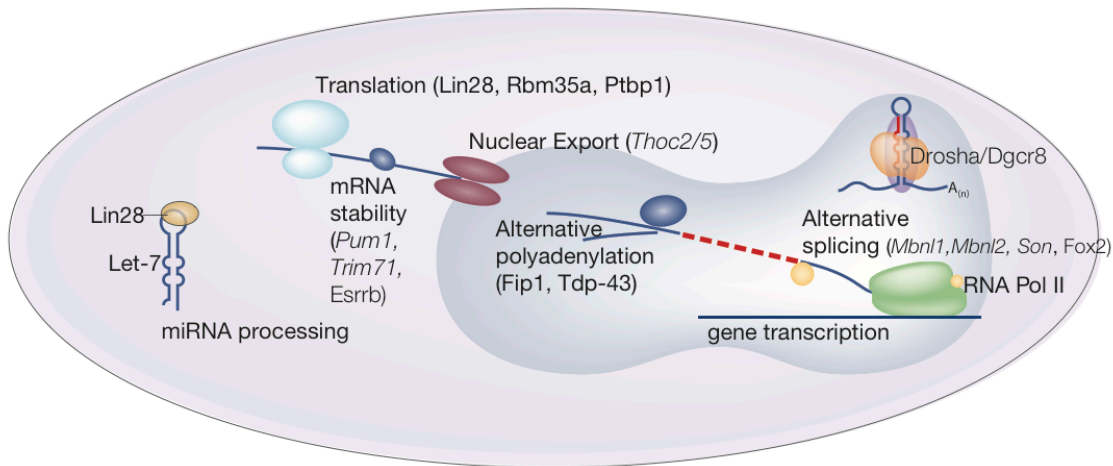


Figure 6: RBPs involved in pluripotency act on many regulatory steps. Negative pluripotency regulators are depicted in italic.

1.3.3 Pluripotency regulation by alternative polyadenylation

Alternative RNA cleavage site usage, better known as alternative polyadenylation (APA) generally regulates the length of 3'-UTRs. According to latest estimates, 70% of human genes undergo APA resulting in different 3'ends with different length of 3'-UTR (Derti et al., 2012) and such extensive regulation is enabling tissue specific expression of ubiquitously transcribed genes (Lianoglou et al., 2013). On top of this, APA lead to altered translational efficiency (Bava et al., 2013) or affect the mRNA stability and localization (Danckwardt et al., 2008) without changing the mRNA coding potential. Although isoforms with different length of 3'UTRs are producing the same protein, APA can even differentially determine the localization of proteins with different functions dependent 3'UTR usage (Berkovits and Mayr, 2015). The length of 3'-UTRs has important biological functions as it also in this way increases coding diversity and has implications in cell metabolism and disease development (Mayr and Bartel, 2010). Since 3'-UTRs harbor miRNA-binding sites, shortening of 3' UTR may lead to the loss of such regulatory elements making these short isoforms less prone to mRNA degradation and translate more efficiently. This can cause an increase in oncogene expression via APA. Recent studies support that cancer cell lines often express substantial amounts of mRNA isoforms with shortened 3'-

UTRs. Such phenomena can also be observed in other proliferating cells such as T lymphocytes (Sandberg et al., 2008) and IPS cells in the process of reprogramming of somatic cells (Ji et al., 2009). Therefore, the proliferation and differentiation status of a cell seem to be tightly associated with its APA profile. Lately, more studies are presenting how APA plays a crucial role in many biological functions. The key regulators, extend and biological roles of APA during stem cell differentiation are to date unclear and it is generally unknown what controls exit of pluripotency and in particular what mechanisms are employed by the relevant RBPs on a transcriptomic level to regulate pluripotency breakdown.

Multiple RBPs function in concert to dictate alternative polyA selection. We have uncovered principles how TDP-43 is committed in this process. TDP-43 protein is involved in several layers of mRNA life cycle, both in nucleus and cytoplasm. In nucleus they play a role in regulating alternative splicing (Lukavsky et al., 2013; Tollervey et al., 2011a) and as we show, it is essential for alternative 3'end processing as well (Eréndira Avendaño-Vázquez et al., 2012). Apart from splicing, TDP-43 acts in several different regulatory steps; TDP-43 was originally identified as transcriptional inhibitor of assembly of general transcription factors on the TATA long terminal repeat of the HIV-1 genome, thereby inhibiting transcription of the TAR DNA element (Ignatius et al., 1995). Binding of TDP-43 regulates mRNA stability as well. That was presented by TDP-43 binding to 3'UTR of human low molecular weight neurofilament (hNFL) promotes stability of these transcripts. Recently it was reported that by regulating its own protein levels, TDP-43 might directly or indirectly affect TDP-43 transcription (Eréndira Avendaño-Vázquez et al., 2012).

1.3.4 miRNA regulation of pluripotency and early stem cell differentiation

There are mounting evidence that specific miRNAs play crucial roles in regulating myriad of cellular processes, including stem cell differentiation. Loss of Dicer, central processor of dsRNAs into mature miRNAs, causes severe defect in stem cell differentiation (Kanellopoulou et al., 2005). DGCR8, an RNA-binding protein that assists the RNase III enzyme Drosha in the miRNA biogenesis, results in silencing of mESCs self renewal (Wang et

al., 2007). Introduction of let-7 miRNA into Dgrc8 KO but not into wt mESCs rescues differentiation potential and suppress self-renewal. This indicates that precise control of let-7 is essential in facilitating stem cell differentiation. While primary transcripts of *let-7* (*pri-let-7*) family is unchanged through development; rather, a post-transcriptional regulatory mechanisms prevent accumulation of the mature miRNAs in undifferentiated stem cells and pluripotent embryo. Namely *pri-let7* family is uniformly expressed during embryonic development, however mature *let-7* miRNAs is found only in differentiated cells (Heo et al., 2008; Suh et al., 2004), indicating a regulation on a post-transcriptional level. Indeed, LIN28 proteins were identified to control processing of let-7 family on several layers of miRNA biogenesis. Primarily LIN28 proteins were thought to regulate let-7 processing by binding to conserved stem loop motif (Wilbert et al., 2012) to recruit TUTase4 (Heo et al., 2008) and in concert mediate 3' terminal uridylation of pre-let7. Such modified precursor transcript is refractory to processing by Dicer and is in turn degraded by exonuclease Dis3l2 (Chang et al., 2013). Additionally, by unknown nuclear mechanism LIN28 may inhibit Drosha processing, which is necessary to cleave the primary transcript into the pre-miRNA hairpin (Newman et al., 2008). This sheds some light on possible let7-dependent nuclear role of LIN28A, confirmed by recent identification of novel LIN28A binding sites to *pri-let7* (Stefani et al., 2015). Independent of its role in let-7 biogenesis, LIN28A promotes translation of pluripotency promoting factors (especially OCT4)(Cho et al., 2012; Wilbert et al., 2012). Similar function has been observed for interaction partner of LIN28 (L1TD1) suggesting it acts in complex with LIN28, thereby facilitating binding to mRNA to influence levels of OCT4 (Emani et al., 2015).

1.3.5 Post-transcriptional regulation by noncoding RNAs

It is becoming evident that a large group of RNAs, arbitrarily referred to as long non-coding RNAs (lncRNAs), are functionally diverse, intimately involved in regulating transcription and in posttranscriptional processes. To name just few examples, lncRNAs have been shown to interact and potentially regulate and orchestrate the activity of proteins including, the Polycomb repressive complex 2 (PRC2) that modulate chromatin structure and accessibility for RNA transcription machinery (Mondal et al., 2015). lncRNAs have been shown to influence splicing decisions either directly by

forming RNA-RNA duplexes or by direct regulation of splicing factors (Gonzalez et al., 2015). In addition lncRNAs were found involved in regulating post-transcriptional RNA modifications through a diverse set of interactions ranging from acting as microRNA sponges to serving as scaffolds for higher-order complexes, such as ribonucleoprotein (RNP) particles (Du et al., 2016). Notably, lncRNAs have been indirectly linked to both gene activation and repression through the organization of nuclear subdomains; *NEAT1*, for example, have been shown to associate with interchromatin granules implicated in post- and -transcriptional regulation (Hirose et al., 2013). lncRNAs therefore can regulate transcription and transcripts through several mechanisms, and it has been postulated that our knowledge of the transcriptional regulatory mechanisms discovered thus far might just be the “tip of the iceberg, with additional means of lncRNA-mediated transcriptional regulation to be uncovered” (Geisler and Coller, 2012, Amaral et al., 2008). Considerable knowledge has been gained about *NEAT1*, crucial scaffold component of subnuclear structures termed paraspeckles (Fox et al., 2002) Two *NEAT1* isoforms exist, short isoform – *NEAT1v1*, (MEN β ; 3 kb) and long isoform – *NEAT1v2* (MEN ϵ ; 23 kb), both sharing the same 5' region, but only long isoform can polymerize paraspeckles (Clemson et al. 2009, Sasaki et al. 2009). While the middle part of *NEAT1v2* localizes to paraspeckle core, its 3' and 5' ends as well as *NEAT1v1* transcripts are found at the periphery (West et al., 2016). Paraspeckles were initially regarded as non-crucial nuclear domains (Nakagawa and Hirose, 2012) due to the generation of mouse *NEAT1* knockout that seemed normal (Nakagawa et al., 2011), but thus view is now changing with the studies implicating paraspeckles in female sterility and lactational defects (Nakagawa et al., 2014). Also, the importance of paraspeckles may differ between mouse and human, also since the latter contains *Alu* inverted repeat sequences, which are present only in higher primates and were shown to serve as paraspeckle retention signal (Chen et al., 2008). The link between paraspeckle functions and regulation of developmental processes has not been fully established yet, but it is enticing to speculate that at least for the case of exit from pluripotency there are strong indications for paraspeckles' central role as paraspeckles are absent in PSCs (both in human and mouse), but form rapidly upon differentiation, and are present in all non-pluripotent cell types and in all mammals tested to date.

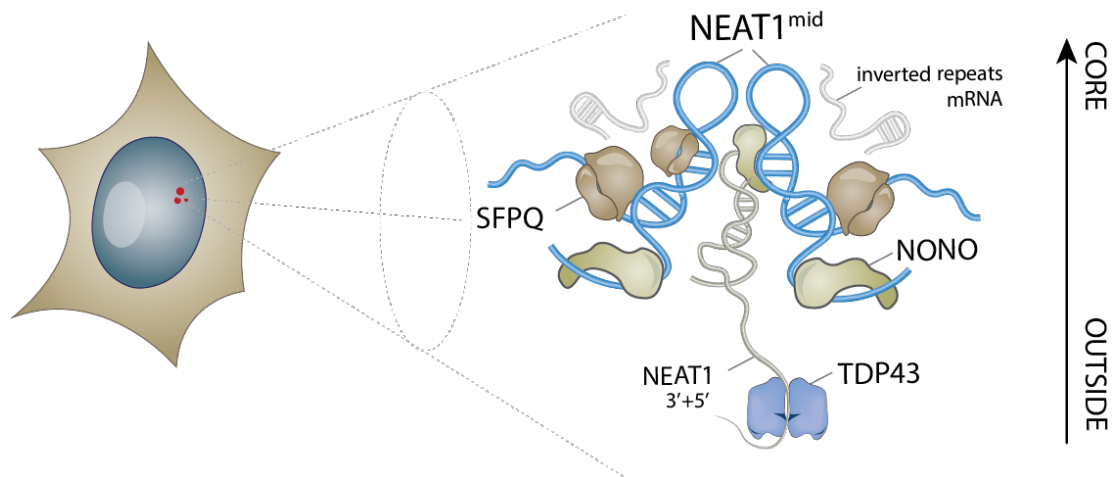


Figure 7: Paraspeckles are subnuclear foci built upon transcription of long isoform of lncRNA Neat1 containing myriad of RNA binding proteins, including TDP-43.

Majority of the proteins confirmed to localize to paraspeckles are involved in splicing and other post-transcriptional processes that are fundamentally important for control of early differentiation, even the paraspeckle core components (Ma et al., 2016). This proposes an existence of a high-order orchestration of posttranscriptional processes involving multiple RBPs. It is enticing to speculate that in this regard paraspeckles are particularly important and function as global hubs for post-transcriptional regulation. The fact that signaling pathways have tight control over early developmental stages, as demonstrated in great detail for the mouse ESC system (Fig. 3) raises the intriguing question of whether the same biochemical cascades directly govern stem cell differentiation. How changes of the cohort of RBPs functionally regulate the co- and post-transcriptome during pluripotency cell state transitions remains to be elucidated in detail, however paraspeckles mediated post-transcriptional regulation indicates existence of post-transcriptional hubs.

2 AIMS OF THE PHD THESIS

The primary objective of my thesis is to identify mechanisms pertinent to posttranscriptional regulation that govern exit of pluripotency. The understanding of PSC biology is fundamentally important because these cells provide us a proxy system to address key questions regarding early human development, and from the medical perspective, we can use knowledge to produce progeny for regenerative therapies and for modeling disease in the dish. My approach relies on global characterization of posttranscriptional changes that take place upon commitment of PSCs and their early differentiation, followed by detailed mechanistic studies. The importance of posttranscriptional regulation is evident because the copy numbers of transcripts can only partially explain the quantity of the respective proteins.

How are RBPs involved in rapid dissolution of the self-renewing reprogramming apparatus in PSCs is to date unknown. To deepen our knowledge concerning post-transcriptional mechanisms governing pluripotency transitions and breakdown of the pluripotency circuitry, I first aimed to identify global changes in alternative polyadenylation (APA) during stem cell differentiation. Upon discovering differentiation-induced pre-mRNA processing changes, I took a candidate approach to identify RBPs that maintains pluripotency or such that dissolve this state by promoting specific post-transcriptional landscapes. To gain direct mechanistic knowledge about respective RBPs, I went on to analyze what *cis*-acting elements in the UTRs that recruit RBPs, and how this regulates gene expression. Finally, I conducted functional 3'end processing analyses of RBPs to prove their link to developmental states. The specific aims of my study therefore include:

1. To globally identify the alternative polyA-sites and what 3' end processing rearrangements take place upon change of cell fate. In this context the choice of PSCs is highly advantageous from the perspective that discrete developmental stages can be propagated and purified *in vitro*, termed naive and primed states, and purified developmental progenitors that exhibit mesoderm and endoderm characteristics.

2. To identify clustered enrichment of regulatory motifs around developmentally regulated polyA sites. This premise is guiding me because RBPs typically bind to such clustered motifs and currently a compendium of 114 human/mouse RBPs has experimentally defined binding motifs¹.
3. To identify key regulatory RBPs involved in the dismantling of pluripotency and/or promotion of lineage commitment. This can be done using information about predicted RBPs according to their known binding motifs, and by integrating mRNA-sequencing data with experimentally defined RBP binding sites by iCLIP.
4. To decipher the biological functions of APA and influence of RBPs on pluripotency and/or differentiation, i.e. how are RBPs promote and dissolve the self-renewing reprogramming state.
5. To further empower identification of developmentally essential RBPs I aimed to identify changes that occur in the RNA binding activity of compendium of RNA binding proteins during subsequent steps of differentiation, from naive to primed stem cell state and to purified progenitors exhibiting mesoderm and endoderm characteristics. In this context my aim is to focus on changes of RBP activity that is mediated by alternations of the binding propensity rather than to the RBP levels in the cells.

From the methodological perspective, my intention here is to integrate high-throughput sequencing approaches with RBPome capture coupled to global proteomics analysis as well as biochemical and functional assays for unearthing how the post transcriptional layer of gene regulation governs developmental stages, particularly those pertinent to the pluripotency circuitry.

By such integration approach I aimed to identify developmental RBP rearrangements to create a model centered on sequestration of the pluripotent state and stabilization of differentiation poised state through the

¹ (http://rbpmap.technion.ac.il/RBPmap_motif_list.html)

concert action. One such hub that has been implicated in post-transcriptional regulation in concert action are nuclear membraneless granules termed paraspeckles. It should be noted that I have set here to address my hypothesis regarding the function of the paraspeckles, which have been reported to exist only in differentiated cell types, but not in undifferentiated PSCs. Although the full composition and cellular functions of paraspeckles have yet to be clarified, they have been implicated in post-transcriptional regulation by association with RBPs. I therefore reason that paraspeckles are at the center of a crosstalk between RBPs, global post-transcriptional regulation and developmental stages. I therefore aim to elucidate mechanisms of paraspeckles formation and to uncover how paraspeckles regulate pre-mRNA processing which I suspect that they promote differentiation. It is my ultimate goal to create a model centered on sequestration of pluripotency factors and stabilization of differentiation poised state through the concert action of hubs of RBPs and RNAs.

3 RESULTS

3.1 DEVELOPMENTAL ONTOGENETIC RBPOME DURING EMBRYONIC STEM CELL PRIMING AND EARLY DIFFERENTIATION

3.1.1 Recapturing early embryonic development *in vitro*

Early embryogenesis in mammals encompasses a cascade of transient pluripotency states along the path from the blastocyst to the epiblast stage. Many studies have dealt with signaling requirements for maintenance of h/m naïve and primed state PSCs, and uncovered an interplay of underlying Wnt and MEK signals (Figure 2) (Bernemann et al., 2011; Joo et al., 2014; Kurek et al., 2015) Kojima et al., 2014). In addition, Wnt signaling has been shown to promote the formation of the primitive streak in mouse embryos or a similar fate *in vitro* (Drukker et al., 2012; Kim et al., 2013; Kurek et al., 2015). As a preparatory step to study the function of RBPs in development in the context of posttranscriptional modifications that govern cell fate, I established protocols for preparation of homogeneous preparations of h/mESCs along the commitment path from naïve pluripotency to primitive streak-like progenitors.

3.1.2 Identification and isolation of naïve-primed pluripotency transitioning states

To take an advantage of the most advanced staging system of pluripotent states, I established culture systems allowing propagate three stages of pluripotency according to unpublished data of the ten Berge lab (Erasmus MC). In collaborative work with the laboratory of Derk ten Berge we sought to identify whether Wnt or MEK signals regulate the transition to the intermediate state of pluripotency found in the embryonic rosette, using an ESC-based model for rosette formation. Together we demonstrated that rosette-stage pluripotency (Fig. 1) can be indefinitely maintained *in vitro* by inhibiting Wnt/ β -catenin signaling in conjunction to active MEK signaling. Immunostaining confirmed that rosette-like cells (termed Limes) undergo lumenogenesis, with the cells arranged in a monolayer surrounding the cavity stained by sialomucin podocalyxin. Additionally, limes cells express naïve marker Klf4 and but are lacking primed marker Oct6, indicative for epiblast

stem cells (Fig. 8), but are already expressing early stem cells priming factor Otx2.(Fig 9A-B).

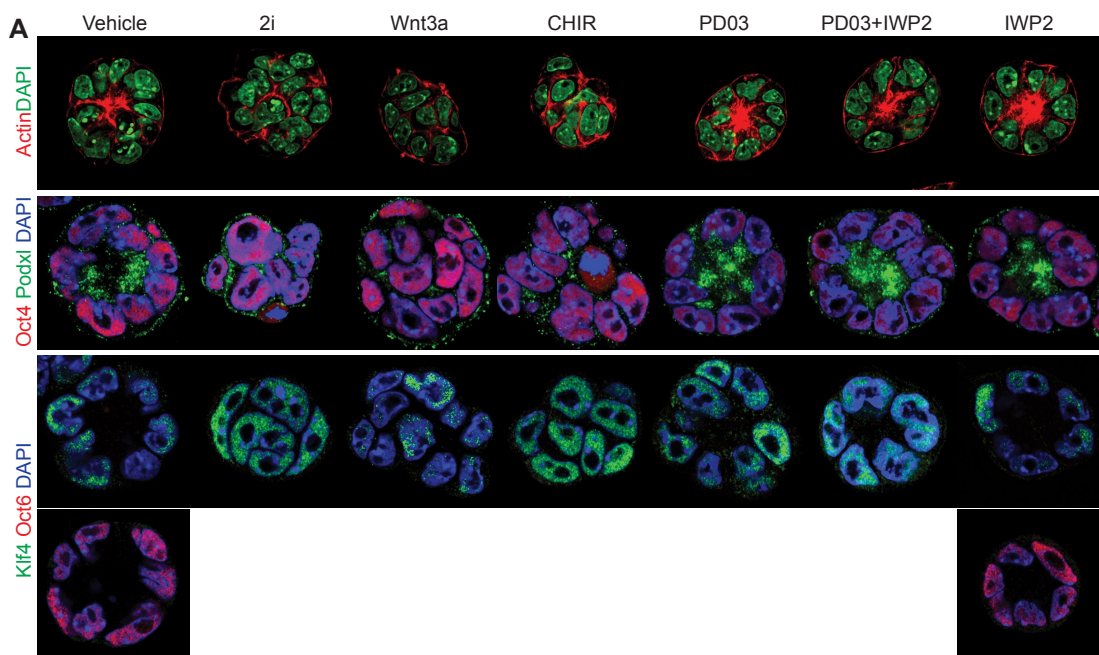


Figure 8: Wnt signalling and Mek inhibition regulate entry to and exit from the limes rosette stage.

(A) Rosettes generated by ESCs 48 hrs after seeding in Matrigel in the indicated conditions using naïve pluripotency conditions (2iLIF) or individual inhibitors: CHIR99021 (GSK3 inhibitor), PD0325901 (MEK inhibitor), IWP2 (Wnt pathway inhibitor) or by adding Wnt3a recombinant protein. Seeded cells were stained for the indicated markers of pluripotency (Klf4, Oct4), stem cell priming (Oct6), lumen formation (Podocalyxin), or actin cytoskeleton (Actin) to label polarized rosettes (courtesy of Derk ten Berge).

In an additional set of experiments, we corroborated the limes stage classification by quantifying the protein content of these cells. Using quantitative proteomic measurements we found that markers of naïve pluripotency including NANOG, ESRRB, REX1, KLF4 and STELLA (DPPA3) are expressed at similar protein levels also in limes cells (Fig. 9A,C), with exception of UTF1 and LIN28A, which were recently identified as first factors that respond to withdrawal of 2iLIF conditions (Galonska et al., 2015; Kalkan et al., 2016). In addition to retained naïve pluripotency circuitry, the proteomics data confirmed that limes PSCs express comparable levels of OTX2 as primed PSCs (EpiSC) (Fig. 8B). In contrast to naïve and limes stage pluripotency, primed state EpiSCs exhibit an alternative pluripotency configuration that lacks naïve markers, but also markers associated with early germ layers such as FOXA2, EOMES, T and SOX17 (Fig. 9B).

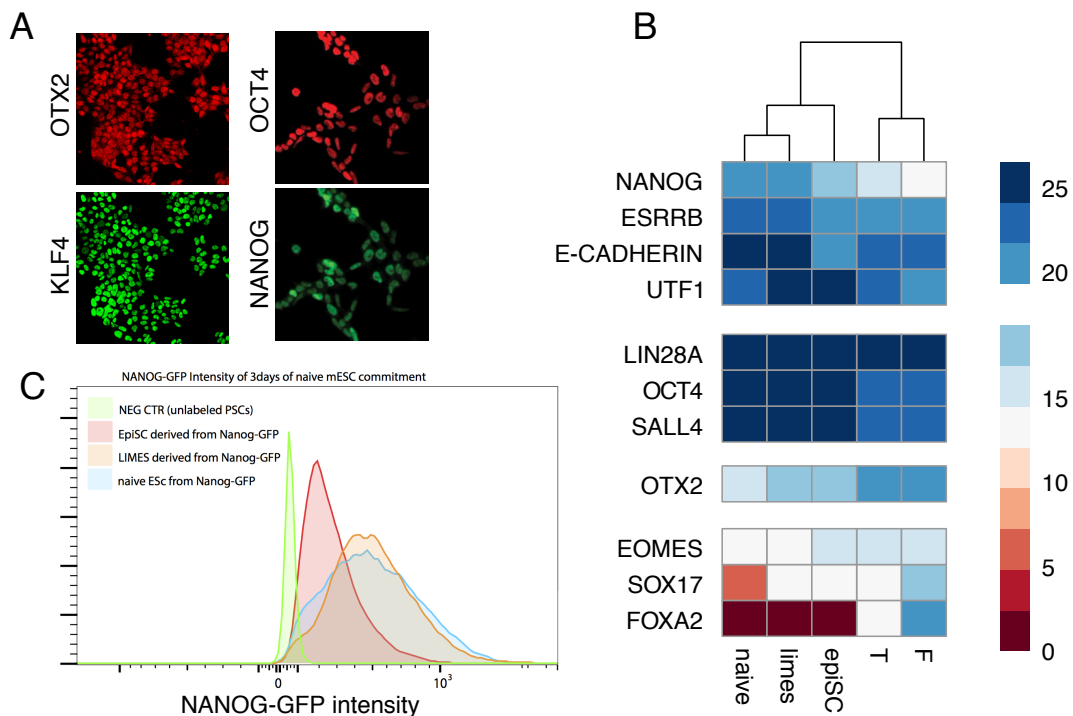


Fig. 9. Limes cells represent an intermediate state of pluripotency.

(A) Immunostaining for the indicated markers plated in limes MEK/Wnt-inhibited conditions containing Lif, PD325901 and IWP2. (B) MS-count (log2) heat map of proteins markers of naïve pluripotency (NANOG, ESRRB), pluripotency factors (OCT4, SALL4, E-CAD), early priming factor OTX2 and lineage specifiers (EOMES, SOX17, FOXA2). (C) A representative flow cytometry diagram of *Nanog-eGFP* reporter line maintained in naïve, limes and primed pluripotency culturing condition.

To observe commitment potential of novel pre-implantation pluripotency layer, I exposed naïve ESCs and limesSCs to a medium containing ActivinA, FGF2 and the Wnt inhibitor IWP2, which promote conversion of naïve PSCs to primed state. I observed accelerated differentiation potential of limesSCs that lose expression of naïve pluripotency markers within 24hrs of stem cell priming. In contrast naïve PSCs revealed depletion of naïve markers first upon 72hrs of priming (Fig. 10). This indicates that limesSCs occupy a bordering state of pluripotency that occurs during the intermediate phase of the naïve to primed pluripotency transition.

Collectively results indicate that absence of Wnt signals with simultaneous MEK inhibition progresses naïve PSCs towards the pluripotency intermediate state of naïve-primed transition with similarities to the rosette stage in embryo development. These stably maintained limes

cells are in vitro counterparts of the embryonic rosette co-expressing naïve and primed pluripotency markers.

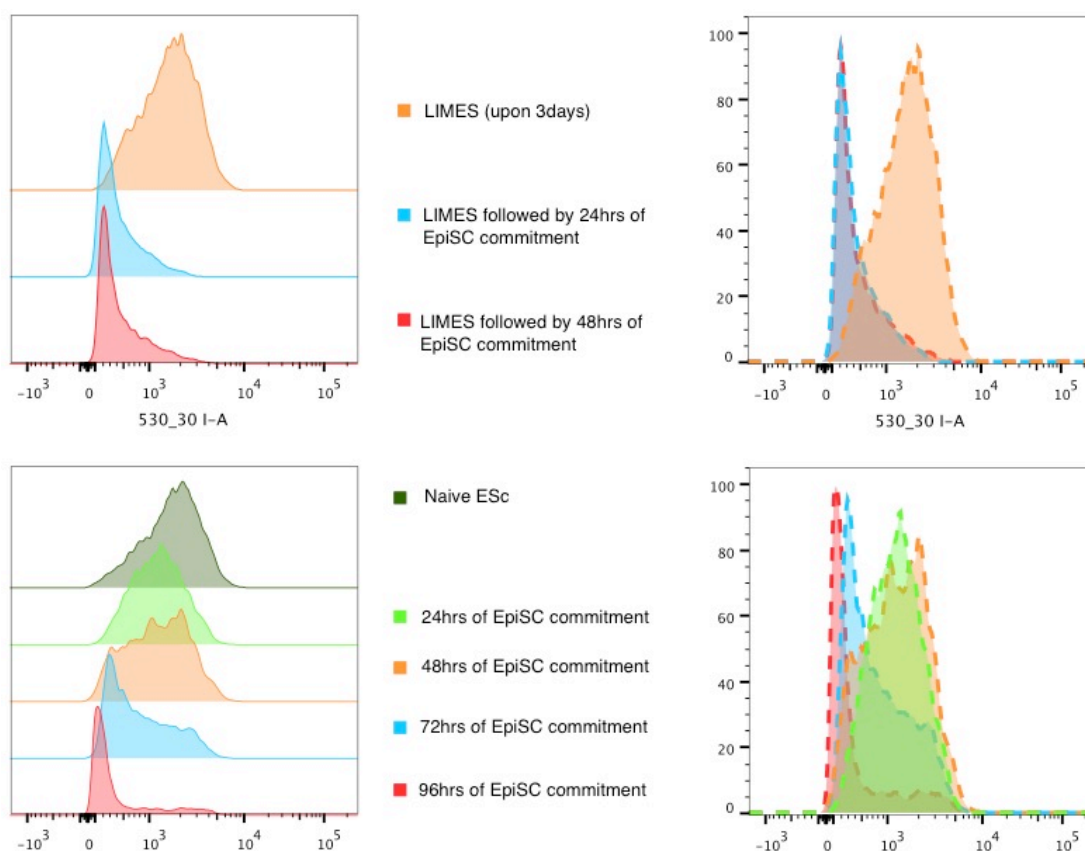


Fig 10: Exit from pluripotency comparing naive and limes PSCs.

A representative flow cytometry analysis of *Rex1-d2GFP* reporter during a time course treatment by priming medium of limes (above) and naïve PSCs (below). Accelerated commitment potential of limes PSCs indicates that limes PSCs are homogeneously accumulated at the edge of the naïve-primed transition.

3.1.3 Isolation of early primitive streak-like progenitors

As a further preparation for analyzing post transcriptional processes using differentiated PSCs I optimized a protocol to generate progenitors that resemble nascent primitive streak (PS) cells. I utilized a double reporter mESC line modified to express distinct fluorescent proteins regulated by the promoters of *T (Brachyury)* and *Foxa2* genes. The expression of *T* and *Foxa2* alone is indicative of mesodermal and endodermal commitment, respectively (Burtscher and Lickert, 2009; Wu et al., 2015) (Fig. 11).

In accordance with recent findings that Fgf/ActA treatment is not optimal for long-term maintenance of EpiSCs (Kurek et al., 2015; Wu et al.,

2015), I observed that historical EpiSCs derived without IWP2 (Tesar et al., 2007) exhibit promiscuous differentiation with high degree (~70%) of *T/Foxa2* positive cells, while addition of IWP2 suppressed lineage differentiation (Fig. 11) and maintained the cells in pluripotency state as indicated by SSEA1 pluripotency marker expression (Fig. 12).

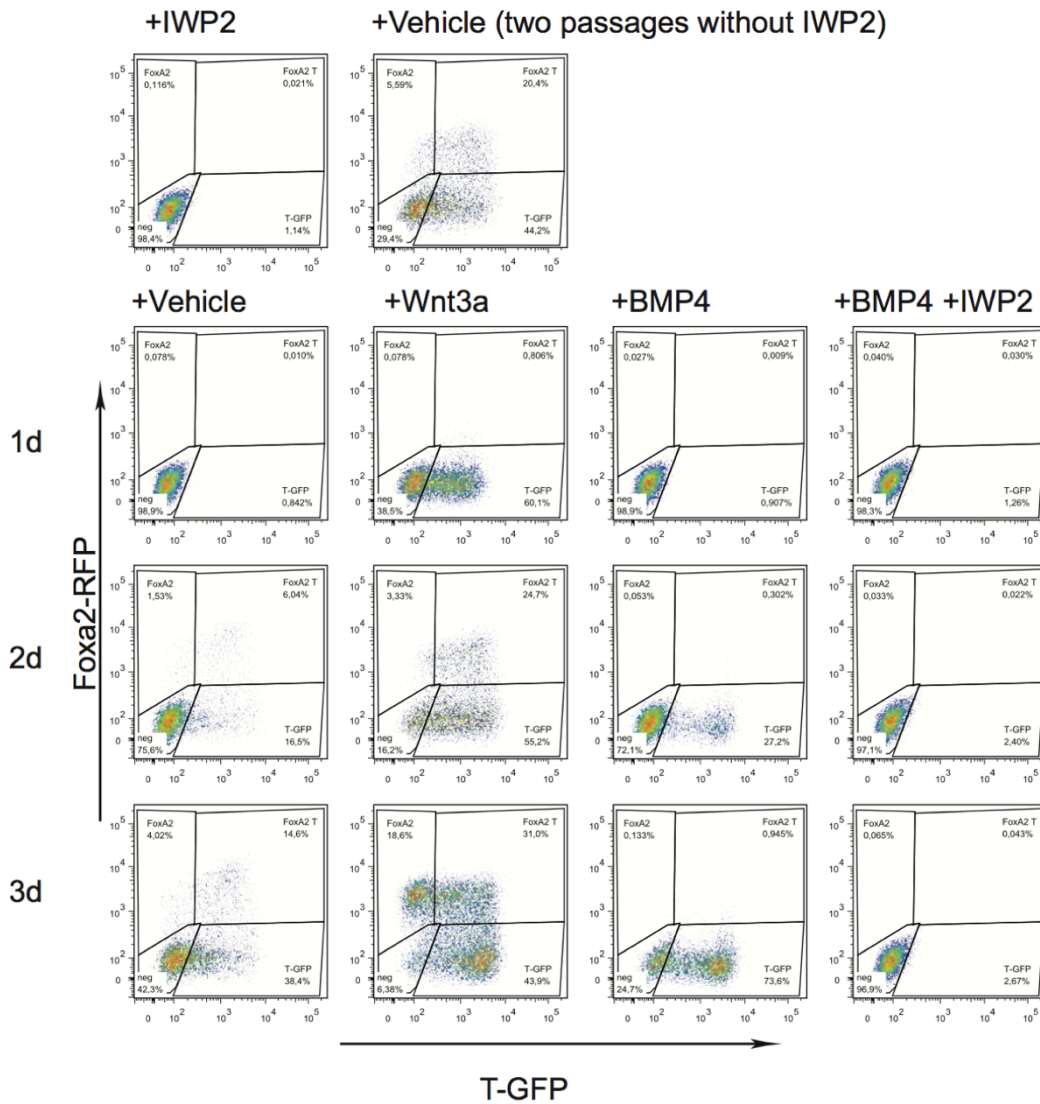


Fig. 11. Differentiation of dual reporter *Foxa2-RFP T-eGFP* EpiSCs. Flow cytometry diagrams of cells treated with the indicated factors and analyzed according to the relevant fluorescent emissions. Cells were maintained in the presence of IWP2 prior to the experiment and vehicle stands for bFGF ActA basal medium.

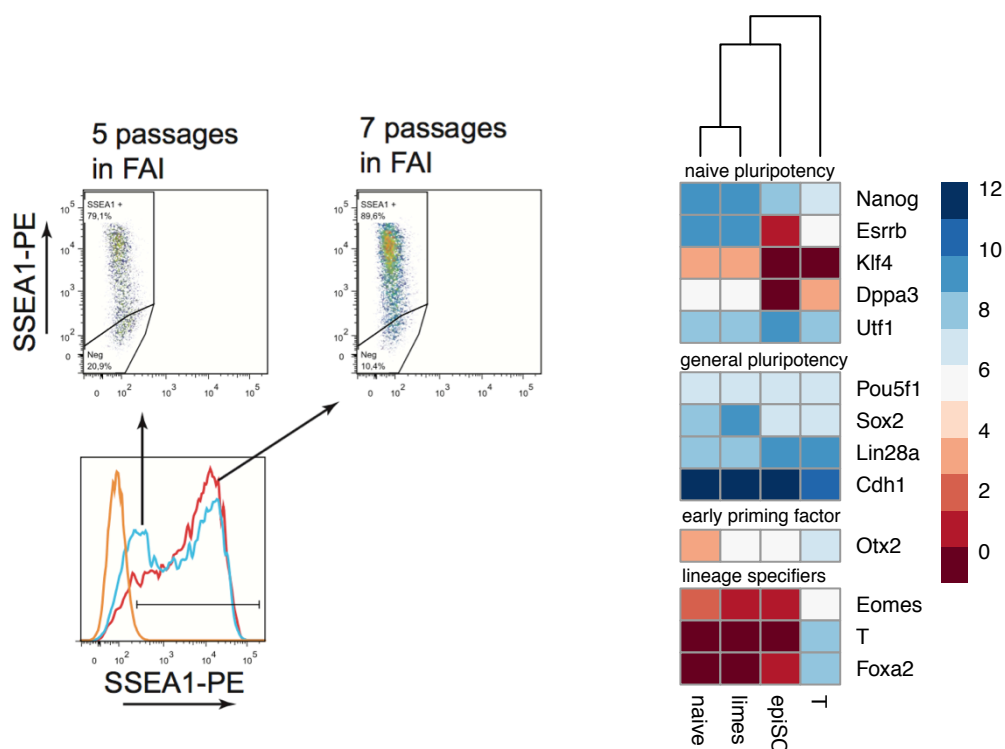


Fig 12: Foxa2-RFP T-GFP EpiSCs are pluripotent.

Flow cytometry plots of Foxa2-RFP T-GFP EpiSCs analyzed for pluripotency marker SSEA1. The cells were maintained in medium containing ActivinA, bFGF and IWP for 5 or 7 passages as indicated.

Fig 13: Expression heat map of main naïve pluripotency factors (*Nanog*, *Esrrb*, *Klf4*, *Dppa3*, *Utf1*), core general pluripotency network (*Pou5f1/Oct4*, *Sox2* and *Lin28a*), epithelial and pluripotency marker E-cadherin (*Cdh1*), earliest priming factor specifying linesSCs (*Otx2*) and lineage specifiers (*Eomes*, *T*, *Foxa2*).

Next, I confirmed that treatment by Wnt3a leads to a rapid upregulation of genes associated with the formation of primitive streak (Kurek et al., 2015). Following 24hours of Wnt3a-driven differentiation I noticed the formation of progenitors with mesodermal characteristics (*T-eGFP+*), then double positive cells and lastly upon 3days of EpiSC differentiation the formation of *Foxa2-RFP+* populations with endodermal characteristics (Fig. 11). This confirms the dynamic expression of these markers in the primitive streak as the embryo undergoes gastrulation as the T expression represents an early event in gastrulation where this gene is expressed in the proximal posterior region of the embryo, while FOXA2 is an anterior marker (Burtscher and Lickert, 2009).

Finally, to confirm the transition from pluripotency to PS progenitor phenotype I performed mRNA sequencing (QuantSeq). This confirms

maintenance of naïve pluripotency network (*Nanog*, *Esrrb*, *Klf4*, *Dppa3*, *Utf1*) and expression of early priming factor *Otx2* in limesSCs (Fig. 13).

3.1.4 Alternative polyadenylation changes accompanying early PSC differentiation

To understand in a greater detail post-transcriptional changes that underlie transitions between pluripotency stages and to primitive streak progenitors, I characterized the global APA landscape in naïve, limes and epiblast PSCs as well as earliest primitive streak like *T-eGFP+* (PS)-progenitors that were produced as outlined above and sorted (Fig 11). I globally analyzed the polyA site (PAS) locations in these populations by expressRNA analysis tools (Gregor Rot et al., 2017). I observed a general trend towards lengthening of transcripts during the transition between pluripotency stages to PS progenitors (Fig. 14A), and identified >400, >700 and >1000 alternatively polyadenylated transcripts comparing naïve, to limes, EpiSCs and *Brachyury+* sorted progenitors, respectively (Fig. 14B). This indicates that RBPs, which regulate APA may be important for transitions between pluripotency stages and differentiation.

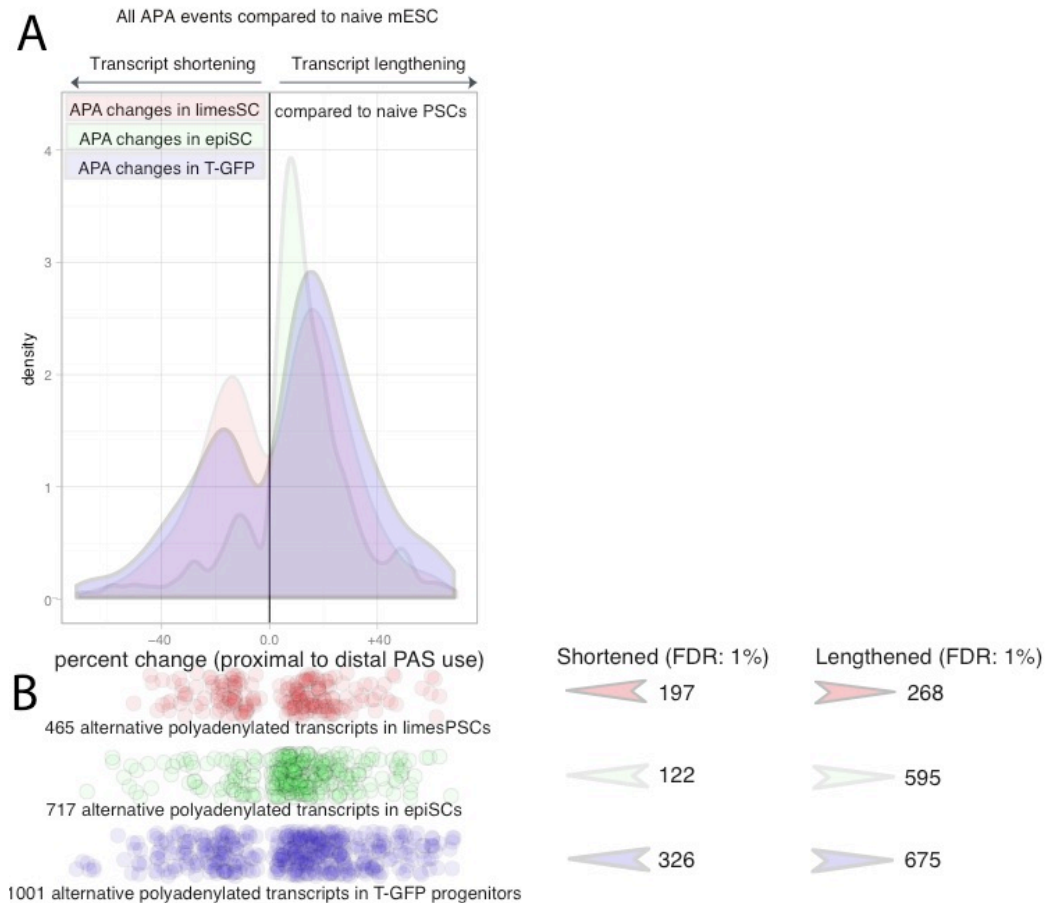
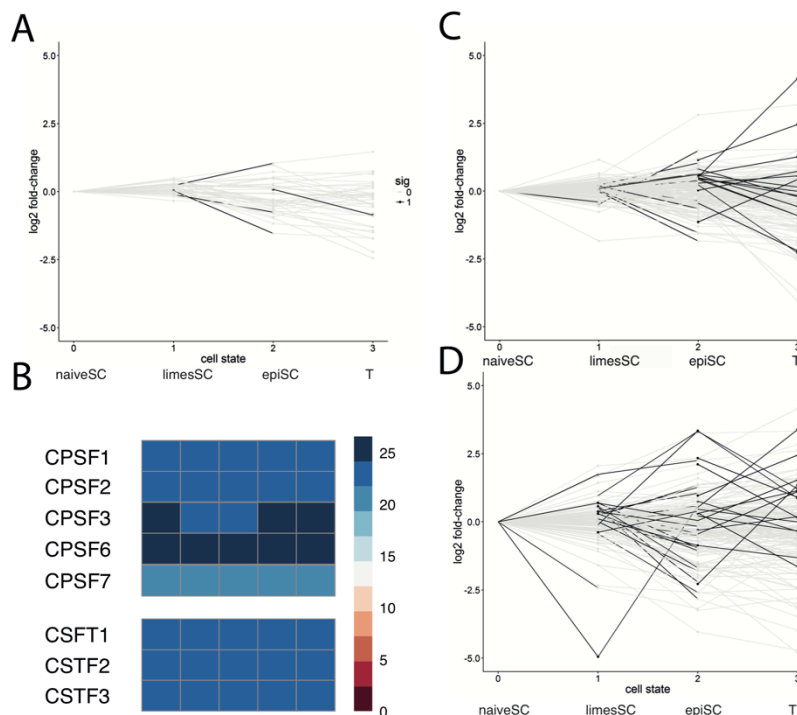


Fig. 14. Developmental alternative polyadenylation.

(A). A density plot depicting the direction and degree of APA upon pluripotency transitioning and primitive streak specification (genes exhibiting APA, $p < 0.01$, Fisher's exact test, $n=5$). (B) The degree of lengthening or shortening per gene detected by global 3'RNA-Seq in mouse naïve, limes, epiblast PSCs and T-GFP progenitors. Each dots depicts individual regulated transcript.

Defined developmental populations provided an insight into cohorts of RBPs that are specifically expressed in continuum of pure pluripotency cultures of naïve-limes-epiblast PSCs, and sorted homogeneous progenitor populations that emerge from epiblastPSCs, presenting a foundation for functional analysis of the leading candidates for the respective developmental pre-mRNA processing changes. In an attempt to explain the observed altered developmental APA landscape, we next compared the protein levels of known polyadenylation (Fig. 15A-B) and splicing factors (Fig. 15C) in the respective stages. Interestingly, we failed to identify significant changes between abundances of polyadenylation or splicing factors (Fig. 15A-C) in early pluripotency transitioning, which are not dynamic to similar degree as observed for transcription factors (Fig. 15D).

**Fig 15: Protein abundances during stem cell transitions and early primitive streak differentiation.**

(A-C) Quantitative MS identified protein changes during naïve->limes->EpiSC->primitive streak (*T-eGFP+*) cell transitions for RNA splicing factors (A), general polyadenylation factors (B), all RNA processing factors (C) and general transcription factors (D).

This indicates that the rearrangement of the RNA processing landscape during pluripotency transitions may be uncoupled of the protein amounts of RNA splicing and processing factors. I thus hypothesized that RBPs can regulate observed RNA processing changes independent of protein amount by dynamic changes of RBP localization, changes in interaction partners of the candidate RBP or its post-translational modifications. To uncover changes in the RNA-binding activity of RBPs I set out to generate ontogenetic mRNA-RBP occupancy assay (RBPome) that represents changes that take place during early embryonic development.

3.1.5 Analysis of the RBPome in PSC states and early differentiation

To uncover the changes in composition of the RNA-bound proteome and the rearrangements that occur in the cohorts of RNA binding proteins bound to mRNAs - termed RBPome - during early embryonic development I modified protocols of RBPome capture and analysis by liquid chromatography–mass spectrometry (LC-MS) to the scale of several million of cells per sample (Fig. 16). This is because 10^7 cells is the upper limit amount of developmental progenitors that can be readily purified by FACS, while existing protocols for RBPome LC-MS analysis utilize $>5^9$ cells per sample (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013; 2016). Briefly, I eliminated RNase digestion of bound RNPs (RNase masks RBP peptides) by instead employing filter aided sample preparation (FASP) on-filter-digestion of the proteins with LysC and trypsin (Zougman et al., 2009), and utilizing stringent washing steps. Additionally, I implemented recently published improvements of the mRNA-interactome protocol by increasing lysis temperature to 60°C and using higher DTT concentrations that allowed the complete removal of major contaminants such as Tubulin (Sysoev et al., 2016).

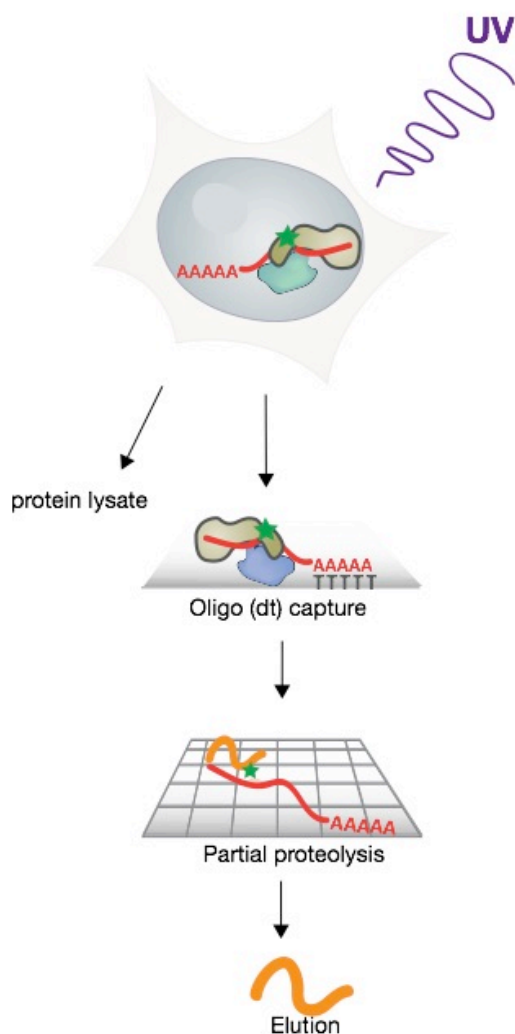


Fig. 16. Schematic overview of the LC-MS RBPome analysis protocol. RNA-protein complexes were cross-linked by exposure of the cells to UV-C, followed by lysis, mRNAs were captured by oligo-dT coupled magnetic beads followed by stringent washing. Eluted RNPs were tryptic digested on-filter and prepared for MS measurements.

To test whether the eluted proteins were indeed specifically bound to RNAs and were not contaminants associated with pulldowns, I carried out silver staining comparing proteins released from oligo-dT beads with or without prior RNase-I treatment. I provided that without the RNase I treatment, the proteins that were eluted from mRNA capture oligo-dT beads were trapped in the gel while with RNase digestion the proteins migrated into the gel (Fig. 17A). In addition, I confirmed the enrichment of RBPs using this protocol by analyzing the bona fide RBP FUS compared to common contaminants such as Beta actin and Tubulin Alpha 1a (Fig. 17A). This proves that the oligo-dT eluted proteins are RBPs.

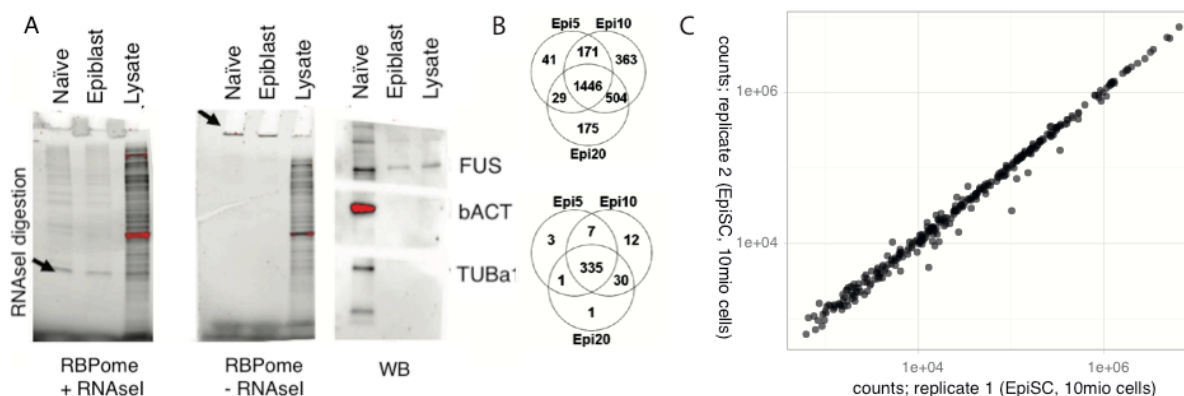


Fig 17: Validation of RBPome pulldown in 10^6 cells.

(A) RBPome samples with RNase I treated or not, left and right, respectively.. Western blot (far right) analysis using antibodies that are specific to ACTB, TUBa1 and FUS. Samples are mouse naive/ epiblast ESCs. (B) Total peptides (above) and proteins (below) identified by MS-LS analysis of 1, 5 and 10 million naive mESCs and presented as Venn diagrams. (C) Scatter point-plot comparing the peptide counts of two RBPome technical replicates of EpiSC generated using 10mio cells.

To determine the lower limit of the cell number for this protocol, I analyzed naive mESCs in the range of 5 to 20 million cells, and noted that the peptide/protein discovery rate was highly correlated even when analyzing 5 million cells (Fig. 17B). Furthermore, my analysis shows that we can reach similar RBP discovery rate by utilizing samples with more than an order of magnitude lower amount of cells. To verify quantitative nature of RBPome capture we measured technical repetitions of LC-MS RBPome measurements and observed near perfect correlation ($r^2=0.993$) (Fig. 17C). To confirm that robust identification of mRNA-bound proteome is quantitative we compared independent biological replicates. We observed near perfect correlation for RBPs between biological replicates ($r^2 = 0.953$) (Fig 18A).

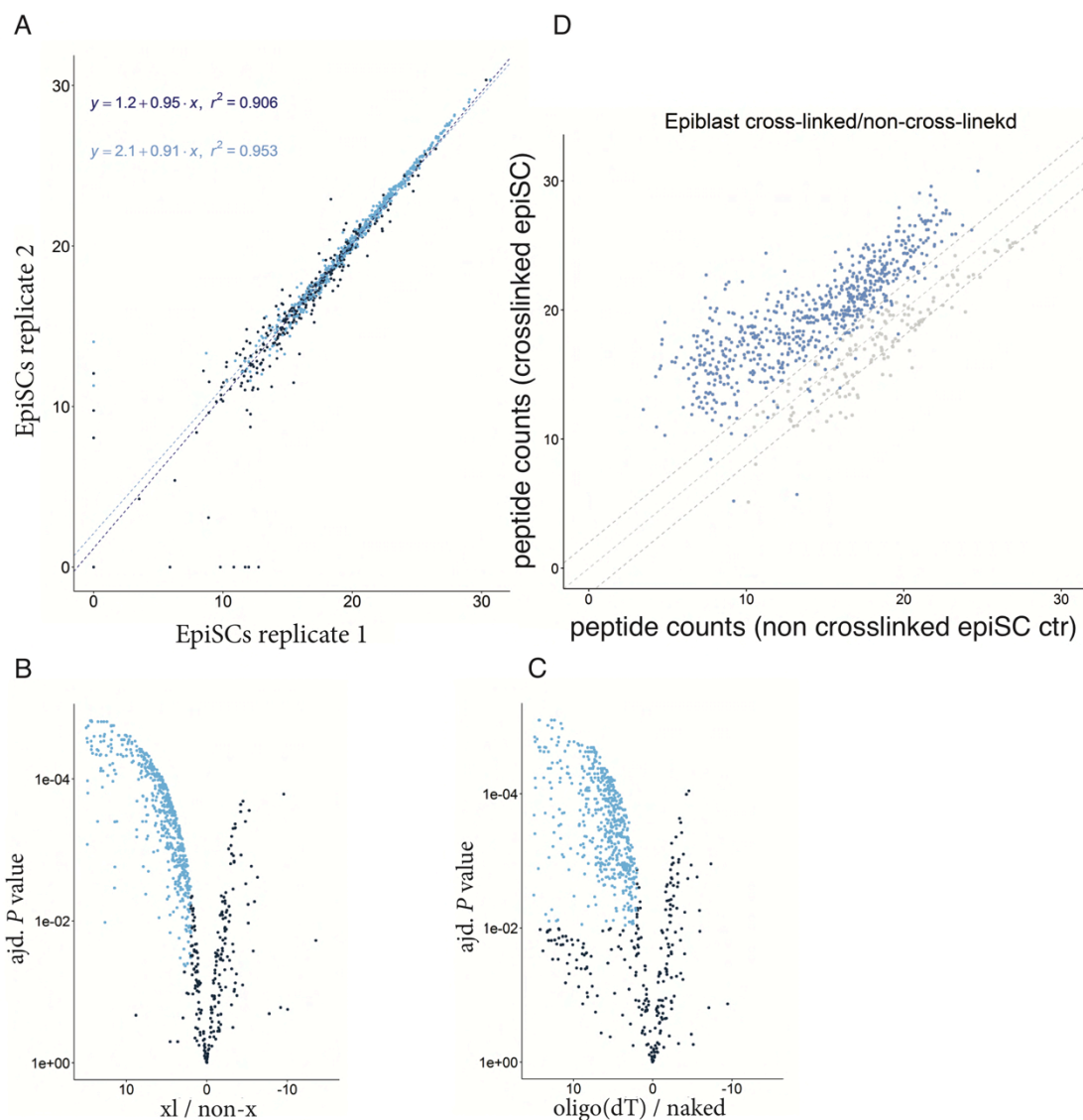


Fig 18: Determining RBPome data integrity.

(A) A scatter plot representing peptide counts detected by LS-MS of corresponding proteins in two EpiSC sample replicates. Blue and black marks correspond to annotated (Gersteberger et al. 2014) ($r^2=0.953$) and novel RBPs (detected in this study, $r^2=0.906$), respectively. The dashed line represents a linear correlation. (B-C) Peptides (blue) passed filtering if they are enriched by a factor of 2 or more compared to non-cross-linked and naked beads samples, left and right, respectively (adjusted P value < 0.01 Benjamini-Hochberg method). (D) A scatter plot representing peptide counts detected by LS-MS of corresponding proteins in one EpiSC sample and one non-cross-linked sample. Blue dots represent proteins fulfilling criteria of statistical enrichment over both negative controls. Axes depict the number of unique peptide identifications. The dashed line represents a linear correlation.

It is noteworthy to mention that measurements and analyses of sample series, including non-crosslinked and naked magnetic bead controls, were run in a single batch. This is because for label-free quantification, which is based on peptide peak intensities in extracted ion chromatograms, it is advantageous to measure all samples within one batch using similar UPLC trapping and separation columns in order to keep sample to sample

variations as low as possible. This is important particularly for assessment of RBPs that are enriched over the controls. Importantly, peptides derived from UV-crosslinked samples typically exhibited higher intensities than contaminants like trypsin and keratins, which also had similar intensities in non-crosslinked control preparations (Fig 18B). To further strengthen a cohort of bona fide RBPome, I implemented additional control using mock pull-downs with control beads lacking oligo(dT) to eliminate not only proteins that unspecifically bind oligo(dT) but also common contaminants of magnetic beads (Fig. 18C).

Further, running all samples in a single batch is important particularly for calculating changes that occur in mRNA-RBP occupancy during early differentiation as outlined below. After completing optimizing of RBPome capture and LC-MS analysis, I conducted differential mRNA-RBP occupancy analysis of homogeneous preparations of naïve-, limes- and primed-state mESCs and purified primitive streak progenitor population by sorting of *T-eGFP* labeled cells. In total, I detected 4798 unique peptides from 853 proteins, and of these 819 were consistently identified (2 of the 3 replicates) in all analyzed developmental stages. Of the latter, 676 proteins were enriched by a factor of 2 or more relative to both non-crosslinked and naked bead controls in at least one of the cell states, 440, 499 and 631 in naive, limes and epiblast PSC states respectively, and 387 in the *T-eGFP* population (Fig. 19A).

Finally I analyzed whole cell amounts (proteome) of the respective stages for normalizing out RBP-mRNA occupancy changes that are due to up- or down-regulation of RBP protein levels. Label-free proteome measurements were highly quantitative, exhibiting almost perfect correlation between replicates (coefficient $r^2=0.999$). By applying a 2-peptide cut-off criterion I identified 3172 proteins in the total proteome measurements, and thereby I narrow down the list of 676 RBPs to 519 RBPs which were identified in both total ontogenetic RBPome and whole cell proteome datasets (Fig. 19C). This set of 519 high confidence RBPs was used to compute what relative proportion of protein amount is identified in RNA bound fraction during cell fate transitions. A GO terms analysis of the final list indicates the expected enrichment of categories related to RNA processing (Fig. 19D).

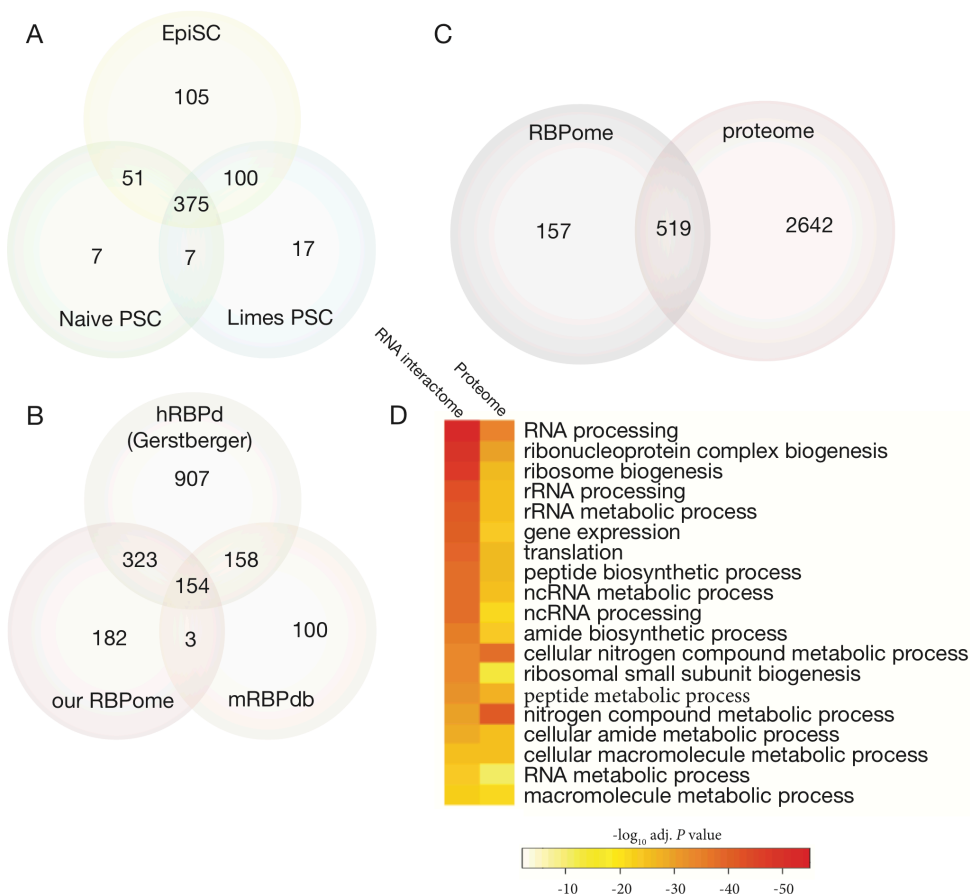


Fig. 19. Proteomic analysis of ontogenetic RBPome.

(A) A Venn diagram depicting the number of proteins identified with high-confidence (enriched over non-crosslinked and naked beads samples) in naïve, limes and epiblast-stage PSCs. (B) A Venn diagram depicting total RBPome list (naïve, limes and EpiSCs states) with the list of annotated RBPs (Gerstberger et al., 2014) and previously identified mESC RBPs grown in LIF and serum (Kwon et al., 2013). (C) A Venn diagram depicting RBPs (enriched over non-crosslinked and naked beads samples) with proteins identified in whole cell lysate proteome label-free MS analysis. The intersect of 519 proteins is further assigned for analysis. (D) Gene Ontology analysis of the ontogenetic RBPome related to (C). “RNA-processing” and other RNA-related “GO terms” were enriched in RBPome compared to total/whole-cell lysate.

3.1.6 Ontogenetic dynamic RBPome

As my primary goal is to identify RBPs that are directly regulated by signaling cascades during cell fate transitions from naïve -> limes -> primed -> progenitor fates, I analyzed the changes in the mRNA-RBP occupancy (RBPome) between states relative to changes in their total amount between the respective states, which was detected by the whole-cell proteomics approach. This process is demonstrated for naïve and epiblast PSCs (Fig 20A). First, comparing the averaged levels of RBPs detected in RBPome between naïve and epiblast PSCs shows only small differences between the states (Fig 20A). Next, calculating the Δ RBPome values for these states and comparing to the Δ Proteome values, shows a complex picture of RBP behavior (Fig. 20B, and C for naïve and limes comparison), including

Class 1: RBPs that do not exhibit abundance differences in the RBPome relative to their whole cell amount between the stages (~80%),

Class 2: RBPs that exhibit correlated changes in both the RBPome and whole cell measurements (~10-13%), and

Class 3: RBPs that exhibit changes in the RNA bound fraction (RBPome) but such that are not correlated with changes of their overall amount between the stages (~5-10%). Importantly, the latter class of proteins exhibits dynamic changes in mRNA binding affinity that doesn't correlate to corresponding protein abundance of given RBP (Fig 20D) either upon transition from naïve PSCs to limes (Fig. 20B) or epiblast pluripotency state (Fig 20C).

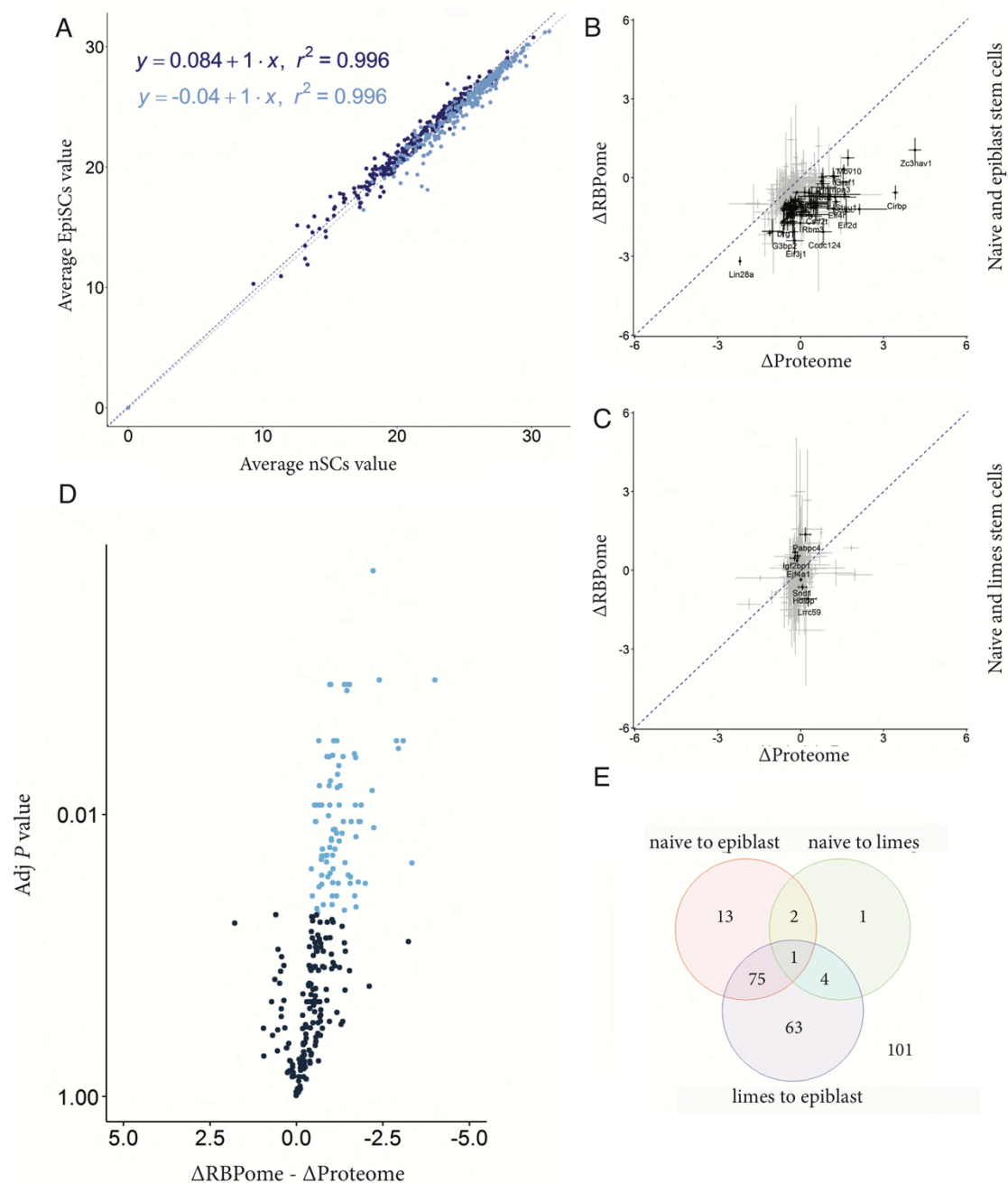


Fig 20. Dynamic associations of RBPs during pluripotency state transitions. (A) A scatter plot depicting averaged RBP intensities of naive and primed-state PSCs ($n=3$, respectively). Blue labels are previously annotated RBPs according to Gerstenberger et al. 2014. Dashed line represents linear correlation. (B-C). A scatter plot depicting Δ RBP amount as detected by proteome or RBPome with respect to naive versus primed (B) or limes states (C). The x-dimension thus represents the difference in the expression level, while the y-dimension indicates the difference in mRNA-binding capacity. Quantitative nature of both methods results in expected values on or near the blue dashed line dissecting quadrants. Proteins that aberrate from correlative proteome/RBPome abundances are characterized as dynamic RBPs (grey) upon limes commitment (B) or epiblast PSCs (C) priming. (D) A Volcano plot depicting proteins from (B) that were counted as dynamic RBPs (blue) if they show that difference abundance ratio naïve/epiblast of each protein in the RBPome versus its naïve/epiblast abundance ratio in the total proteome is statistically enriched (P values were adjusted using Hommel method). (E). A venn diagram depicting the number of RBPs in analysis of pairs of states (with adjusted P values <0.01).

By analyzing the GO terms of the group of RBPs that collectively exhibits a change in RBPome without corresponding changes of expression levels (class 3), I found that they have an annotated role in regulation of translation process (Fig 21A). In contrast, static RBPome (proteins with unaltered abundance of the mRNA-RBP occupancy) is enriched for RNA processing and regulation of gene expression. Of the RBPs that exhibit enrichment in the primed state (Fig. 21B), I noted in particular the pluripotency factor LIN28A, which exhibits higher mRNA-RBP occupancy in primed EpiSC limes PSCs ($p=1.9E-08$, two sided Welch's t-test).

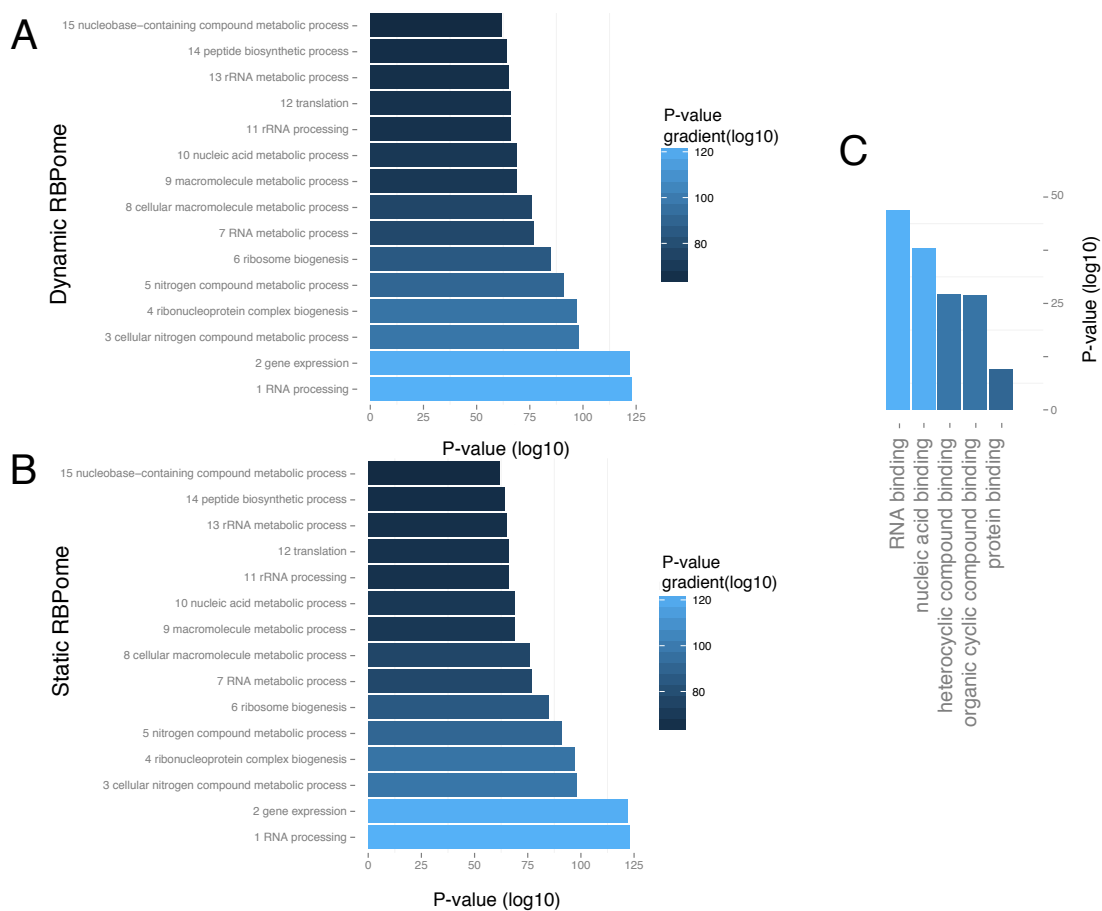


Fig. 21. Functional annotations of dynamic and static RBPs and functional enrichment of proteome PTMs. (A-B) Analysis of the RBPs in the category that exhibits changes in the mRNA-RBP occupancy (A) or all RBPs (B). GO terms analysis by gene ontology.

3.1.7 Signaling-induced RBP translocation

Observing the changes in the mRNA-RBP occupancy during the transition from naive to epiblast pluripotency (Fig. 21B) led me to postulate the direct involvement of signaling pathways in the regulation of cell states via modulation of RBP functions. This is because key developmental signaling pathways, including the MAP kinase cascade, activate the naïve->primed state transitions and lineage commitment. The classical view is that MAPK-ERK regulates gene expression by phosphorylation of transcription factors, however by analyzing functional enrichment of phosphopeptides between naïve and primed cell state I observed that dynamically phosphorylated are mainly RBPs (Fig. 21C). Combined this is a strong evidence to argue that MAPK-ERK also regulates RBPs (Fig. 21C), indicating that there is a tight link between signaling and post-transcriptional regulation. MAPK-ERK, for example, in diverse settings phosphorylates several heterogeneous ribonucleoproteins (hnRNP, Habelhah et al., 2001, Chang et al., 2007, Shimada et al., 2009, Reinhardt et al., 2010), leading to their cytoplasmic accumulation. Recently MAPK/ERK activation was also found to led to phosphorylation of LIN28A in ESCs, which increases its stability (Tsanov et al., 2017).

Because I discovered differential mRNA-RBP occupancy of LIN28A comparing naive and primed states, I set to decipher how LIN28 phosphorylation contributes to the observed increase in its mRNA binding. I first monitored changes of LIN28A locations during transitions of the pluripotency stages and early primitive streak differentiation. To confirm the redistribution of LIN28A by a method that does not depend on antibody staining, I utilized the CRISPR/Cas9 system to recombine (HDR-targeting) a fluorescent gene into the endogenous locus of *LIN28A* and *Lin28a* in human and mouse PSCs (Fig 22). To generate fused LIN28A/lin28a-eGFP proteins without an intervening linker, I introduced the coding sequence of *eGFP* into the *LIN28A/Lin28a* open reading frame right before stop codon. This enables to analyze the location of endogenous LIN28A during the transitions of the pluripotency states. Importantly, I found that LIN28A localization is consistent in the respective stages of h/mPSCs; primed state human ESCs and mouse EpiSCs exhibit predominantly cytoplasmic LIN28A, while in naive-stage mouse ESCs, the protein is predominately present in the nucleus. Importantly, following hPSCs primitive streak differentiation, LIN28A is

extensively translocated to the nucleolus. Consistent, with antibody-mediated tracing of m/hLIN28A, the eGFP fused version also exhibited subnuclear compartmentalization that was observed with antibodies (Fig. 23).

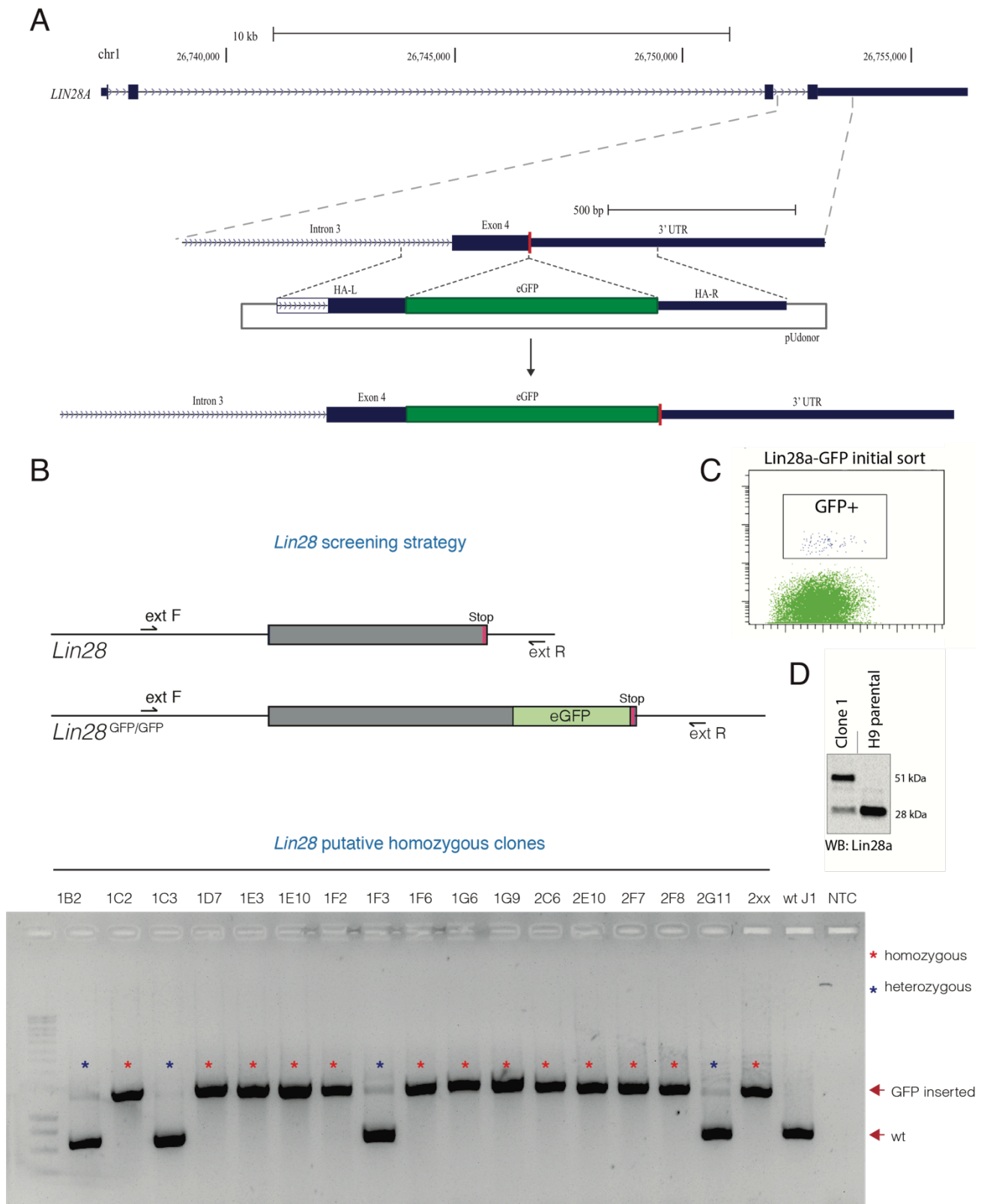


Fig. 22. Generation of LIN28A-GFP fusion PSC lines.

(A) Schematic overview of eGFP insertion into the *Lin28a/LIN28A* locus via CRISPR/Cas assisted gene editing. Target site followed by PAM motif (5'-AGG) is located on the top strand (positive strand) upstream of LIN28A stop codon. Integration is facilitated by double strand breaks created by Cas9 directed to the target sequence by a specific gRNA. (B) Screening PCRs of all CRISPR generated *Lin28a-eGFP* fusion cell lines showing 2 distinct PCR products that are specific either for WT mPSC (lower band) or for Lin28a-eGFP fusion mPSCs. (C) Cells that express a functional Cas9 complex can then be identified by expression of GFP and enriched via FACS. Shown are the sorting of GFP-positive cells following Cas9 and LIN28A gRNA transfection (5 days post-nucleofection). (D) Western blot analysis of clone 1 and parental H9 cells using anti-LIN28A antibody demonstrating the presence of wild type LIN28A and a LIN28A-GFP fusion protein. One clone harbored replacement of both LIN28A alleles.

Importantly, I found that LIN28A exhibit strong enrichment in nucleoli in the naïve PSCs. This was reported previously for the equivalent stage of mouse embryos where LIN28A is exclusively nucleolar in the ICM of early blastocysts (Vogt et al., 2012). Intriguingly, I found that LIN28A translocate to cytoplasm during the transition from naïve to primed state, and upon formation of the epiblast (E5.5) *in vivo* (Fig 23A).

Next, I hypothesized that the developmental translocation of LIN28A alters its biological function. To analyze this possibility, I analyzed interactions of the LIN28A with other proteins. For that I performed the Co-IP of LIN28A-GFP (using an anti-GFP antibody) and LC-MS/MS nuclear extracts of mesoderm derivatives of human ESCs (which exhibit nuclear LIN28A translocation). In total I detected 469 putative proteins enriched by a factor of 6 or more over no-GFP control PSCs (Fig. 5B). Interactome bioinformatics showed that LIN28A interacting proteins include spliceosome components (e.g. U2AF65, SF3B and SF3A), the polyAdenylation machinery as well as other mRNA processing pathways (Fig. 23B).

To understand the functional connection between these interacting RNAs and LIN28A, I next conducted iCLIP using an anti-GFP antibody for immunoprecipitating the LIN28A-eGFP fusion protein. I found that in mesoderm derivatives of human ESCs, nuclear LIN28A accumulation correlates with binding at intron-exon boundaries, including enrichment 3' to splice sites, and immediately downstream to the intron branch point (Fig. 23C), a pattern that overlaps with U2AF65 binding. Because U2AF65 binds to polypyrimidine tract at 3' splice sites, often acting as a repressor of splicing, I hypothesize that LIN28A recruitment to these sites serve a similar function (Fig 23C). In addition we know that other factors that bind near branch point often lead to silencing effects (Witten and Ule, 2011).

Furthermore, because I identified U2 snRNP protein components as putative interactors of LIN28A, this indicates that LIN28A may be recruited to these sites by U2 snRNP and we show enrichment of TDP-43 binding sites at the intron-exon boundaries (Fig 23D-E). Taken together this provides a novel direction to understand nuclear LIN28A role in cell fate decisions.

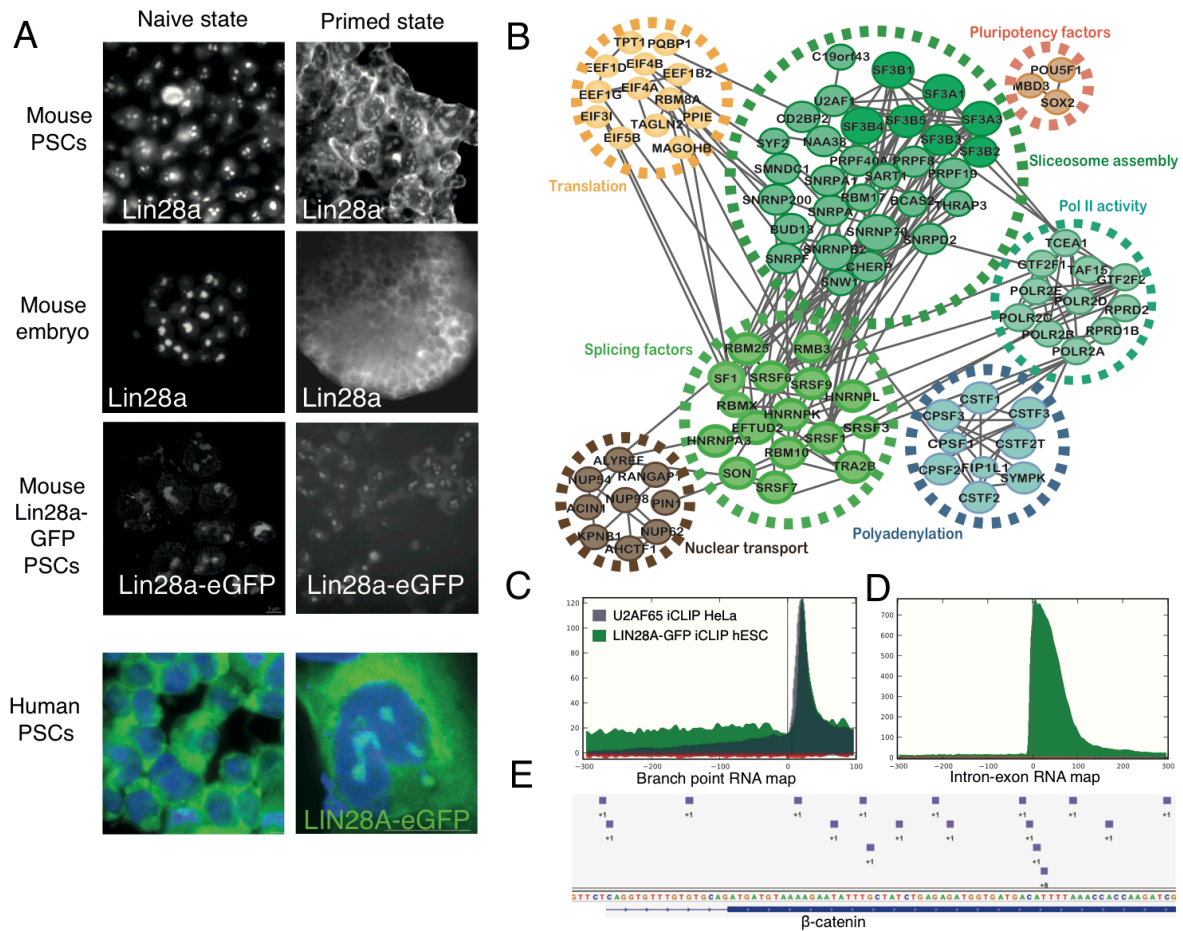


Fig 23: LIN28A redistribution and novel nuclear functions

(A) LIN28A and LIN28A-eGFP (mouse and human, respectively) exhibit naïve and primed pluripotency-dependent distribution in vitro and in mouse embryos. Note that LIN28A localization shifts from nuclear (primarily nucleolar) to cytoplasm upon PSC priming. Mouse blastocyst Lin28a localization taken from (Vogt et al., 2012). Inset shows CRISPR/Cas9 generated LIN28-GFP biallelic fusion protein expressed in human and mouse PSCs, respectively. (B) LIN28A nuclear interactome includes categories of majority of the RNA processing pathways engaged in nucleus. Specifically enriched are polyadenylation factors and components of the U2 snRNP multi-protein spliceosomal complex. (C-D). RNA maps of LIN28A cross-linked sites of regulated pre-mRNAs. (C). Positions of cross-linked nucleotides were plotted for intron-exon boundaries of alternatively utilized exons (C), and surrounding the branch point sequence (D, E). An example of LIN28A binding sites in the mesoderm-inducer β -catenin, reveals intronic binding adjacent to 3'splice site, indicating potential involvement in β -catenin alternative splicing.

3.2 CROSS-REGULATION BETWEEN TDP-43 AND PARASPECKLES PROMOTES PLURIPOTENCY- DIFFERENTIATION TRANSITION

What mechanisms govern and maintain cell states during the process of differentiation is a pivotal question in science. Alternative polyadenylation (APA) is a process that diversifies the 3' end of mRNAs and such that appears to be closely related to cell state. Changes of polyadenylation patterns have been observed upon mouse myogenesis of C2C12 cells (Ji et al., 2009) and adipogenesis in 3T3-L1 cells (Hoque et al., 2013). However, it is not clear whether the APA landscape is modulated upon exit of pluripotency, how and if APA contributes to the dismantling of pluripotency and which factors regulate APA during this transition.

3.2.1 Alternative polyadenylation landscape in early PSC differentiation

To study the APA regulatory program, I differentiated mouse and human PSCs into precursors of the embryonic layers called mesoderm progenitors. I used mouse embryonic stem (ES) cell line bearing an insertion of GFP into the *T(Brachyury)* locus (*Bra::eGFP*), and sorted the eGFP+ population following 72hrs of Wnt3a and Activin A treatment, which is a condition known to promote differentiation of primitive streak cells that mimic nascent mesoderm (Kurek et al., 2015) (Fig. 24a). To confirm this embryonic commitment, I compared the gene cohorts in the treated cells to undifferentiated PSCs. This comparison is based on the Quant-Seq technique, which characterizes the PolyA sites of mRNAs and can be used for quantitative gene expression measurements (Fig. 24a). Similarly, I derived a population of human primitive streak-like progenitors by exposing human PSCs to Wnt signaling for 72 hrs using a small-molecule inhibitor of GSK3b (CHIR99021) according to a published protocol (Blauwkamp et al., 2012). RNA sequencing and analysis of the differentially expressed genes with respect to undifferentiated cells confirmed the lineage restriction of early primitive streak progenitors (Fig. 24b).

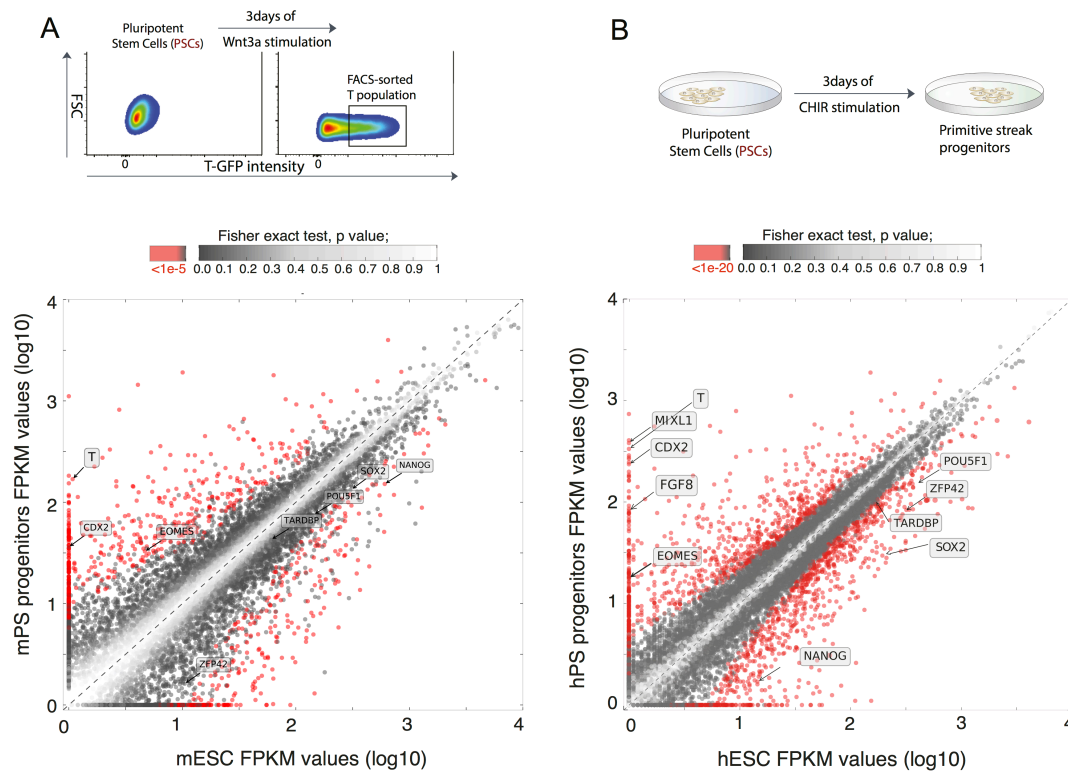


Fig 24. Differentiation of mouse and human PSCs into primitive streak precursors.

(A) representative scatter plot showing transcripts detected by RNA-Seq in PS-like 72hrs progenitors and undifferentiated mouse (A) and human (B) PSCs. Differentially expressed genes with adjusted p-value $< 1e-5$ and $< 1e-20$ (Fisher's exact test, false discovery rate 1e-15) are labeled red ($n=2$ /condition) in mouse and human comparison, respectively.

Next in collaboration with Gregor Rot, we used the computational integrative web platform called apaExpress (Rot et al., 2017), to analyze the locations of PASs (Fig 25A) and total number of PASs extracted from the Quant Seq data in h/mPSCs and PS progenitors (Fig. 25B). We validated the identification of PASs by examining the overlap with a published PAS dataset (Derti et al., 2012) (Fig. 25C), by analyzing the efficiency of cleavage (Fig. 25D), and by comparing the nucleotide composition of newly identified PASs (Fig. 25E) with known PASs (Fig. 25F).

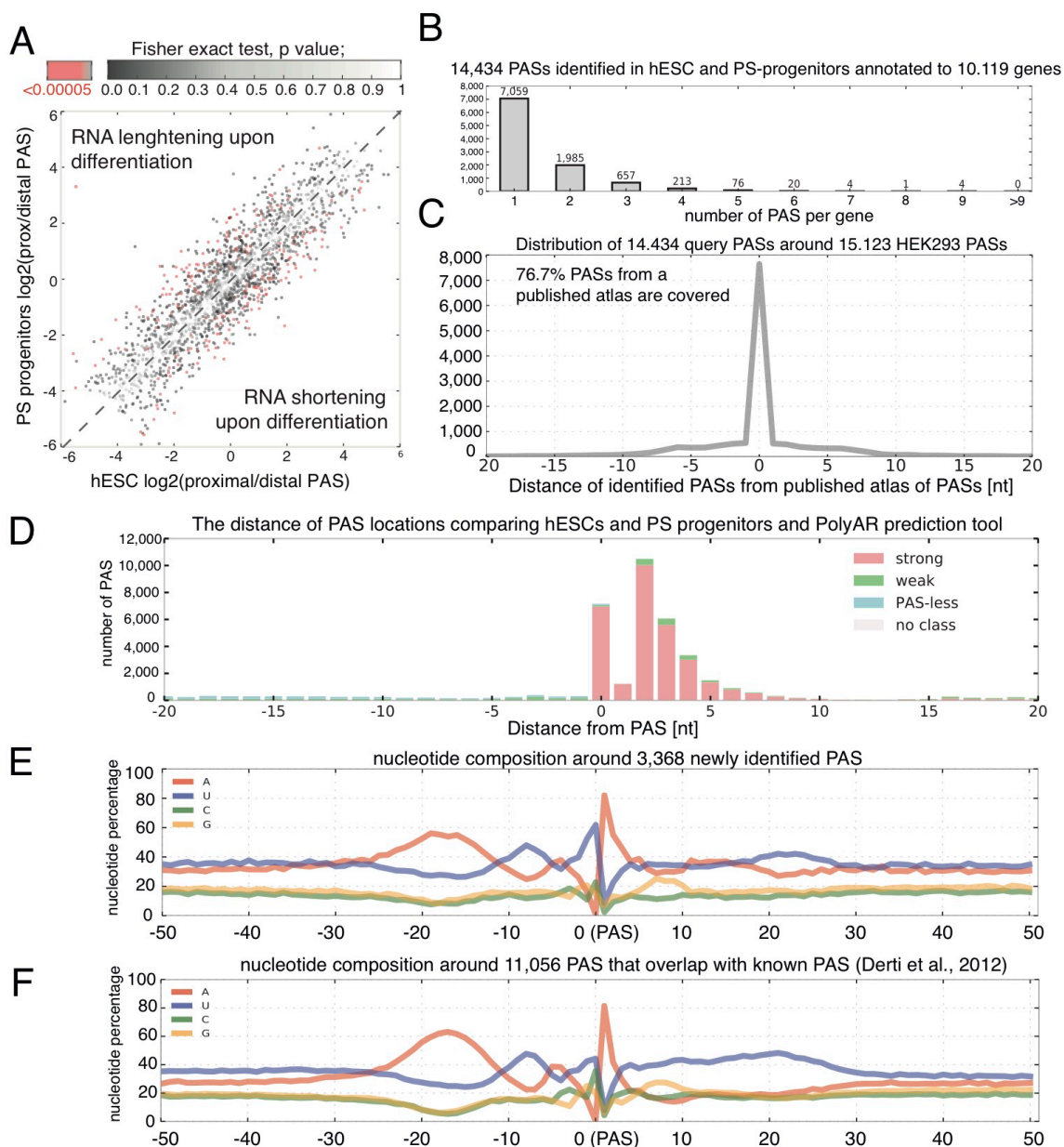


Fig 25: Global analysis of polyA sites.

(A) Quantitative analysis of PAS positions in undifferentiated hESCs and PS progenitors that includes lengthened and shortened 3' UTRs. Significant changes (adjusted p-value <0.001 , Fisher's exact test) are colored red ($n=4/\text{condition}$). (B) The number of PASs per gene detected by global 3'RNA-Seq in hESCs and PS progenitors. (C) The PAS positions comparing PASs detected in hESCs and PS progenitors versus PASs identified in HEK293 cells (1). (D) Locations of PASs detected in hESCs and PS progenitors versus predicted PASs (PAS-strong, PAS-weak and PAS-less, based on the presence of poly(A) signals) classified using polyAR (2). (E-F) The nucleotide composition around newly identified PASs in hESCs and PS progenitors (E), or PASs that overlapped with those of Derti et al., (1) (F). The close homology supports the *bona fide* classification of novel PASs in this study.

To identify mechanisms that regulate the APA events observed during PS differentiation, I analyzed sequence motifs that are enriched close to the PASs in undifferentiated hESCs versus PS progenitors. We noted that GUG

was the most significantly enriched motif in the region upstream of the poly(A) signal of the regulated PASs, where accessory regulatory sites typically reside (Di Giammartino et al., 2011) (Fig. 9A). As TDP-43, an important regulator of alternative splicing, binds to such motifs (Lukavsky et al., 2013; Tollervey et al., 2011b), I next assessed if it underlies the regulation of APA during hESC differentiation.

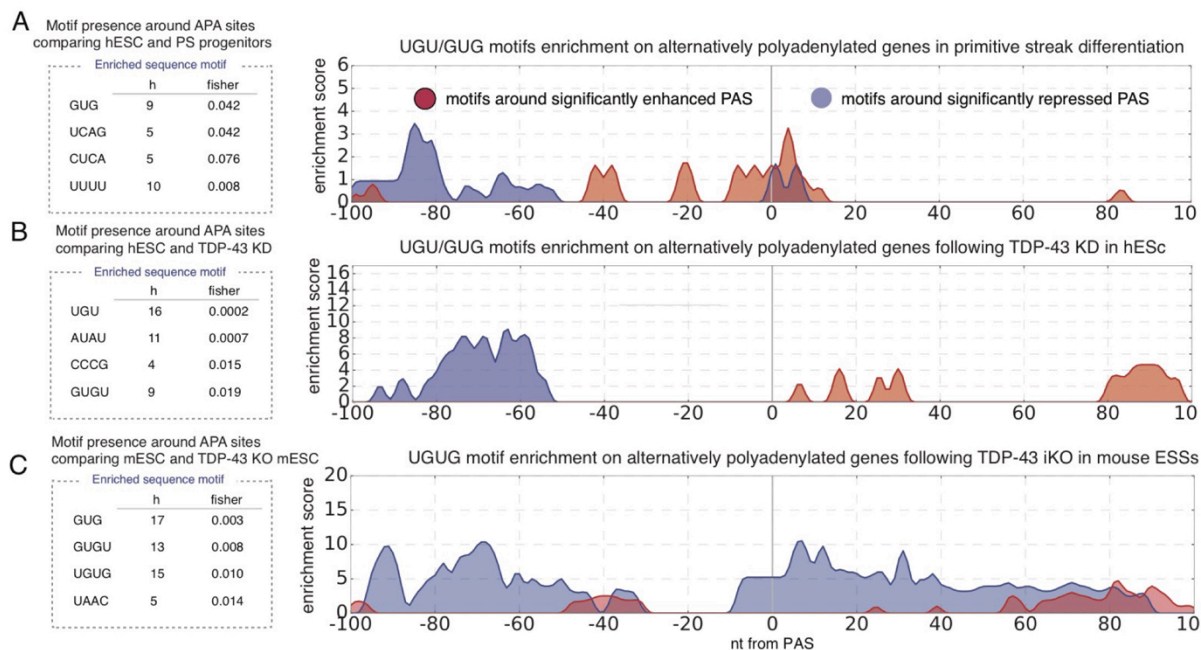


Fig. 26. Analysis of sequence motifs enriched around differentiation- and TDP-43-regulated PASs. (A) The sequence motifs that are most significantly correlated with the enriched PASs in hESCs compared with PS progenitors, hESCs compared to sh*TDP-43* KD (B), and in mESCs compared with KO of *Tdp-43* (C).

We found that the knockdown (KD) of *TDP-43* in undifferentiated hESCs also gives rise to a significant enrichment of the GUG motif in the region upstream of the poly(A) signal of the regulated PASs (Fig. 26B) and similar enrichment was observed also by inducing a knockout (KO) of *Tdp-43* in mouse ESCs (mESCs) (Fig 26C), indicating evolutionary conserved role TDP-43 in regulation of APA.

3.2.2 TDP-43 regulated developmental alternative polyadenylation

Strikingly, nearly two thirds of the differentiation-induced APA events took place in the same genes and the same direction as following *TDP-43* KD (Fig. 27A), and were highly correlated (Fig. 27B) ($R=0.62$, Pearson correlation, p -value $< 2.2e-16$, Fig. 27A). In contrast, we observed no correlation in transcript abundance between PS progenitors and TDP-43 depleted cells (Fig 27C), indicating that TDP-43 regulates only APA but not transcriptional changes during differentiation of primitive streak progenitors. Finally, we found that the level of TDP-43 protein decreases upon PS differentiation (Fig. 27D-E). Taken together, this indicates that the expression of TDP-43 promotes global PAS rearrangement in ESCs.

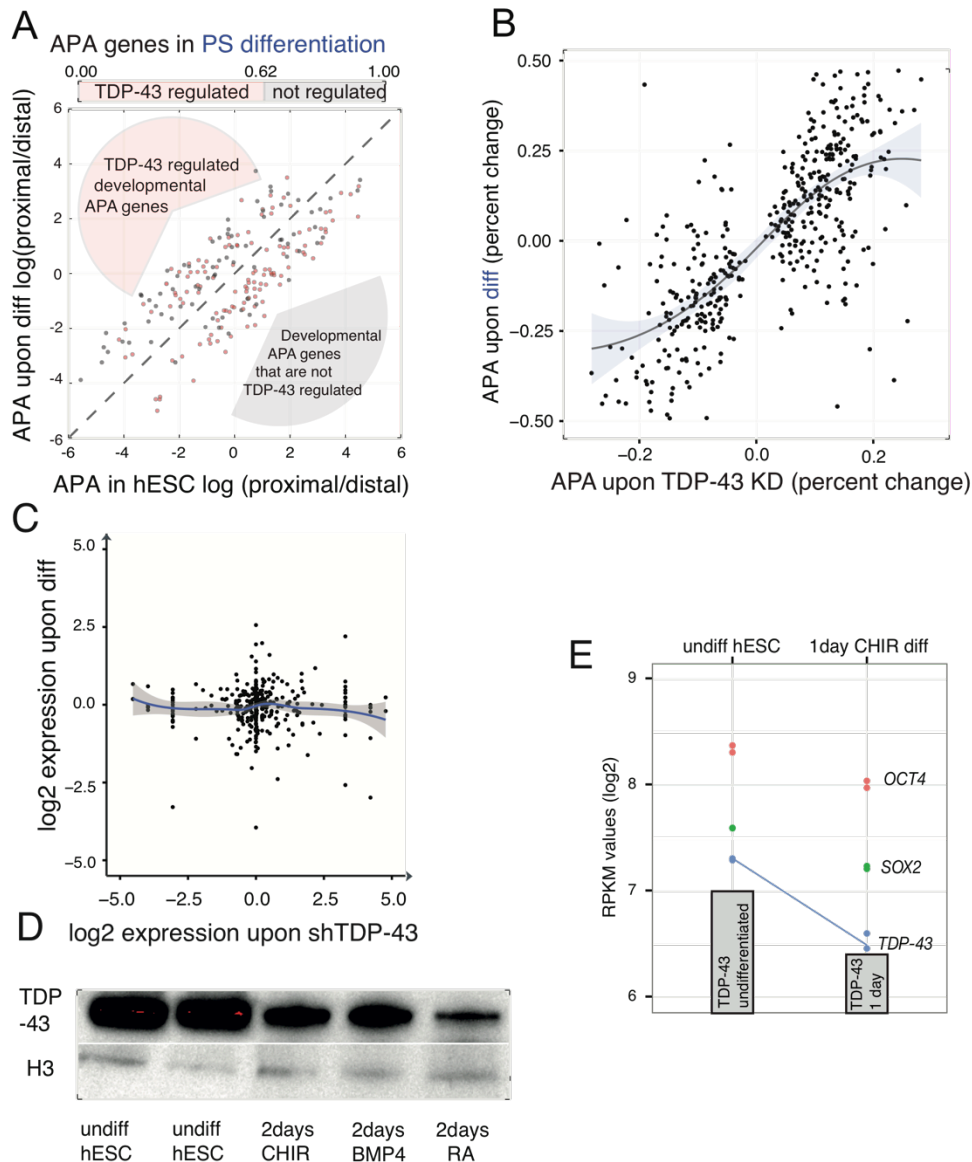


Fig 27: TDP-43 does not directly regulates levels of genes exhibiting APA upon differentiation.

(A) scatterplot indicating differentially polyadenylated genes that are also regulated by TDP-43 (red) or are not TDP-43 dependent (grey) (B) A scatterplot showing the direction of APA, with preference for proximal (positive values) or distal PASs (negative values) upon PS progenitor differentiation of hESCs and sh*TDP-43* KD versus control shRNA. Linear regression (grey line) and the 90% confidence interval region (light blue) are shown (Pearson's correlation coefficient, $R=0.62$). Threshold set for TDP-43 regulated PASs is p-value < 0.05. (C) A scatterplot of associations in transcript level changes upon *TDP-43* KD relative to undifferentiated hESCs, and in PS progenitors relative to undifferentiated cells. Limited to TDP-43 APA to genes exhibiting APA following *TDP-43* KD ($n=2$ / category). (D) A Western blot of TDP-43 and Histone H3 using samples obtained in hESCs and upon 48hrs of directed differentiation towards mesoderm, trophoblast, neuroectoderm progenitors using CHIR99021, BMP4, retinoic acid (RA) treatment, respectively. (E) A combined plot showing changes in expression of *OCT4*, *SOX2* and *TDP-43* mRNAs and TDP-43 protein after 24hrs of CHIR99021 treatment compared to undifferentiated hESCs.

Congruently with the prevalent direction of APA in Zebrafish gastrulation (Li et al., 2012), we observed also a general trend of transcript lengthening of transcripts in the transition from pluripotency to PS progenitors (Fig. 28A). Similar transcript lengthening was observed also upon *TDP-43* knockdown (Fig 28A-B).

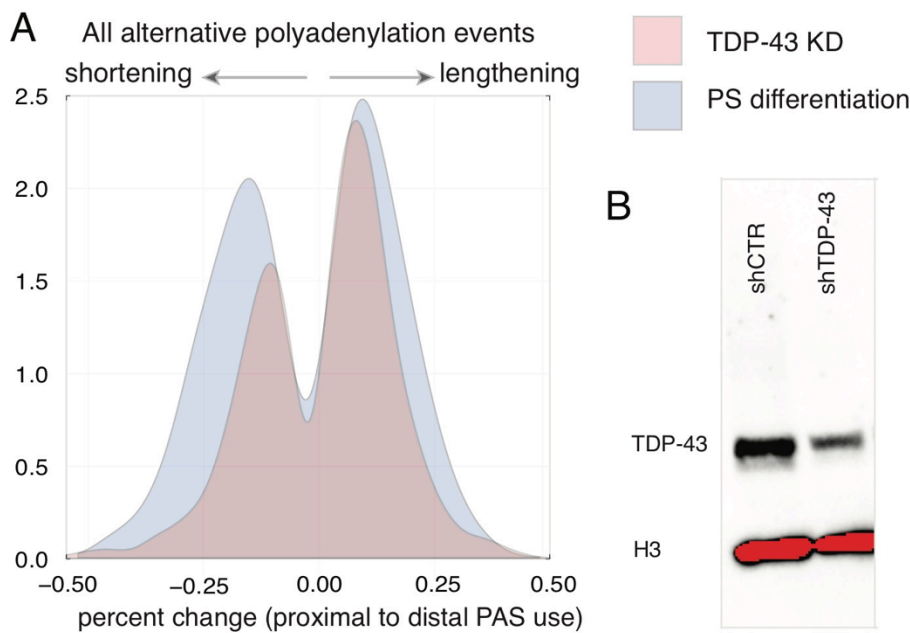


Fig 28: alternative polyadenylation landscape changes upon human PSC differentiation and TDP-43 KD.

(A) A histogram depicting the direction and degree of APA during PS differentiation and following *TDP-43* KD in human ESCs. (B) A representative Western blot analysis of TDP-43 shRNA-transduced hESCs and cells treated with control shRNAs. Histone H3 is used as a loading control.

To decipher if TDP-43 directly regulates APA during early differentiation, we analyzed the RNA-binding profile of TDP-43 in hESCs by individual-nucleotide resolution crosslinking and immunoprecipitation (iCLIP)

(Tollervey et al., 2011b) (Fig. 29A). We assessed the enrichment of pentamers within 50 nucleotides surrounding crosslink sites. iCLIP revealed that the most significantly enriched pentamers are GUGUG and UGUGU (Fig 29B). Analysis of the RNAs bound by TDP-43 showed that 7,5% of all crosslinked sites are located in 3'UTRs (Fig 27C). I therefore reasoned that TDP-43 has an important role in direct regulation of APA.

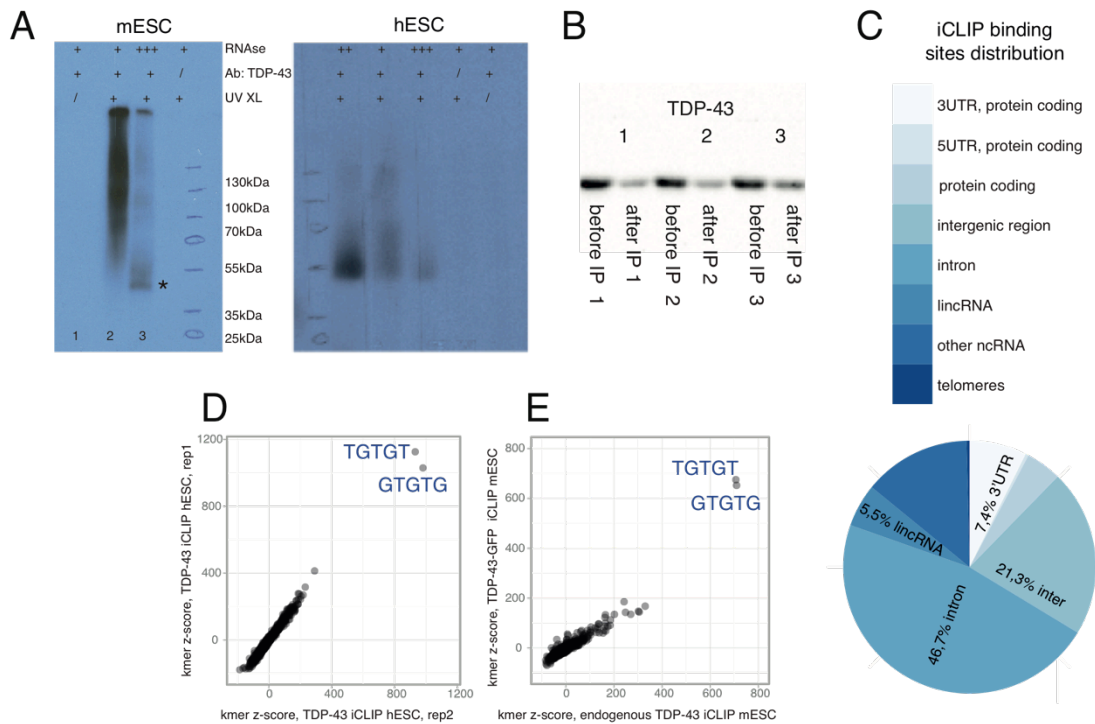


Fig. 29. TDP-43 iCLIP PSCs

(A) An autoradiograph of ^{32}P -labelled RNAs in the presence or absence of UV crosslinking and an anti-TDP-43 antibody in mESC (left) and hESC (right). High and low concentrations of RNase were used to confirm the presence of RNA bound to TDP-43. The asterisk marks the position in the gel corresponding to the size of a TDP-43 monomer. (B) A representative TDP-43 Western analysis of mESC extracts before and after crosslinking and TDP-43 immunoprecipitation. (C) A pie chart depicting the regional distribution of TDP-43 binding sites identified by TDP-43 iCLIP in mESCs. (D) Z-scores of pentamer occurrences surrounding (-30 nt to $+30$ nt) all TDP-43 cross-linked sites in hESCs as determined by iCLIP. The sequences of the two most enriched pentamers are depicted. The Pearson's correlation coefficient between the two shown samples is $R=0.91$. (E) Z-scores of pentamer occurrences surrounding (-30 nt to $+30$ nt) all TDP-43 cross-linked sites in mESCs as determined by iCLIP, using either TDP-43 antibody targeting the endogenous protein (x-axis) or using antibody targeting GFP of the overexpressed TDP43-GFP (y-axis). The sequences of the two most enriched pentamers are depicted.

We further plotted TDP-43 binding sites detected by iCLIP around genomic locations of alternatively polyadenylated mRNAs and marked the polyA site at the center of the RNA maps by a red line (Fig. 30A). The approach for visualizing an RNA map of APA is analogous to any other region of interest, such as for instance the intron-exon boundaries in the

context of alternative splicing (Ule et al., 2006). An RNA map of summarized TDP-43 crosslinking, or of GU-rich motifs, around the PASs that undergo APA following *TDP-43* KD, suggests position dependent dual role of TDP-43 in regulating APA. The RNA map detects enriched TDP-43 binding in the 25-100 nucleotide (nt) region upstream of the PASs that are repressed by TDP-43, and to a lesser extent further than 100nt upstream and downstream of the PASs that are enhanced by TDP-43 (Fig. 30A). This indicates that TDP-43 regulation of APA is position-dependent, exhibiting a repressive effect on polyadenylation when TDP-43 binds close to a PAS, and conversely, an enhancing effect on APA when TDP-43 binds further away from a PAS (Fig. 30B-C). By inducing a KO of *Tdp-43* in mouse ESCs (mESCs) we noted that the dual regulatory role of TDP-43 is evolutionary conserved (Fig. 30D).

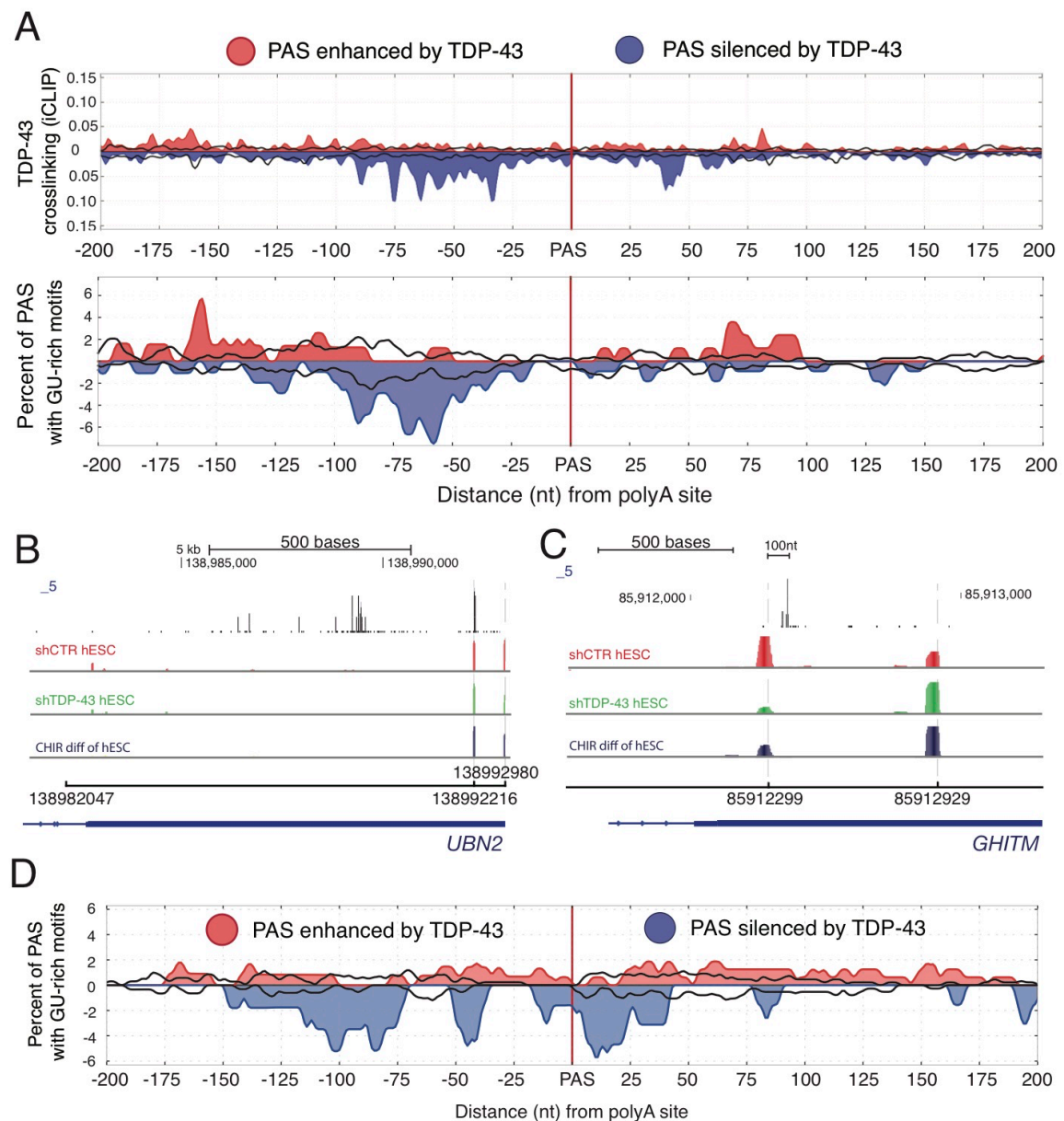


Fig. 30. TDP-43 is a conserved position-dependent regulator of APA.

(A) An “RNA map” showing the positions of TDP-43 crosslinking (top) or GU-rich motifs (bottom) summarized across all the APA sites comparing scrambled shRNA control hESCs (n=4) with TDP-43 KD (n=8). Crosslinking at the PAS that are enhanced by TDP-43 (89) is shown in red, at the repressed PAS in blue (119), and at the unchanged PASs (893) in black lines. (B-C) Representative diagrams of UBN2 (B) and GHITM (C) 3' UTRs showing cross-linked positions of TDP-43 in undifferentiated hESCs (top, black bars), and two alternative mRNA isoforms as determined by 3'RNA-Seq (bottom). The TDP-43 KD effect in hESCs (green) indicates that TDP-43 represses the proximal cleavage site of UBN2, and conversely, enhances the use of the proximal PAS of GHITM. The same trends were noted in human ESC-derived PS progenitors (red). (D) An “RNA map” showing the positions of TDP-43 crosslinking (top) or GU-rich motifs (bottom) summarized across all the APA sites comparing scrambled shRNA control mESCs (n=4) with *Tdp-43* KO (untreated samples N=12, and tamoxifen-induced samples N=10). Crosslinking at the PAS that are enhanced by TDP-43 (316) is shown in red, at the repressed PAS in blue (127), and at the unchanged PASs (1616) in black lines.

3.2.3 Alternative polyadenylation of pluripotency circuitry mRNAs

To understand whether the regulated APA events contribute to the exit from pluripotency, I focused on the conserved targets of mouse and human TDP-43. I identified 523 such common targets (38% of the human compendium) in h/mESCs by shRNA KD and conditional KO, respectively. This includes the h/m pluripotency core factor *SOX2/Sox2* transcript, which undergoes a 10-fold lengthening of its 3' UTR upon *TDP-43* KD and hESC differentiation (Fig. 31A). Importantly, I found that the TDP-43 crosslinking pattern and the APA direction of *Sox2* are consistent with the modes of TDP-43 regulation indicated by the RNA map (Fig. 30): TDP-43 binds close to the repressed distal PAS and far from the enhanced proximal PAS (Fig. 31B), indicating bimodal regulation, respectively. The functional link between TDP-43 and *Sox2* is supported by further findings including *Sox2* transcript and protein down-regulation upon conditional KO of *Tdp-43* in mESCs (Fig. 31C), which also promotes accelerated differentiation, as evident by the loss of the dome-shaped compact colony morphology (Fig. 31D).

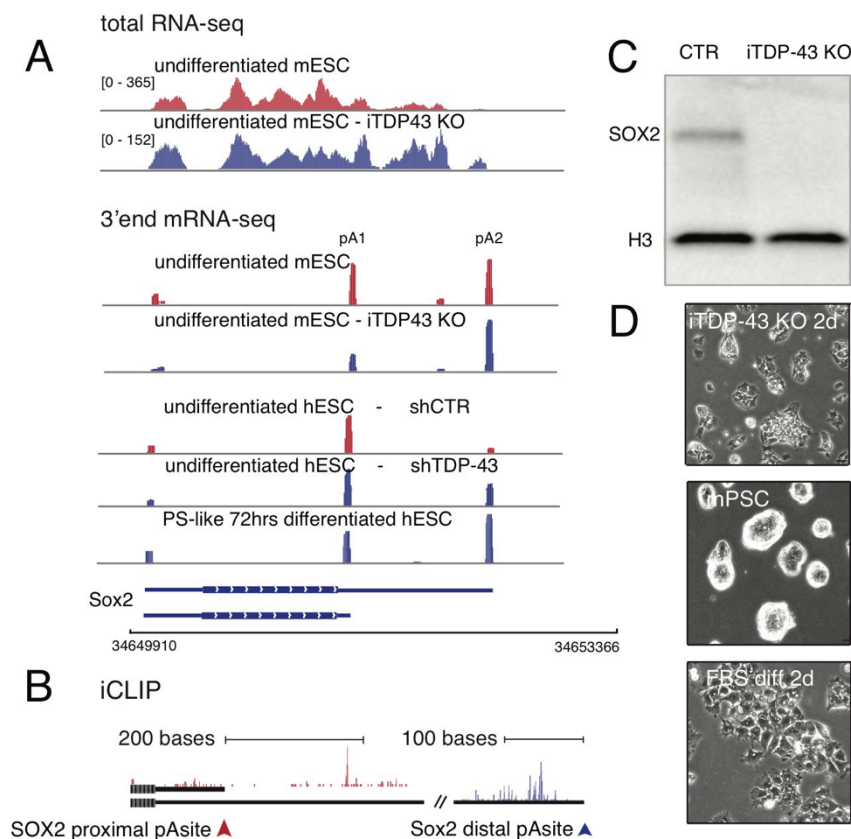


Fig. 31: TDP-43 regulates SOX2 via APA.

(A) Representative diagrams showing the relative frequencies of the PASs in *Sox2/SOX2*

and the 3' mRNA isoforms (mESCs FPKM levels in brackets; n=3) upon tamoxifen-induced *Tdp-43* KO in mESCs (iTDPKO, fig. S5D-F), and *TDP-43* KD, 72hrs of PS differentiation by CHIR99021 treatment, and control human ESCs. **(B)** Representative TDP-43-binding clusters (analyzed by iCLIP, Fig. S4) surrounding the proximal and distal PASs of *Sox2*. **(C)** A Western blot analysis of SOX2 levels in iTDPKO mESCs: control cells and following tamoxifen treatment to induce KO and 48hrs of spontaneous differentiation. **(D)** Representative photomicrographs of iTDPKO mESCs maintained in naive media, treated with FBS to induce initial differentiation, or following tamoxifen treatment in naive mESCs media 48hrs.

To further study the effect of TDP-43 on *SOX2* expression, we used a minigene reporter containing the 3' UTR of *SOX2* located downstream of the *eGFP* coding sequence (Fig. 32A). Flow-cytometry confirmed that *TDP-43* overexpression in HEK 293T cells increases the GFP signal (Fig. 32B). This indicates that TDP-43 enhances the pluripotency circuitry through conserved direct regulation of PAS choice in *SOX2* mRNA, which is required for efficient translation. To show that APA switch governs GFP expression via regulation of the 3'UTR of *SOX2*, I mutated the proximal PAS. This reverted the effect of TDP-43 (Fig. 32C). Next, I speculated that the mechanism governing down-regulation of the long *SOX2* 3'UTR isoform is mediated by microRNAs, because I conducted Targetscan analysis (Lewis et al., 2005), which showed a putative *miR-21* target site in the long isoform. Consistent with this hypothesis is the observation that *miR-21* is upregulated in exit of pluripotency (Singh et al., 2015). I therefore mutated the *miR-21* binding site in the reporter construct, and found that this leads to a loss of response to *TDP-43* overexpression (Fig. 32D). This indicates that TDP-43 decrease upon early differentiation causes *SOX2* mRNA lengthening, which coincides with *miR-21* increase (Singh et al., 2015), leading to a decrease in *SOX2* transcript. Taken together, this confirms that TDP-43 is a conserved direct regulator of global APA patterns safeguarding pluripotency circuitry.

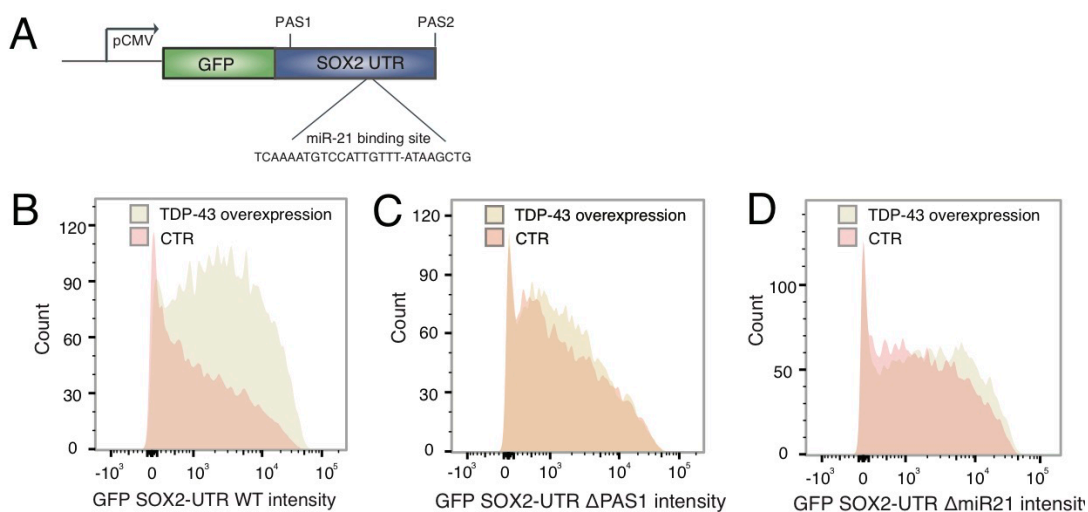


Fig 32: TDP-43 coordinates translational regulation of Sox2 through APA.

(A) A scheme of the eGFP-SOX2 3' UTR reporter minigene showing the positions of the proximal and distal PASs and of the miR-21 binding site. (B-D) Representative flow cytometry analyses of HEK293T cells harboring a doxycycline inducible cassette of TDP-43 (Budini et al., 2015) transfected with miR-21 and the respective GFP-SOX2 3' UTR constructs: unmodified (B), with deletion of the proximal PAS (C) or with deletion of the miR-21 binding site (D) in the presence or absence of doxycycline.

GO-term analysis of the APA transcripts that are regulated by TDP-43 in h/mESCs demonstrates that this factor regulates primarily genes implicated in post-transcriptional processes (Fig. 33). Consistent with this are findings that post-transcriptional regulation governs the production of the long non-coding RNA (lncRNA) *NEAT1*, which serves as a hub for formation of paraspeckles, and that undifferentiated h/hESCs lack paraspeckles. Because *NEAT1* has two isoforms, short polyadenylated and long non-polyadenylated, coined respectively, *v1* and *v2*, and only the latter isoform serves as the scaffold paraspeckles (Naganuma et al., 2012). The decrease of TDP-43 expression level upon dissolution of pluripotency coincides with the *NEAT1* isoform switch. Thus, I hypothesized that TDP-43 regulates the formation of paraspeckles upon exit from pluripotency. Furthermore, because TDP-43 has been shown to localize to paraspeckles, I speculated that these nuclear domains that are thought to coordinate post-transcriptional processes by association with RBPs (Naganuma et al., 2012; Nishimoto et al., 2013; West et al., 2016), contribute to a TDP-43 down-regulation during exit from pluripotency.

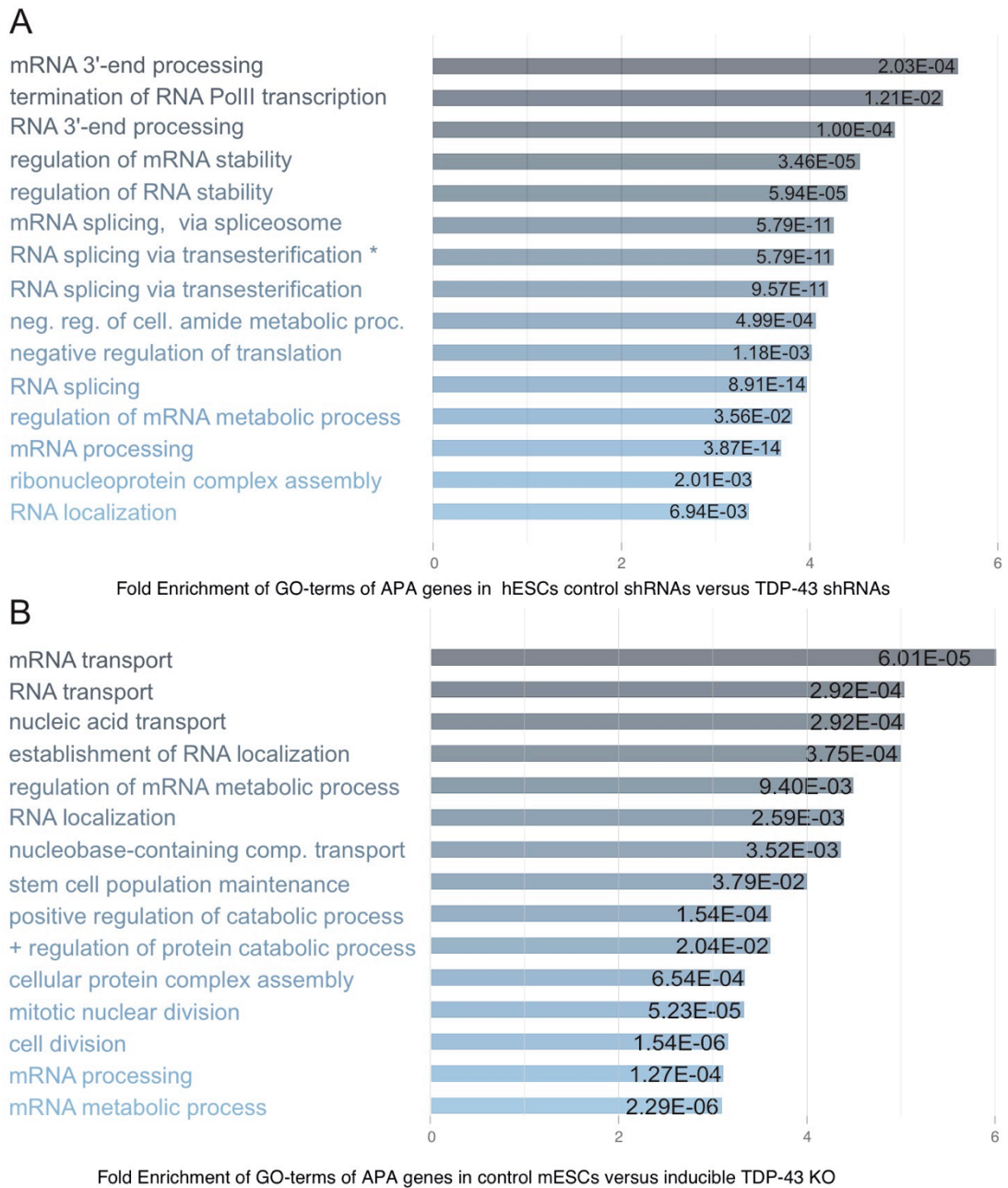


Fig. 33. TDP-43 preferentially regulates APA of genes involved in post-transcriptional processes, and in particular other RBPs that are involved in 3' end mRNA processing, both in (A) hESCs and (B) mESCs. p-values are given in tissue categories. GO terms analysis by gene ontology.

3.2.4 TDP-43 governs formation of paraspeckles upon exit of pluripotency

Paraspeckles have been reported to exist in various cell types with the exception of ESCs (Chen and Carmichael, 2009), but the mechanisms controlling paraspeckles formation in differentiation have not been discovered. It has been reported that one of the RNAs with most abundant TDP-43 binding is *NEAT1* (Tollervey et al., 2011b). By analyzing the TDP-43 iCLIP of h/m*NEAT1* I found that TDP-43 binds to a vertebrate-conserved cluster of UG-rich motifs, present 100-200 nt upstream of the PAS that produces the short *Neat1v1* isoform (Fig. 34A). According to the pattern defined by the RNA map (Fig. 30, 34B), such distal binding pattern suggests that TDP-43 enhances the production of *Neat1v1* in ESCs, thereby preventing paraspeckles polymerization in these cells. In accordance to this prediction, I found that KO of *Tdp-43* in mESCs leads to expression of *Neat1v2* (Fig. 34C), and formation of paraspeckles even in undifferentiated mESCs (Fig. 34D-E; 8-fold increase, χ^2 test, $p < 6e10^{-6}$). Conversely, overexpression of *Tdp-43-eGFP* fusion protein in mESCs following 72hrs of spontaneous differentiation prevents the formation of paraspeckles (Fig. 34F).

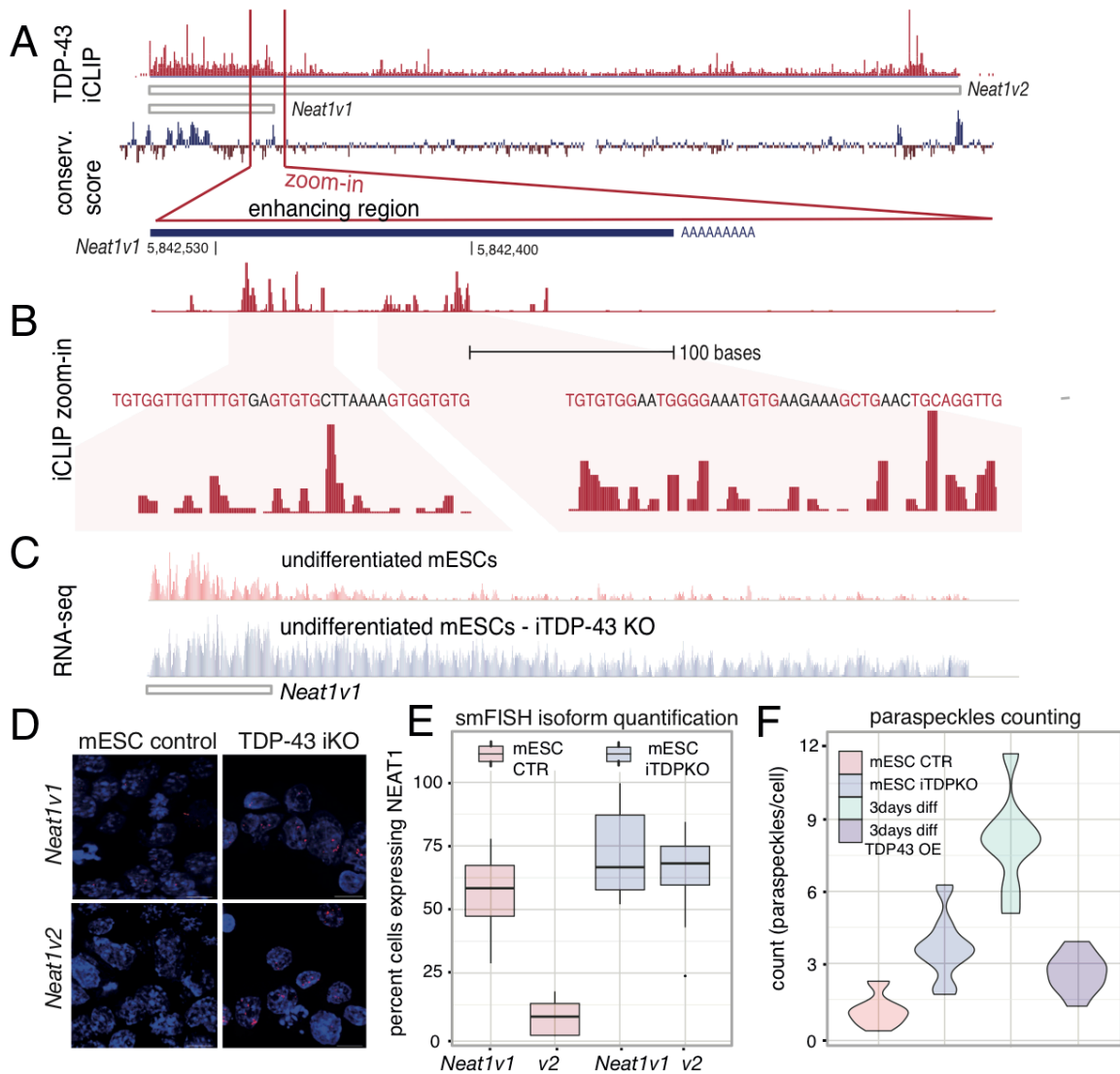


Fig. 34. TDP-43 regulates paraspeckle formation through a conserved PAS choice mechanism.

(A) Cross-linked positions (red) of TDP-43 in mESCs (n=8) plotted on a diagram of *Neat1v1* and *v2* isoforms. Blue bars correspond to regions of high sequence conservation among vertebrates. (B) A close-up of the area upstream of *Neat1v1* PAS, highlighting the GU-rich motifs that correspond to the TDP-43-bound sites. (C) Representative RNA-Seq data showing *Neat1* mapped reads following tamoxifen induced KO of Tdp-43 and of untreated control cells (iTDPKO line, fig. S5D). (D) Representative maximum projection photomicrographs of *Neat1v1* and *v2* isoforms analyzed by single molecule fluorescent in situ hybridization (smFISH, fig. S7). Red: *Neat1v1* and *v2* probes; blue: DAPI. (E) Quantification of paraspeckles (number of *Neat1v1* and *v2* foci/cell) in mESCs 72hrs following induction of KO of Tdp-43 by tamoxifen treatment and in untreated mESCs. (F) Paraspeckle count in control mESCs, 72hrs following induction of KO by tamoxifen, in spontaneously differentiated mESCs by 72hrs FBS treatment, and in differentiated cells that were prepared by the same treatment with the overexpression of *Tdp-43-GFP*. >200 cells were analyzed per group in E-F. T-test, $p < 0.001$.

To confirm evolutionary conserved action of TDP-43 in regulating paraspeckles via APA, I repeated the experiments in human ESCs. This confirmed that hESC don't efficiently form paraspeckles (Fig 35A-B), while

TDP-43 KD lead to expression of *Neat1v2* and paraspeckles even in undifferentiated cells (Fig 35A-B). I thus concluded that TDP-43 maintains the production of *NEAT1v1* in ESCs, and that the decrease of TDP-43 upon differentiation stimulates the production of *Neat1v2* and consequently the formation of paraspeckles.

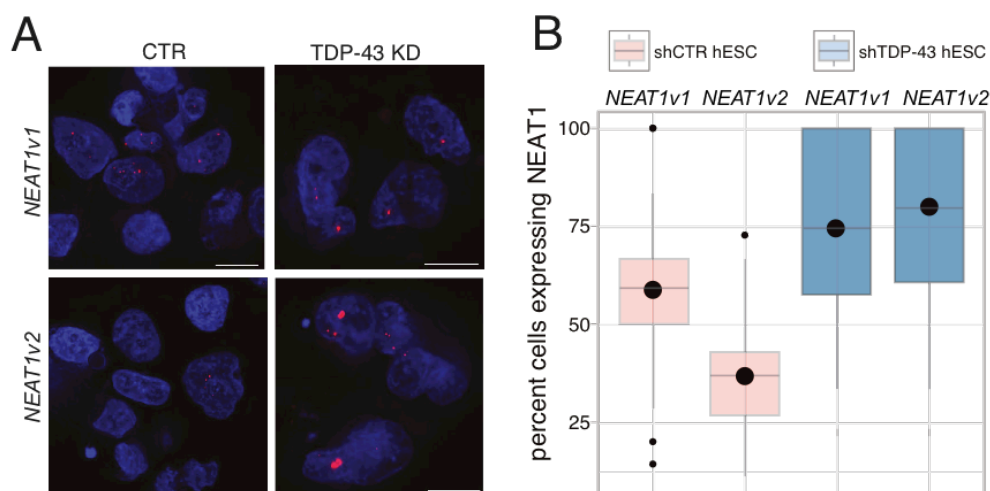


Fig 35: TDP-43 regulates paraspeckle formation in hESCs.

(A) Representative maximum projection photomicrographs of NEAT1v1 and v2 isoforms (red) in spontaneously differentiated hESCs treated for 72 hrs with a medium supplemented with knockout serum replacement (KSR) without bFGF, comparing control and cells following TDP-43 KD. Blue: DAPI. **(B)** Quantification of the frequency of cells exhibiting paraspeckle foci in A; ~100 cells analyzed per group.

The nature of the stimuli that induce the formation of paraspeckles during ESC differentiation are unknown, as is the timing of this and the lineage direction in which it takes place. To address these questions I conducted RNA-Seq and counted paraspeckles in hESCs prompted to differentiate to diverse fates. I found that *NEAT1* is amongst the 18 up-regulated transcripts that are common to 24hrs treatment by purified Wnt3a protein, a small-molecule inhibitor of GSK3 α/β , or BMP4 (Fig. 36A). Furthermore, the paraspeckle frequency and kinetics in early trophoblast, neuroectoderm, and mesendoderm hESC progeny was similar (Fig. 36B-D).

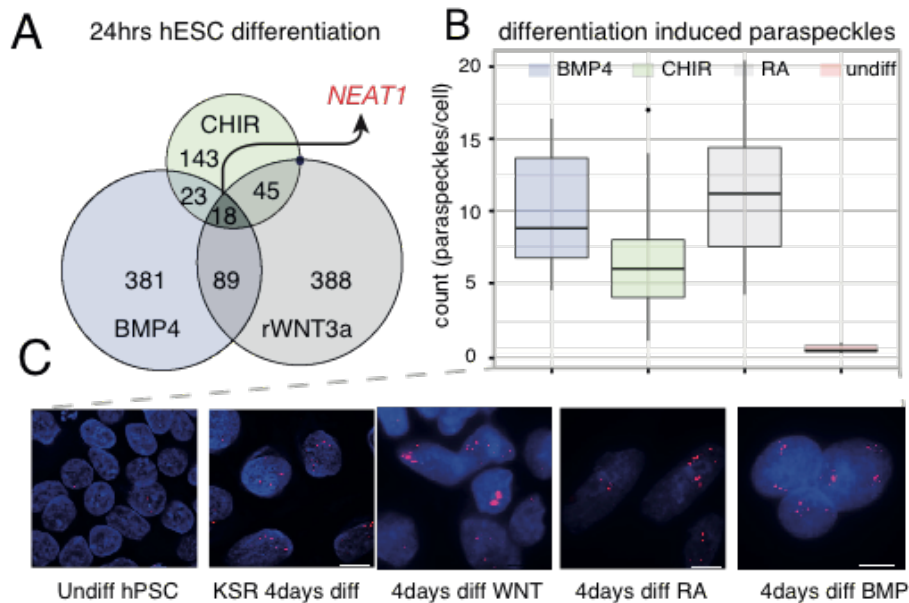


Fig 36: Kinetics of paraspeckle formation in early hESC differentiation.

(A) A Venn diagram depicting differentially expressed transcripts in hESCs exposed to the indicated differentiation stimuli for 24hrs relative to untreated cells (n=2 per condition). (B) Paraspeckle count in hESCs as a function of the differentiation treatment, 72hrs BMP4, CHIR99021 and retinoic acid (RA), promoting trophoblast, mesendoderm and neuroectoderm, fates, respectively, and in control untreated cells (~250 cells analyzed per group, t-test, $p < 0.001$). (C) Representative maximum projection microphotographs of differentiation-stimulated hESCs (conditions as in H, 72hrs). Red: *NEAT1v2* probe; blue: DAPI.

Next, I analyzed whether the production of *NEAT1* is post-transcriptionally regulated in ESCs. I first generated CRISPR-Cas9 modified h/mESCs harboring targeted deletions of regulatory regions surrounding the proximal PAS of *Neat1/NEAT1* (Fig. 37A-C), and confirmed that both h/m *NEAT1 Δ PAS* and *Neat1 Δ PAS* ESC lines efficiently transcribe the full length *v2* (Fig. 37D-E) and form paraspeckles already in pluripotency conditions I mouse and human ESCs. I found that deletion of internal poly(A) site within *NEAT1/Neat1* leads to a similar effect on *NEAT1* processing as TDP-43 depletion (Fig. 37F-I). I therefore concluded that the production of the long *NEAT1v2* isoform is inhibited if the internal polyadenylation site (polyA-site) within *NEAT1* pre-mRNA is used, which produces the short *NEAT1v1* isoform.

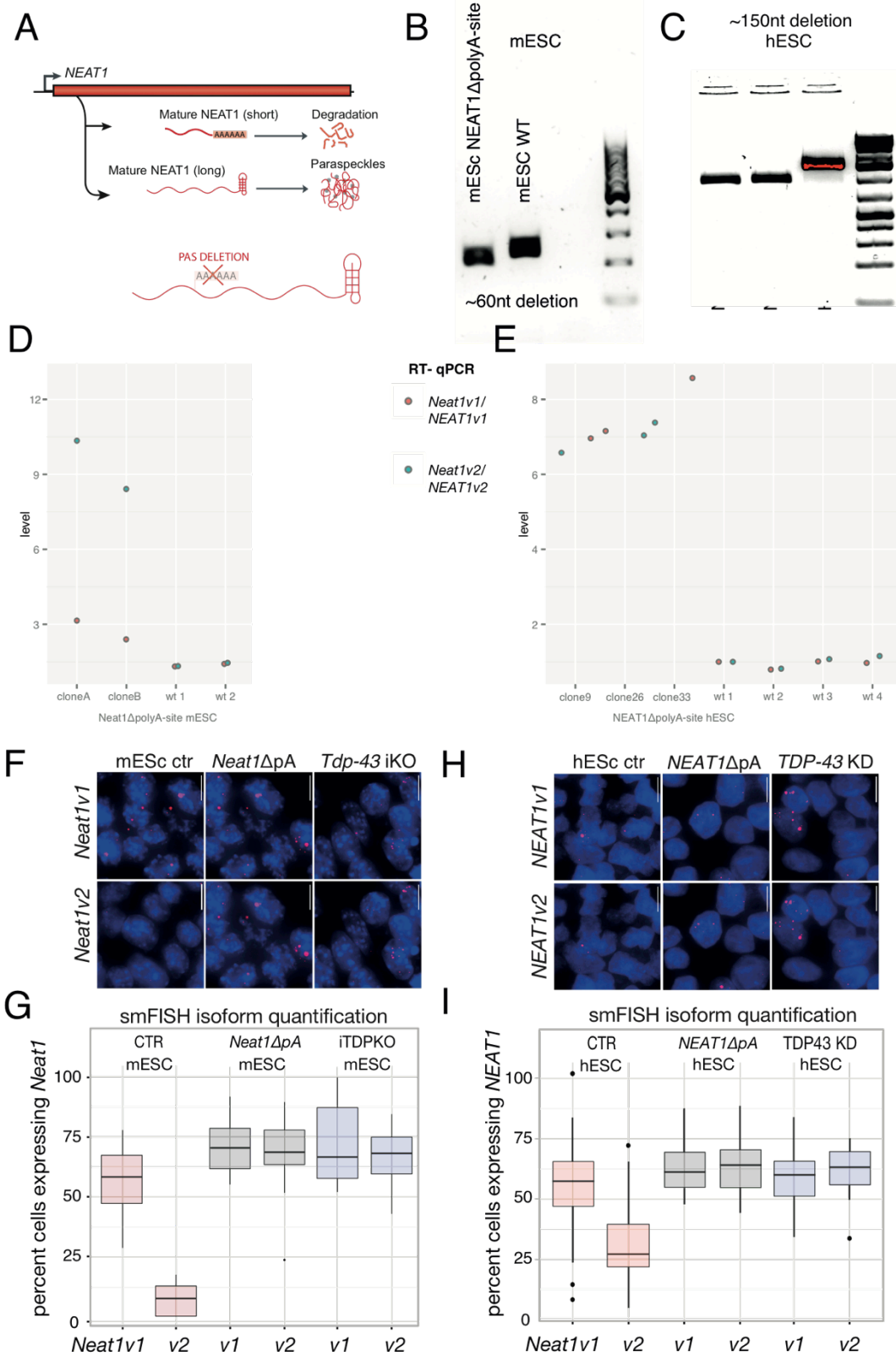


Fig 37: CRISPR/Cas9 deletion of internal poly(A) site of *Neat1/NEAT1* in mouse and human PSCs.

(A) Schematic overview of CRISPR-Cas9 generation of paraspeckles-forming full length *Neat1/NEAT1*ΔPAS m/hESCs (B-C) Gel electrophoresis of creening PCRs confirming

deletions of internal pA site in mouse **(B)** and human **(C)** *Neat1/NEAT1ΔpA* generated ESC lines. (D-E) RT-qPCR quantification of *Neat1/NEAT1* v1 and v2 in *Neat1/NEAT1ΔpA* and parental mouse (D) and human (E) ESCs. **(F-I)** Representative maximum projection photomicrographs of *Neat1v1/v2* **(F)** and of *NEAT1v1/v2* **(H)** isoforms analyzed by single molecule fluorescent in situ hybridization (smFISH) quantified in single cells **(G and I)**, respectively). **(F-H)** Undifferentiated mESCs, *Neat1ΔpA* mESCs, and mESCs 72 hr following induction of Tdp-43 KO by tamoxifen treatment (iTDPKO). **(H-I)** Undifferentiated hESCs, *NEAT1ΔpA* hESCs, and hESCs following si*TDP-43* mediated KD (>200 cells were analyzed per group). Red: *Neat1/NEAT1v1* and v2 probes; blue: DAPI.

To confirm that *NEAT1* is post-transcriptionally repressed in pluripotent stage I analyzed the nascent RNAs in hESCs using pulse labeling and sequencing (Michel et al., 2016). While *NEAT1v1* was not detected in total RNA-Seq, it is detected by the nascent RNA-Seq, indicating that this isoform has a short half-life in hESCs (Fig. 38A). This confirms that *NEAT1* is transcribed in PSCs, but is post-transcriptionally repressed and explains why paraspeckles are absent from undifferentiated PSCs. Furthermore, I analyzed the intranuclear localization of *NEAT1* using a subcellular fractionation protocol that enables discrimination of nascent chromatin-associated RNAs from free nucleoplasmic RNAs. By applying this protocol in conjunction to RNA-Seq, I confirmed that *NEAT1v2* is chromatin-bound (Mao et al., 2011) as opposed to *NEAT1v1* which is enriched in the nucleoplasm (Fig. 38). I conclude that the expression of TDP-43 in ESCs maintains production of the unstable *NEAT1v1* isoform, and its decrease upon early differentiation underlies an isoform switch to chromatin bound stable *NEAT1v2* and formation of paraspeckles.

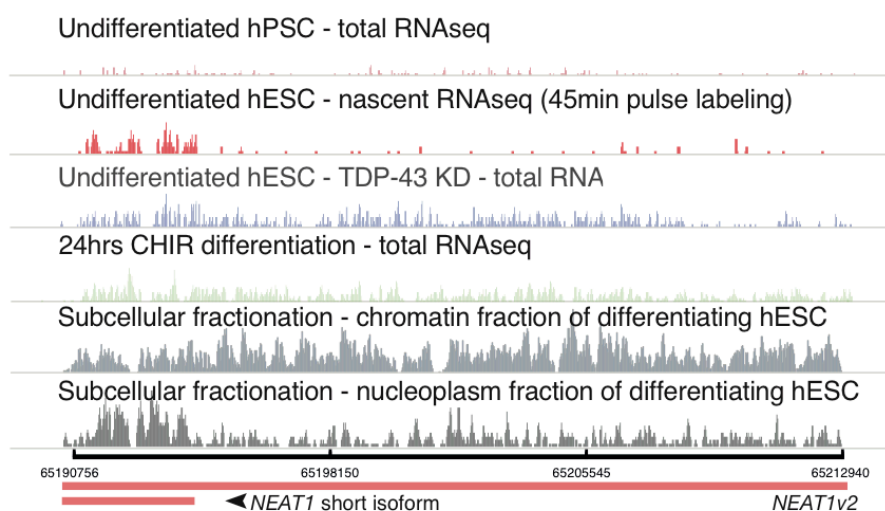


Fig 38: NEAT1 isoforms and intra-nuclear localization.

Representative outlines of *NEAT1* RNA-Seq mapped reads in the ~20kb gene locus; the two isoforms are indicated. The samples are undifferentiated and 24hrs CHIR99021-treated hESCs, nascent RNA of undifferentiated hESCs, hESCs 48hrs following KD of *TDP-43* and chromatin- and nucleoplasm-enriched fractions from 48hrs BMP4-treated hESCs, all as indicated.

3.2.5 Composition of paraspeckles

The cellular functions of paraspeckles have yet to be clarified. To assess if in a developmental context paraspeckles can modulate post-transcriptional processing by relocating RBPs from their other binding sites in mRNAs, I conducted mRNA-RBP occupancy analysis, coined RBPome, to assess how induction of paraspeckles influences binding of RBPs to their mRNA binding sites by using the global mRNA-RBP occupancy analysis (Castello et al., 2012) (Fig. 39).

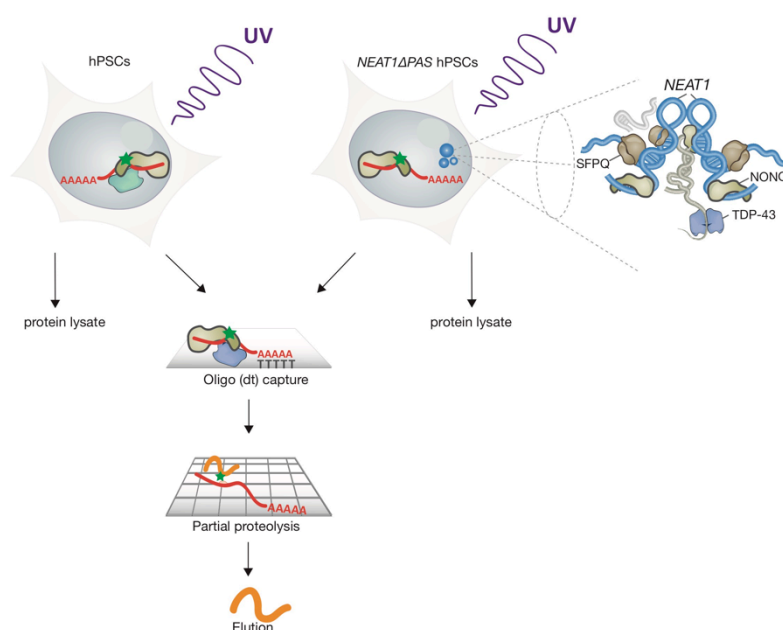


Fig. 39: Schematic overview of NEAT1-RBPome strategy.

This approach uncovers mRNA binding rearrangements of RBPs via paraspeckle associations.

For that I compared the mRBPome of *NEAT1Δ1APo* hESCs and control cells. While the total abundance of RBPs in the mRNA-RBP occupancy assay remained unaltered (Fig. 40A), the induction of *NEAT1v2* in *NEAT1ΔpA* cells led to decreased mRNA-RBP occupancy of known paraspeckle proteins, including NONO, SFPQ and TDP-43, but not of CPEB1 which does not localizes to paraspeckles (Fig. 40C). In addition by coupling RBPome to LC-MS, we noticed that the mRNA occupancy of TDP-43, which has been shown to localize to paraspeckles (Naganuma and Hirose, 2013; West et al., 2016), was reduced to a similar extent as NONO/SFPQ, which is of additional interest since *NEAT1* is one of the top RNA targets of TDP-43 in somatic cells (Tollervey et al., 2011b). This indicates that *NEAT1v2* has a

capacity to relocalize a significant portion of TDP-43 from mRNAs, which likely influences its regulatory activity.

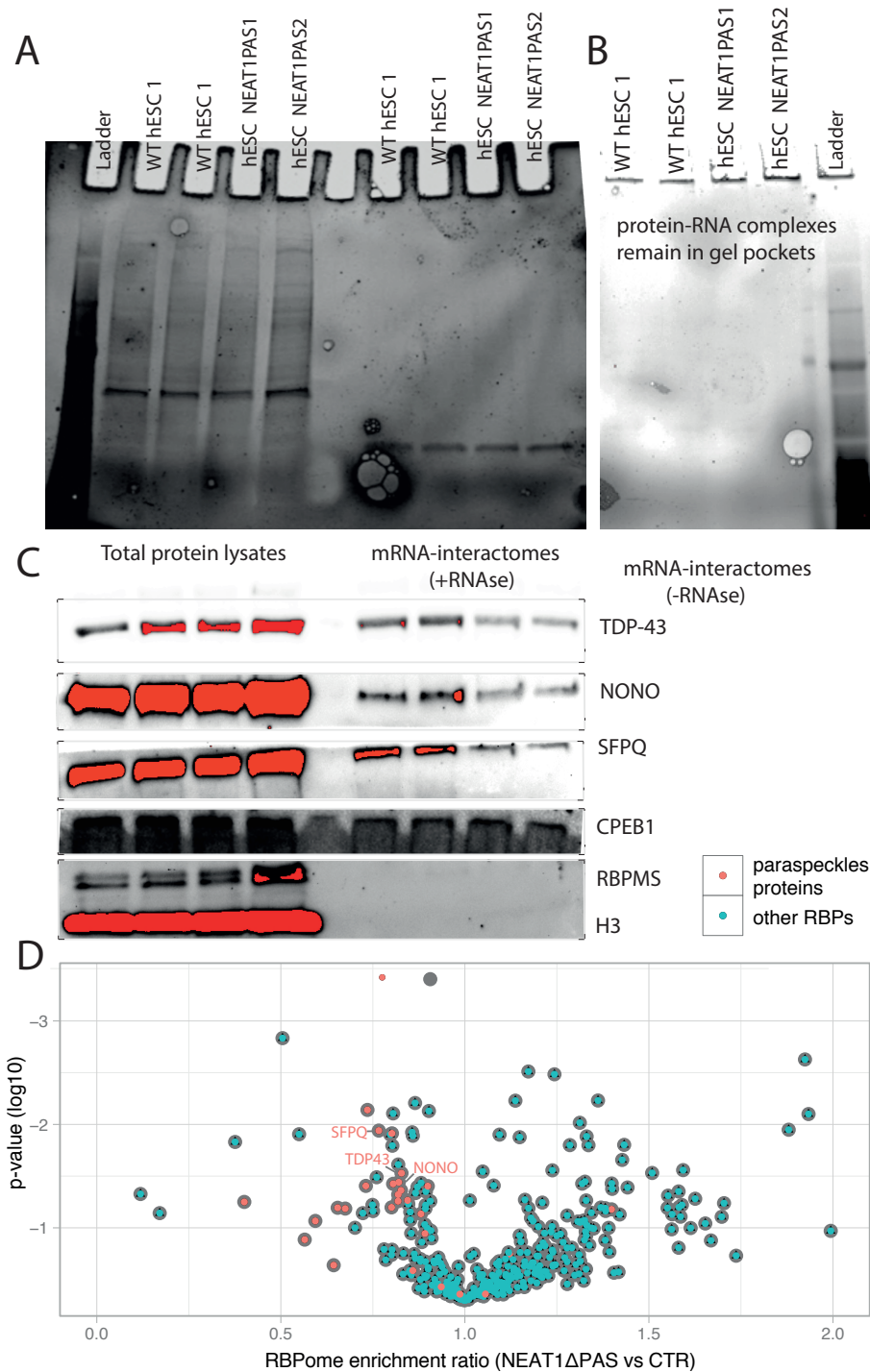


Fig. 40: Paraspeckle-mediated depletion of RBPs.

(A-B) A gel migration assay of covalent RNP complexes (Fig. 39) treated (A) or undigested (B) with RNase I. The specific pulldown of RNP complexes is confirmed because proteins migration is apparent only following RNase I treatment, while undigested RNPs bound to mRNAs remain in gel pockets. (C) Western blot analysis of TDP-43, NONO and SFPQ using crosslinked (UV-C irradiated) whole cell lysates derived from control and *NEAT1* Δ pPAS human ESCs. H3 serves as loading controls. (D) A Volcano plot exhibiting changes in RBP occupancy in the presence of *NEAT1v2* in undifferentiated hESCs. Annotated paraspeckles proteins are labeled red.

3.2.6 Feedback cross-regulation between TDP-43 and paraspeckles governs PSC differentiation and reprogramming

In order to understand how the mutual regulation between paraspeckles and TDP-43 may promote distinct cellular identities in the interface of pluripotency and termination of this state, I analyzed their respective effects on somatic cell reprogramming and differentiation kinetics. I first compared the differentiation kinetics of induced pluripotent stem cells (iPSCs) derived from *Neat1*-deficient mouse embryonic fibroblasts (Nakagawa et al., 2011) and control wt cells. Importantly, in spontaneous differentiation conditions the *Neat1*^{-/-} miPSCs retain the stereotypical PSC colony morphology and expression of the pluripotency cell surface marker SSEA-1 for a longer time compared to control hPSCs (Fig 41A-B). Moreover, *Neat1*^{-/-} miPSCs display delayed expression of the early differentiation marker SSEA-4 (Fig. 41C), meaning that pluripotency exit is slower in the absence of full length *Neat1*. Surprisingly we didn't observe an overt effect on key pluripotency factors, albeit the levels of TGF- β superfamily core factors *Lefty1* and *Gdf3*, but not of other canonical pluripotency factors in *Neat1*^{-/-} cells were higher compared to control wt mESCs (Fig. 41C-F). In accordance, expression of *Lefty1* is retained in 72hrs spontaneously differentiated *Neat1*^{-/-} miPSCs (Fig. 41G). As expected (Kim et al., 2014), loss of LEFTY1 coincides with higher SMAD2 phosphorylation (Fig 41H), which promotes mesendoderm differentiation potential without influencing the self-renewal of PSCs (Lee et al., 2011). This indicates that paraspeckle formation poises exit from pluripotency by antagonizing nodal signaling through phosphorylation of SMAD proteins.

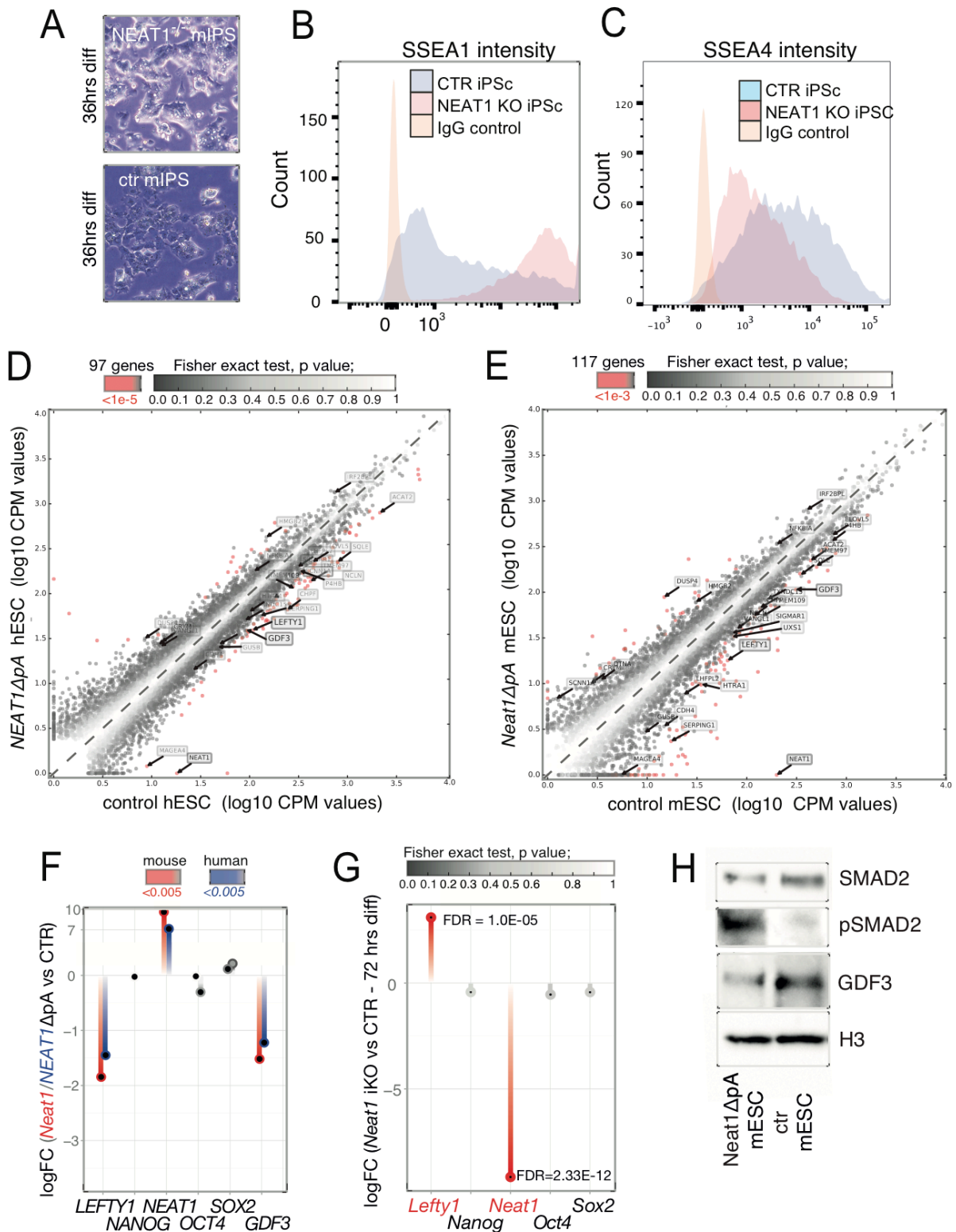


Fig 41: Paraspeckles poise PSC differentiation.

(A) Representative photomicrographs of 36hr *Neat1*^{-/-} and control wt mIPSCs treated with spontaneous differentiation medium (+FBS -LIF). (B-C) Representative flow-cytometry analyses of 48hr *Neat1*^{-/-} and control wt mIPSCs stained with SSEA-1 (B) or SSEA-4 (C), culture conditions as in (A). (D-E) Representative scatter plots showing transcripts detected by RNA-Seq in *NEAT1/Neat1ΔpA* h/mESCs (mESC= 2iLIF, hESCs=mTeSR; see methods; n=3/condition). Differentially expressed human genes with adjusted p-value < 1e-5 (Fishers exact test, false discovery rate <0.001) in (D), and mouse genes with adjusted p-value < 1e-3 (Fishers exact test, false discovery rate <0.05) in (E) are labeled red. (F) RNA-Seq

observed changes of transcripts in *Neat1/NEAT1ΔpA* ESCs compared to undifferentiated m/hESCs. Differentially expressed genes are labeled red in mESCs and blue in hESCs (n=4 per condition). **(G)** RNA-Seq observed changes of transcripts in *Neat1^{-/-}* KO miPSCs compared to isogenic control cells (n=2 per condition), both upon 3days of spontaneous differentiation. **(H)** A Western blot of pSMAD2, SMAD2, GDF3 and Histone H3 in *Neat1ΔpA* mESCs and control cells supports the notion that paraspeckles antagonize TGFbeta/activin/nodal signaling.

Next, we analyzed the role of TDP-43 in the context of reprogramming using human secondary fibroblasts harboring a doxycycline-inducible OKSM (*OCT4*, *KLF4*, *SOX2*, *C-MYC*) gene cassette (Cacchiarelli et al., 2015) (Fig. 42A). We found that overexpression of *TDP-43* results in a nearly twofold increase in the number of colonies expressing the pluripotency marker alkaline phosphatase (AP) and TRA1-60 (Fig. 42B-D), and confirmed that the resulting colonies had a stereotypical human induced pluripotent stem cell (hiPSC) morphology. In accordance, we noted a threefold increase in the number of cells expressing the pluripotency marker TRA-1-60 (Fig. 42E), and that >90% of the cells overexpressing *TDP-43* expressed endogenous *OCT4* at an earlier reprogramming stage compared to control cells (Fig. 42F).

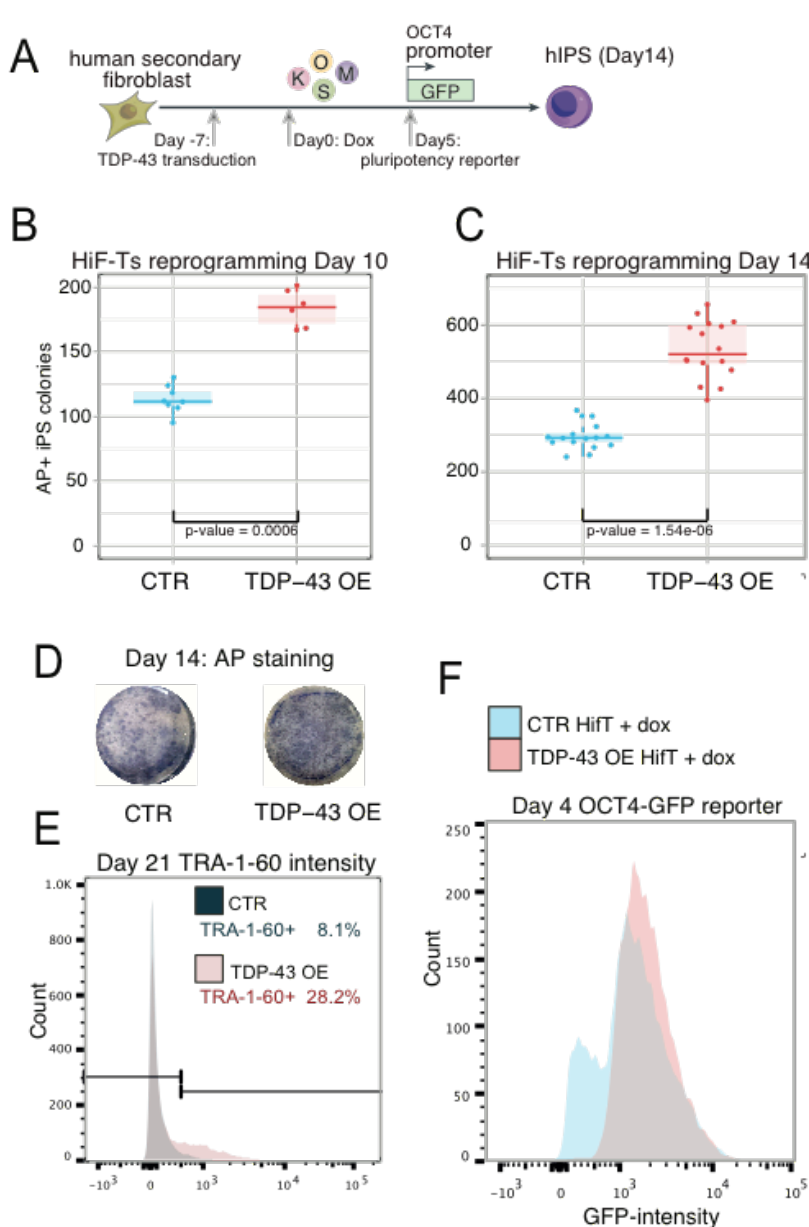


Fig. 42. TDP-43 enhances somatic cell reprogramming.

(A) A schematic representation of the reprogramming strategy that relies on hiF-T secondary fibroblasts harboring inducible reprogramming factors (iOKMS) controlled by the reverse doxycycline transactivator. (B-C) Analysis of the number of AP-positive nascent hiPSC colonies, at days 10 (B) and 14 (C) of reprogramming (with total 24 technical replicates per condition). (D) AP staining of representative hiF-T reprogrammed cultures. TDP-43 overexpression as indicated, control - empty vector. (E) Representative flow cytometry analysis using Tra-1-60 of HiF-T cells at reprogramming day 21 (n=3). (F) A representative flow cytometry analysis of *OCT4-GFP* reporter on day 4 following doxycycline treatment of HiF-T fibroblasts.

To validate the role of TDP-43 in promotion of the pluripotent state, we transiently overexpressed *TDP-43* during the initial phase of reprogramming of primary adult human fibroblasts (Fig. 43A), and found that this leads to a 5-fold increase in the number of iPSC colonies (Fig. 43B). Moreover inducible overexpression of *Tdp-43* in *mESCs*, which inhibits paraspeckle formation upon differentiation (Fig. 34F), leads to an effect of delay in differentiation (Fig. 43C) that is similar to the KO of *Neat1* (Fig 41C). Taken together, we find that the mutual negative feedback between paraspeckles and TDP-43 promotes distinct cellular identities; paraspeckles promote differentiation, while TDP-43 inhibits differentiation and lowers the barrier to somatic cell

reprogramming.

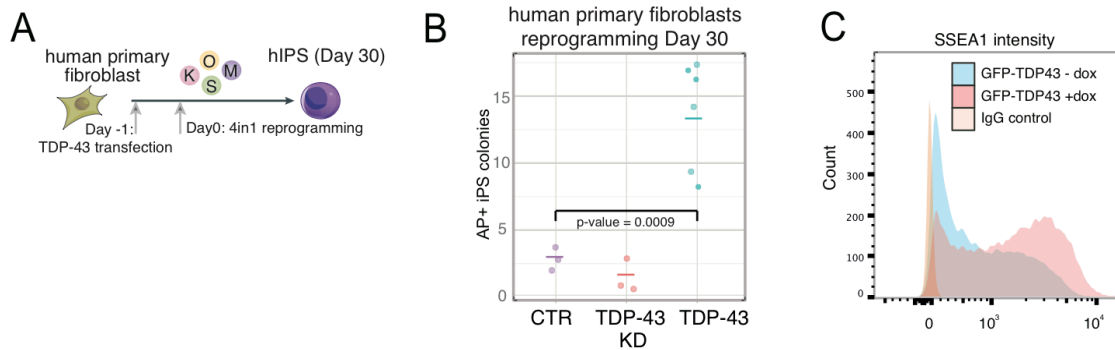


Fig. 43. TDP-43 enhances somatic cell reprogramming and delays differentiation of PSCs: (A). A schematic representation of the reprogramming strategy for primary fibroblasts transfected with the OKMS 4-in-1 episomal vector (Diecke et al., 2015) (B) Analysis of the number of AP positive hiPSC colonies, at reprogramming day 30. (C) Representative flow-cytometry analyses of mESCs undergoing differentiation (48hr treated with +FBS -LIF medium) in the presence of induced *Tdp-43* (+dox) and of control cells (-dox).

Taken together, we discover how post-transcriptional regulation concomitantly dissolves PSC self-renewal and promotes differentiation in a lineage independent fashion. This is a primary proof of uncoupling of exit from pluripotency and fate commitment (Fig. 44) through mutual feedback regulation between paraspeckles and TDP-43, which promotes the efficient transition between pluripotent and differentiated cellular fates.

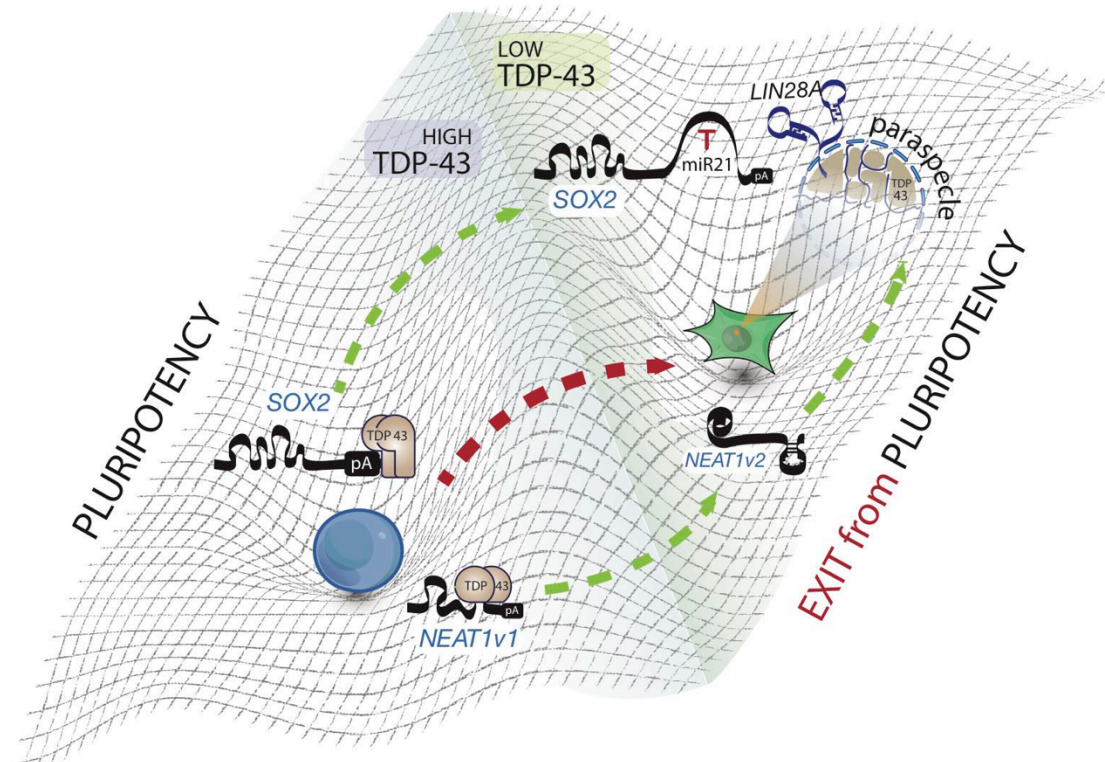


Fig. 44. A model depicting TDP-43- and paraspeckle-mediated mutually exclusive stabilization of pluripotency and differentiation states. The model outlines the possibility that pluripotency mRNAs, including *LIN28A* and *REX1*, and the TDP-43 protein promote exit from pluripotency by becoming sequestered in paraspeckles. Consistent with this model, our data confirmed the chromatin enrichment of the *NEAT1v2* isoform in early-differentiated hESCs, and conversely the nucleoplasmic localization and the rapid degradation of the *NEAT1v1* isoform in undifferentiated ESCs (Fig. 38).

4 DISCUSSION

4.1 DEVELOPMENTAL ONTOGENETIC RBPOME UNCOVERS DYNAMIC RNA BINDING PROTEINS

Numerous studies have recently clarified that gene expression in eukaryotes involves tight regulation on the RNA level ensuring that correct complement of mRNAs and protein quantities are produced in cells (Schwanhausser et al., 2011). It is an established theme that the copy number of a given transcript in the cell can only partially explain the quantity of the respective protein, but little is known to what degree the amount of protein explains its functionality. In particular, the open question remains to what degree are RBPs bound to RNA targets and in this way functional. I used the mRNA-RBP occupancy method for this purpose to assess what proportion of the RBP is mRNA bound at certain cell stage and to address a central question whether we can determine level of RBP functionality in given cell type and if there is a switch in functionality between different cell states. Stem cells are ideal biological system to study functionality of cohorts of RBPs that are specifically expressed in continuum of pure pluripotency cultures of naïve-limes-epiblast PSCs, and sorted homogeneous progenitor populations that emerge from epiblast PSCs. This system also presents a foundation for functional analysis of the leading candidates for the respective developmental pre-mRNA processing changes. Intriguingly, upon generating first global analysis of changes in RBP functionality, I didn't observe significant changes in abundances of splicing factors in early pluripotency transitioning and commitment despite global changes in post-transcriptional landscape. This hints that possibly dynamic nature of RBPs is sufficient to remodel developmental alternative splicing and polyadenylation landscape. This called for in depth analysis of altered RBP functionality during stem cell transitions without changes in RBP abundance.

A global mRNA-RBP occupancy has already been conducted with mESCs (Kwon et al., 2013), however only in steady state, providing mainly static information on RBPome composition. Hence, little is known about the developmental dynamics of changes in the RBPome composition, whether these changes are the result of signaling pathway activity, whether the RBPs are developmentally dynamic in its function and whether we can determine

level of RBP functionality in given cell type. Normalization of RBPome analysis to corresponding protein abundance reveals the extent of RBP functionality to form dynamic mRNPs in different developmental stages, hinting the change in functionality of the protein.

Together, my mRNA-RBP occupancy assay constitutes the first ever definition of RBPome in an ontogenetic system of an unbiased and systematic approach for detection of the entire repertoire of RBPs in early stem cell priming and differentiation. In addition, my work for the first time explains multiple cases of mRNA occupancy changes in relation to pathway activation scenarios thereby presenting dynamic composition of RBPs in early embryonic development. Insight into RBPome changes during early embryonic stem cell development was enabled by further improvement of mRNA-interactome approach resulting in reducing protein background and thereby generation pure set of 676 RBPs. This discovery rate is very high considering that my analysis utilized <5% of the starting material that was used in a previous study, which identified 555 RBPs in mESCs grown in serum and LIF (Kwon et al., 2013), conditions allowing spontaneous differentiation leading to heterogenous population of various cell states (Kolodziejczyk et al., 2015). Moreover, my RBPome data overlapped with majority of an annotated list of RBPs (Gerstberger et al., 2013) (Fig. 19B). By including two technical controls for determination of RBP candidates, I stumbled to computational problem as the background signal in both negative control couldn't be quantified for many identified RBPs. This contributed to more challenging data analysis, but at the same time it also confirms that the generated RBPome indeed represents bona fide RBPs including hundreds of previously known RBPs.

During stem cell commitment and early differentiation I for the first time confirmed high dynamicity of the mRNA-RBP occupancy by comparing changes in the RNA-bound fraction by comparing ratios the total proteome changes. I observed that vast majority of RBPs (>80%) do not exhibit abundance differences in the RBPome relative to their whole cell amount between any stages of ESC commitment and differentiation. These results are in line with my expectations because many RBPs have "house keeping" functions, and are therefore either not expected to change or exhibit changes that are scaled between developmental states. I aimed to shed more light on dynamic RBPs that exhibit changes in the RNA bound fraction

but such that are not correlated with changes of their overall proteome amount between the stages (~5% of all RBPs, called dynamic RBPs).

4.2 SIGNALING INDUCED CHANGES IN RBP DYNAMICITY AND FUNCTION

Data integration of RBPome with proteome enabled insight into new links between RNA processing and early development in a dynamic ontogenetic time course. It has already been proposed that differentiation signals affect diversity of protein function in different conditions by altering its protein interaction partners (Kloet et al., 2016) or posttranslational modifications (Heyd and Lynch, 2010). To my knowledge, this is the first study of the effect of differentiation signals on the landscape and activity of mRNPs. First I observed that dynamic RBPs exhibit enrichment of translation related molecular functions (Fig. 20). Of the RBPs that exhibit enrichment in mRNA binding in the primed state (Fig. 19), we noted in particular LIN28A, which was recently identified as MAP kinase target that efficiently phosphorylate S200 LIN28A to elevate its protein levels (Tsanov et al., 2016). In pluripotent stem cells MAP kinase activity primes PSCs for the pluripotency transition and lineage commitment.

Any posttranslational modification can influence, activity and function and even localization of modified RBP. I postulated that the functions of dynamically bound RBPs are modified and orchestrated via signaling pathways that dictate stem cell commitment and priming. Therefore I assumed that MAPK dependent LIN28A phosphorylation upon PSC priming correlates with changes in LIN28A mRNA binding. I first confirmed that LIN28A phosphorylation elevates the protein stability, but intriguingly results in even 2fold higher mRNA occupancy. LIN28A defines stemness by repressing biogenesis of the let-7 family, but beyond this role, LIN28A is a strong translational regulator of pluripotency and splicing factors (Cho et al., 2012), making it an RBP with profound effect on pluripotency. Despite being generally well-studied protein, studies reported opposing observations on LIN28A subcellular localization. The protein is generally considered to be predominantly cytoplasmic, where it binds pre-let-7 to block its processing and instead induces 3'-end oligouridylation (Balzer and Moss, 2007)(Piskounova et al., 2011). In contrary to established LIN28A cytoplasmic role in let-7 repression, systematic study of early mouse

embryonic development observed accumulation of LIN28A exclusively in the mature nucleolus in mouse preimplantation embryos and partial nucleolar localization also in PSCs where it colocalizes with the nucleolar markers (Vogt et al., 2012).

4.3 IDENTIFYING NOVEL SUBNUCLEAR ROLES OF DYNAMIC RBPS THAT ACT AS STEM CELL DIFFERENTIATORS

As RBPome identified LIN28A as one of most dynamic RBPs, I monitored its localization dynamics during pluripotent stem cell transition and exit. To circumvent limitations of antibody staining and fluorescent fusion-proteins and to cross-validate between the two techniques, I used both CRISPR generated fusion fluorescent proteins and a set of 3 different antibodies against LIN28A to compare and confirm the localization patterns. Both in vivo and in fixed cell cultures I observed extensive LIN28A translocation to cytoplasm that reflects also increased mRNA occupancy following exit of naïve and suggests active regulation of RNA binding of some RBPs.

Whether LIN28A has nuclear roles is to date still unknown. The Daley lab hypothesized that nuclear LIN28 could also directly regulate alternative splicing (Shyh-chang and Daley, 2013), for what I provided first initial evidence (Fig. 23). Additionally, by integrating information that LIN28A is predominantly nucleolar in naïve PSCs, but *let-7* is nevertheless repressed in naïve PSCs (Tsanov et al., 2016), I postulate dual LIN28A-dependent mechanisms of *let-7* repression. I first measured *let7* miRNA levels that were in total not significantly changed between naïve and primed pluripotent stem cells. This hinted even nuclear dependent mechanism of *let-7* repression, since LIN28A is localized to nucleoli in naïve PSCs (Figs. 45A, 23A). To further dissect the mechanism of the nuclear Lin28A-mediated *let-7* processing block I tested whether LIN28A co-localizes with pri-*let7*. Indeed, I observed that primary *let7* transcript is sequestered to nucleoli (Fig 45B), where it cannot be processed by DGCR8 and Drosha, which are excluded from nucleoli (Piskounova et al., 2011).

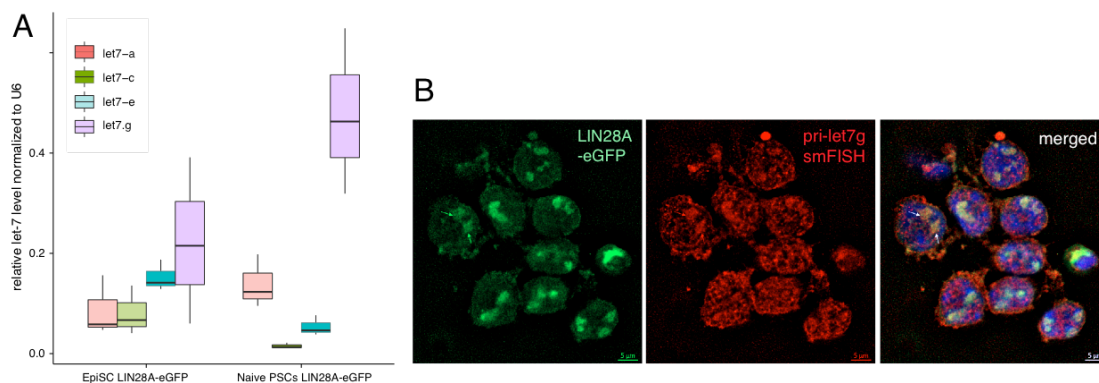


Figure 45: LIN28A represses let-7 biogenesis in the nucleolus of naïve PSCs. (A) Confirmatory RT-qPCR analysis of let-7 miRNA isoforms showing similar expression levels of let-7 isoforms in naïve and primed PSCs (Error bars are SEM of n=3). (B) Representative deconvolved maximum projection photomicrographs taken with smFISH probes used for detecting pri-let7g on LIN28-eGFP mouse naïve PSCs. LIN28 fusion protein and pri-let7g co-localize with each other as indicated with arrows.

Furthermore, LIN28A has nucleolar localization pattern also in non-human primate preimplantation embryos, but is exclusively cytoplasmic protein in non-human primate PSCs (Vogt et al., 2012). I confirmed that LIN28A is exclusively cytoplasmic also in human PSCs, but following stem cell differentiation LIN28A accumulates in the nucleus with preferential localization in the nucleolus (Fig. 23A). Taken together, LIN28A has species-specific localization patterns, which is intriguing for highly conserved pluripotency factor.

Additionally, by simultaneously mutating both LIN28A RNA binding domains (CCHC and CSD) LIN28A exhibits entirely nuclear localisation in cells where it normally acts as cytoplasmic RBP (Balzer and Moss, 2007). These findings suggest RNA-dependent RBP transport. Whether this active and cell state regulated transport is dependent on PTM, RNA binding or formation of specific RNP complexes remains to be elucidated.

It remains to be elucidated whether the protein has nucleolar/cytoplasmic redundant function in repressing let-7 biogenesis. However, specificity of LIN28A is not limited to miRNA molecules as LIN28A binds a myriad of mRNAs to modulate their translation. In this way LIN28A functionally exhibits two distinct molecular mechanism, let-7-dependent (nucleolar and cytoplasmic) and let-7-independent (cytoplasmic). Further in depth mechanistic studies of LIN28A will advance our understanding what is essential LIN28A function for pluripotency maintenance and somatic reprogramming. Furthermore, as already known pluripotency factor LIN28A

was identified as highly dynamic RBP, it is intriguing whether a subset of novel dynamic RBPs has a general role in maintenance of pluripotency and if so, how is their function altered upon pluripotency breakdown.

4.4 DISSECTING THE POST-TRANSCRIPTIONAL MECHANISM OF LINEAGE RESTRICTION BY DYNAMIC RBPS

In further attempt to dissect the post-transcriptional mechanism of lineage restriction, I showed that TDP-43 is an evolutionary conserved global developmental regulator of APA regulatory programs in PSCs and their progenitor progeny (Fig. 14 for mouse and Figs. 26 for human PSCs). TDP-43 decrease takes place during early differentiation of h/mESCs (Fig. 27D-E). To independently identify the primary mechanisms that regulate the APA changes during PS differentiation, we analyzed the sequence motifs around PAS and identified enrichment of GUG, canonical TDP-43 binding site (Fig. 26). Mechanistically TDP-43 has a position dependent bimodal effect on determining PAS locations (Fig. 30), both towards lengthening and shortening of 3' UTRs and thereby it regulates an evolutionary conserved switch in APA of hundreds of transcripts during early differentiation of mouse and human PSCs (Fig. 27A). Amongst others, TDP-43 developmental depletion subsequently leads to APA of *SOX2*, which exposes this core pluripotency factor to miR-mediated decay (Fig. 31). While the effect of TDP-43 on *SOX2* expression can explain why TDP-43 is essential for early development (Sephton et al., 2012), and why it promotes cellular reprogramming and inhibits PSC differentiation, it is likely that additional RNA targets contribute to this phenotype. Important TDP-43 target is *NEAT1* on which TDP-43 promotes production of a short isoform, which is quickly degraded, and thus cannot act as a scaffold for paraspeckles (Fig. 32). The decline of TDP-43 during differentiation ensures that the decrease of *SOX2* coincides with the production of full-length *NEAT1* that forms paraspeckles. These subnuclear aggregates then in turn recruit TDP-43 and relocalise it away from its other RNA targets.

4.5 CROSS-REGULATION BETWEEN TDP-43 AND PARASPECKLES PROMOTES PLURIPOTENCY- DIFFERENTIATION TRANSITION BY FORMATION OF ATTRACTOR FOR PLURIPOTENCY DISSOLUTION

Collectively, my data indicate that paraspeckles functionally oppose TDP-43, and this cross-regulation creates a bimodal network of antagonizing dual stable states that mediate paraspeckle inhibition in pluripotent state and formation of paraspeckles upon differentiation. In this way TDP-43-paraspeckle axis serves to stabilize mutually exclusive states of pluripotency and differentiation. I predict that the components of paraspeckles act in a feed-forward manner (non-linear amplification) to assemble when the components reach a certain threshold, and below the threshold at the same components repress paraspeckle assembly and contribute to stabilization of the pluripotency state. Taken together, in my thesis I uncover how post-transcriptional regulation concomitantly dissolves stem cell self-renewal and promotes differentiation in a lineage independent fashion. This is to my knowledge a primary proof of uncoupling of exit from pluripotency and fate commitment (Fig. 46).

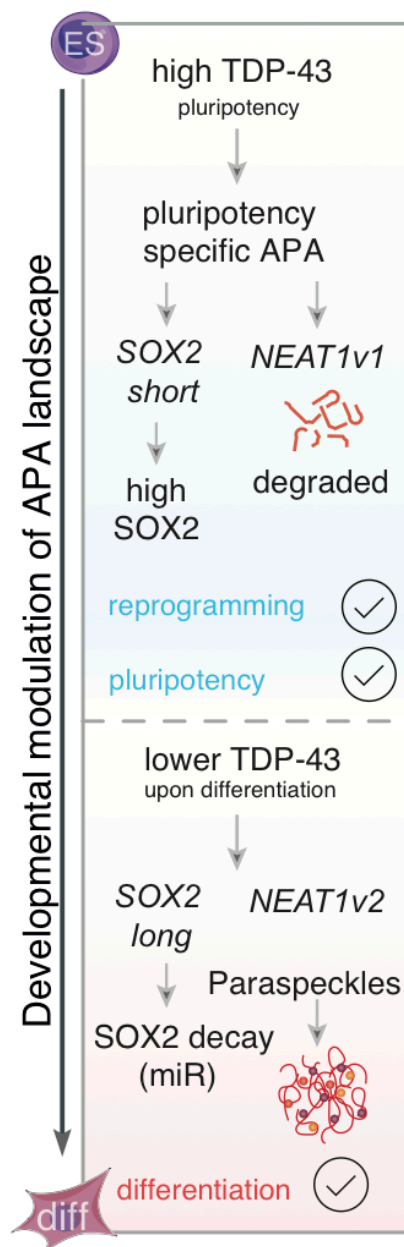


Figure 46: The model for the cross-regulating role of TDP-43 and paraspeckles in the regulation of APA in ESCs, differentiation and reprogramming.

One molecular behavior module that possibly could explain the role of TDP-43 regulated paraspeckles in exit from pluripotency is that they serve to

generate an attractor state for differentiation. Such an attractor state can negate pluripotency in a way that further suppresses pluripotency. The existence of differentiation attractors has been postulated for PSCs (Enver et al., 2009), but to date none have been identified experimentally and I here provide first biological insight into a possible attractor mechanism that leads to reduction of pluripotency through paraspeckle formation.

4.6 DISCOVERING THE PROCESS OF LIQUID PHASE TRANSITION IN CELL FATE TRANSITIONS

Moreover, paraspeckles have been shown to interact with proteins that contain prion-like domains (Hennig et al., 2015). It is likely that the scaffolding activity of *NEAT1* coordinates the tendency of prion-like domains to undergo phase transitions (March et al., 2016), thus leading to the membraneless organelle formation within paraspeckles (Naganuma et al., 2012; Nishimoto et al., 2013; West et al., 2016). This unravels a bimodal equilibrium that regulates cell states via a network of RBPs and a macroscopic RNP granule formed on a lncRNA. It has been proposed that paraspeckles are liquid droplets, implicating the recently discovered process of liquid phase transition in cell fate transitions (March et al., 2016). To the best of my knowledge, PhD thesis discoveries constitute the first proof that such a network is based on the formation of RNA granules.

Furthermore, paraspeckles have previously been linked to diverse etiologies, including cancer and neurodegeneration (Adriaens et al., 2016; Nishimoto et al., 2013). To date, more than 40 proteins are known to associate with paraspeckles (Naganuma et al., 2012), including several key molecular players involved in pathogenesis of ALS and FTL, like TDP-43, FUS, TAF15, EWSR1 and hnRNP-A1 are components of paraspeckles (Shelkovernikova et al., 2014). Furthermore, six paraspeckle proteins are known to cause ALS when mutated. In addition, among known paraspeckle proteins hnRNP-A1, hnRNP-H, hnRNP-K and FUS already been reported to colocalize with $(G_4C_2)_n$ RNA foci, most common ALS hallmark, further strengthening the role of paraspeckles in *C9ORF72* ALS pathology (Cooper-Knock et al., 2014, 2015; Lee et al., 2013; Rossi et al., 2015). To further expand on the role of paraspeckles on neurodegeneration, we in collaborative efforts with Rogelj lab showed that ALS-causing $(G_4C_2)_n$ RNA foci form paraspeckle-like structures, which modulate nuclear

compartmentalization of paraspeckle-bound RNAs. We even show that $(G_4C_2)_n$ RNA foci form paraspeckle-like structures, which may function in similar fashion as paraspeckles and modulate nuclear compartmentalization of paraspeckle-bound RNAs (Darovic et al., in review, data not shown). Our results add additional evidence to the role of paraspeckles in pathogenesis of ALS and FTL, however the exact mechanism linking paraspeckles to disease in *C9ORF72*-positive and negative cases still remains to be discovered.

It would be intriguing to determine sequence requirements of RNAs that can attract paraspeckle proteins and form paraspeckle-like structures. Our results even indicate that $(G_4C_2)_n$ RNA foci might replace NEAT1 and scaffold the formation of paraspeckle-like structures. This is even contradictory with the established dogma, where NEAT1v2 is required for paraspeckle formation and integrity (Naganuma et al, 2012) and raises a question whether NEAT1v2 is truly indispensable for paraspeckle formation. Although this comes a step closer in connection of paraspeckles to neurodegeneration, it still needs to be further elucidated whether nuclear accumulation of RNAs is a critical step in disease mechanism.

4.7 ROLE OF TDP-43/NEAT1 AXIS IN DISEASE

It should also be noted that just like the perturbations in *LIN28/let-7* network, which are manifested in developmental disorders and cancer (Powers et al., 2016; West et al., 2009), perturbations in TDP-43/*NEAT1* cross-regulation might contribute to amyotrophic lateral sclerosis and other diseases (Adriaens et al., 2016; Hennig et al., 2015; Nishimoto et al., 2013; Tollervey et al., 2011b). Paraspeckles are likely to have additional functions beyond recruiting TDP-43, since they are a non-membrane bound organelle that might induce phase separation of multiple RBPs (Hennig et al., 2015). Therefore, I speculate that paraspeckles remodel the nuclear RNA regulatory landscape in a way that facilitates the exit from the pluripotent state. Several mRNAs that are crucial or are specifically expressed in PSCs and are subsequently sequestered in paraspeckles (Chen and Carmichael, 2009; Chillón and Pyle, 2016). A notable example is the transcript of *LIN28A* that contains an *Alu* inverted repeat (IR) region which was proposed to serve as a template for RNA-editing and paraspeckle retention (Chen et al., 2008; Prasanth et al., 2005). As exogenous expression of *LIN28A* in somatic cells

enhances the translation of TDP-43 (Wilbert et al., 2012), this hints to even expanded feed-back loop between LIN28 and TDP43, centered around *NEAT1* regulation.

5 MATERIALS AND METHODS

5.1 CELL CULTURE

Mouse ESC and iPSCs were maintained on gelatin-coated plates in 1:1 neurobasal medium (Life technologies, 21103049) and DMEM (Life technologies, 11320033), supplemented with N2 / B27 (Life technologies, 17502048, 17504044), supplemented with MEK inhibitor PD0325901 (1 μ M), GSK3 inhibitor CHIR99021 (3 μ M) (1408 Axon Medchem and 4953/50 Tocris, respectively), 1,000 U/ml LIF (in-house production), 1% Glutamax (Life Technologies, 35050038), 0.1mM 2-mercaptoethanol (Thermo Fisher Scientific, 31350-010), 1% nonessential amino acids (Thermo Fisher Scientific, 1140050). Cells were passaged by trypsinization (Life Technologies, 25200056). Spontaneous differentiation was induced by the omitting LIF and replacing N2B27 with HyClone™ Fetal Bovine Serum (SV30160.03HI GE Healthcare).

H9 human ESCs (H9 line) were cultured in mTeSR1 medium (05850, Stem Cell Technologies) on Matrigel (diluted 1:1000, 5ml, coating of 10cm plates for 1hr at 37°C., 356234 BD Biosciences)-coated plates. Colonies were passaged by gentle cell dissociation reagent (07174, StemCell Technologies).

HESCs differentiation to PS progenitors was induced by dissociation of colonies into single cells using Accutase (A6964, Sigma) and seeding on Matrigel-coated plates with mTESR1 medium supplemented with 10 μ M ROCK inhibitor (ROCKi Y-27632, 1254/10, R&D Systems) at 100% confluency. Next day cells were washed with PBS, and the medium was changed to RPMI1640 medium (11875093) supplemented with L-Glutamine (25030081), B27 minus insulin (A1895601) all from Thermo Fisher Scientific, 10 μ M ROCKi and 10 μ M CHIR99021. Differentiation into trophoblasts was induced by dissociation of colonies using 0,25% Trypsin-EDTA (25200056, Thermo Fisher Scientific) and seeding as single cells (105,000 cells/cm²) on Matrigel-coated plates in DMEM/F12 medium (11320082), supplemented with 20% KnockOut Serum Replacement (Thermo Fisher Scientific, 10828028), Glutamax, nonessential amino acids, beta-mercaptoethanol, and 50 ng/ml BMP4 (314-BP, R&D systems). Medium was changed daily.

EpiSCs were cultured on gelatin and fetal calf serum-coated plates in N2B27 supplemented with 20 ng/ml ACTIVIN A (338-AC-050, RnD), 12 ng/ml FGF2 (100-18B, Peprotech) and 3 μ M IWP2 (sc-252928, Santa Cruz Biotechnology). EpiSC differentiation to PS progenitors was induced by dissociation of colonies with

collagenase and seeding on gelatin-coated plates in N2B27 supplemented with 20 ng/ml ACTIVIN A (338-AC-050, RnD), 12 ng/ml FGF2 (100-18B, Peprotech) and 200ng/ml Wnt3a (in house production, ErasmusMC, lab of Derk ten Berge).

Mouse embryonic fibroblasts (MEFs) were maintained in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin on gelatin-coated plates.

The HEK293 Flag-TDP-43 inducible cell line was cultured as previously described ¹

5.2 HUMAN FIBROBLAST REPROGRAMMING

1. Reprogramming of primary neonatal human dermal fibroblast (NHDF)

For reprogramming of primary neonatal human dermal fibroblast (NHDF), the cells were harvested using 0.25% Trypsin-EDTA, counted and 1.5x10E6 cells were resuspended in MEF 1 Nucleofector® Kit (VPD-1004, Lonza) solution with 6µg of plasmid DNA, MIP 247 CoMiP 4in1 with shRNA p53 and pCXLE-hMLN (63726 and 27079 Addgene). The cells were pulsed with T-020/N-024 programs using the Nucleofector™ 2b Device (AAB-1001, Lonza). One transfection was divided and plated in three wells of a Matrigel (FALC354230 Schubert&Weiss) coated 6 well tissue culture dish with fibroblast medium composed of 10% HyClone™ Fetal Bovine Serum (SV30160.03HI GE Healthcare), 0.2 mM sodium butyrate and 50 µg/mL ascorbic acid. After the first iPSC-like colonies appeared, between days 15-20, medium was switched to Essential 8 (Thermo Fisher, A1517001) supplemented with 0.2 mM sodium butyrate and 50 µg/mL ascorbic acid in hypoxic conditions. Around days 25–30, the iPSC colonies were large enough for manual selection and counting under the microscope

2. Reprogramming of Secondary human fibroblasts (HiF-Ts)

Secondary human fibroblasts (HiF-Ts) were cultured and reprogrammed as described previously ². Alkaline phosphatase staining (Blue membrane substrate solution AB0300-1KT, Sigma) was used to detect reprogrammed colonies, and reprogramming efficiency was determined as the ratio of reprogrammed colonies relative to the number of the initial cells.

5.3 RNA EXTRACTION AND (Q)RT-PCR ASSAYS

Total RNA was extracted from cell pellets using miRNeasy mini kit (Qiagen, 74104) according to the manufacturer's instructions. First-strand cDNA synthesis of 1 µg total RNA was performed using SuperScript III Reverse Transcriptase (18080085, Invitrogen), according to the manufacturer's instruction. RT-qPCR

reactions were performed using FastStart Universal SYBR Green Master (4367659, Thermo Fisher). Primers used in this study include:

TaqMan Gene Expression Assay, Assay ID: Hs03924655_s1 (human Neat1v2)
TaqMan Gene Expression Assay, Assay ID: Hs01053049_s1, Gene Symbol: SOX2
TaqMan Gene Expression Assay, Assay ID: Hs02758991_g1, Gene Symbol: GAPDH
TaqMan Gene Expression Assay, Neat1, Assay ID: Hs01008264_s1 (Neat1v1)
TaqMan Gene Expression Assay, Assay ID: Hs00999632_g1, Gene Symbol: Pou5f1
TaqMan® Gene Expression Assay, Assay ID: Hs03309805_s1, Gene Symbol: SCARNA10
TaqMan® Gene Expression Assays, Assay ID: Mm03455878_s1 (mouse Neat1v1)

5.4 SINGLE MOLECULE FISH AND PARASPECKLE QUANTIFICATION

Single molecule FISH in mammalian cells was done according to Raj *et al.*^β. Briefly, cells that were grown on sterile 2-well μ -Slides (Ibidi) were washed twice with 1x PBS (5 min each) and fixed with 4 % Paraformaldehyde (methanol-free; Fisher Scientific) for 10 min at RT. After two washes with 1x PBS (5 min each), cells were permeabilized in 70 % ethanol at 4 °C for >12 h. Cells were washed twice with 1x PBS (5 min each) and then incubated for 15 min in 2x saline-sodium citrate (SSC) buffer with 10 % formamide (deionized; Merck Millipore). Hybridization with Stellaris® FISH probes was done in 50 μ l hybridization buffer containing 2x SSC, 10 % formamide, 50 μ g competitor *E.coli* tRNA (Roche Diagnostics), 10 % Dextrane Sulfate (VWR), 2 mg/ml BSA (UltraPure; Life Technologies), 10 mM vanadyl-ribonucleoside complex (NEB) and 1 ng/ μ l probe mix for NEAT1 5`end and middle segment. Cells were covered with parafilm and incubated at 37 °C for 5-6 hours. After washing twice with pre-warmed 2xSSC/10 % formamide for 30 min at 37 °C, cells were washed twice with 1x PBS at RT and then mounted with 10 μ l ProLong® Gold Antifade Reagent containing DAPI (NEB). Cells were imaged by Markus Grosch after 12 hours when the mounting medium was fully cured.

Probes were designed using the Probe Designer software provided by Biosearch Technologies. Probes used were hNEAT1 middle segment conjugated to Quasar®570 (pre-designed by Biosearch Technologies), hNEAT1 5` segment conjugated to Quasar®670 (custom-made), mNEAT1 middle segment conjugated to Quasar®670 (custom-made) and mNEAT1 5` segment conjugated to Quasar®570 (custom-made).

Paraspeckle quantification was done by Markus Grosch in a 3D stack using the spot detection algorithm Airlocalize⁴. 7-10 images containing 40-150 cells each were averaged to calculate number of paraspeckles per cell.

5.5 QUANTSEQ T-FILL 3'MRNASEQ

3 µg of RNA was treated with TURBO DNase (am2238, Thermo Fisher Scientific) according to manufacturer's instructions followed by RNeasy Minelute RNA cleanup (74204, Qiagen). Microcapillary electrophoresis on Agilent 2100 Bioanalyzer with RNA Pico 6000 kit (Agilent, 5067-1513) was used to analyze RNA quality (RIN values >8). 0.5 µg of RNA was used for making libraries for sequencing using Lexogen QuantSeq-REV kit (016.24, Lexogen GmbH) according to manufacturer's protocol, using a poly(T) primer for reverse transcription. The library was sequenced using Illumina HiSeq, producing 60 nt single-end reads and 10 nt index reads.

5.6 TOTAL RNA SEQUENCING

3 µg of RNA was treated with TURBO DNase (Life Technologies, am2238) according to manufacturer's instructions followed by RNeasy Minelute RNA cleanup kit (Qiagen, 74204). Microcapillary electrophoresis on Agilent 2100 Bioanalyzer with RNA Pico 6000 kit (Agilent, 5067-1513) was used to analyze RNA quality (RIN values >8). Per RNA-seq library, 1 µg of DNase-treated RNA was treated with RiboZero Gold (Human/Mouse/Rat) kit (Illumina, RS-122-2301) to remove rRNAs, followed by RNA cleanup using the RNeasy Minelute RNA cleanup kit. Sequencing libraries were prepared using TruSeq Stranded total RNA LT kit (Illumina, RS-122-2301) according to manufacturer's instructions using 11 cycles of PCR followed by purification with Agencourt Ampure XP beads (Beckman-Coulter, A63881). Libraries were evaluated on an Agilent 2100 Bioanalyzer using the DNA 1000 kit (Agilent, 5067-1504). DNA concentration was measured using a Qubit dsDNA HS Assay Kit (Life Technologies, Q32854). Samples were sequenced using a HiSeq2500 instrument to generate 50-nt single-end reads, sequencing depth was 20–40 Mio reads per library. Multiplexing of libraries was performed according to manufacturer's instructions.

5.7 NASCENT RNA-SEQ

Metabolic RNA labelling in ES cells was done by adding 0.4mM EU (5-ethynyl uridine) into the growth medium for 40 min immediately before cell lysis at the indicated time points. Total RNA was prepared using RNeasy mini kit (Qiagen,

74204). For RNA sequencing of nascent transcripts, 10 µg of total RNA was first depleted of ribosomal RNA using Ribo-Zero rRNA removal kit (human/mouse/rat) (Illumina, RS-122-2301) and purified using RNeasy MinElute cleanup kit (Qiagen). rRNA-depleted, EU-labelled RNA (0.5-1 µg) was biotinylated and captured using Click-it Nascent RNA Capture Kit (Thermo Fisher Scientific, C10365) according to manufacturer's instructions. Briefly, RNA was biotinylated with 0.5 mM biotin azide in Click-iT reaction buffer followed by ethanol precipitation. Purified biotinylated RNA was incubated with Dynabeads MyOne streptavidin T1 magnetic beads (Thermo Fisher Scientific, 65601), at 68°C for 5 min and room temperature for 30 min. Beads with bound RNA were washed with Click-iT wash buffers 1 and 2. Washed beads were used for cDNA synthesis and subsequent RNA-seq library preparation using TruSeq Stranded total RNA LT kit (Illumina, RS-122-2301), according to manufacturer's recommendations. Sequencing libraries for all the time points were pooled and sequenced on Illumina NexSeq 500 instrument using high output reagent cartridge producing 75 cycles of single-end reads.

5.8 WESTERN BLOT

Cells were trypsinized and lysed using RIPA buffer, containing phosphatase (Sigma-Aldrich, 4906837001) and protease (Merck, 539134) inhibitors. After addition of 2x SDS loading buffer with 2-Mercaptoethanol (Sigma-Aldrich, M3148) samples were heated at 95°C for 5 minutes. Samples were run on Mini-PROTEAN TGX Stain Free Gels, 4-15% (Bio-Rad Laboratories, 456-8086) and blotted using the Mini Trans-Blot Cell (Bio-Rad Laboratories, 1703930). Following 3x 5 min washing steps with TBS-T, membranes were blocked with 5% milk powder (Carl Roth, T145.1) in TBS-T. Membranes were then incubated over night at 4°C with 5% milk powder in TBS-T containing the primary antibody. After 3x 5 minutes TBS-T washing steps the membrane was incubated with goat anti-rabbit IgM-HRP (Santa Cruz, sc-2030) in 5% milk powder in TBS-T. Following 4 washing steps, 15 minutes each, with TBS-T the membrane was incubated for 1 min with Clarity Western ECL Substrate (Bio-Rad Laboratories, 170-5060) and imaged with ChemiDoc™ MP System (Bio-Rad Laboratories) or exposed to x-ray film. Antibodies used are listed below.

Human Pluripotent Stem Cell Array Dot blot (ARY010, R&D Systems) was used to measure pluripotency factors according to manufacturer's instructions.

5.9 IMMUNOFLUORESCENCE

Cells were grown in Matrigel coated 8 well chamber slides (Ibidi, 80826) and fixed with 4% PFA/DPBS solution (Pierce™ 16% Formaldehyde (w/v), Methanol-free

28906 Thermo Fisher Scientific™) for 15 min at RT and permeabilized with 0.2% TritonX-100/DPBS (9002-93-1, Sigma Aldrich) solution for 15 min at RT. Primary and secondary antibodies were diluted to data sheet indicated concentrations in 10%FBS/0.2% TritonX-100/DPBS and incubated respectively overnight at 4°C and 1 hour at RT. The samples were washed with DAPI (50ug/ml) solution and imaged using a Zeiss Axiovert 200M epifluorescent microscope.

All antibodies used for western blot and immunofluorescence:

Name	Type	Reactivity	Host	Cat. #	Company
Oct-04	mAB	human, mouse	rabbit	2840	Cell Signalling
SOX2	mAB	mouse, human	rabbit	23063	Cell Signalling
NANOG	mAB	human	rabbit	4903	Cell Signalling
TDP-43	polyclonalAB	human, mouse	rabbit	1078-2-AP	Proteintech
TDP-43	polyclonalAB	mouse, human	rabbit	SAB4200006	Sigma
SFPQ	mAB	mouse, human	mouse	MA1-25325	Thermo Fisher
NONO	polyclonalAB	mouse, human	rabbit	ab70335	Abcam
CPEB1	polyclonalAB	mouse, human	rabbit	3583	Cell Signalling
pSMAD2	mAB	mouse, human	rabbit	8828	Cell Signalling
SMAD2	mAB	mouse, human	rabbit	8685	Cell Signalling
GDF3	polyclonalAB	mouse, human	goat	AF958	RnD

5.10 SHRNA AND SIRNA KNOCKDOWN

We used shRNA targeting human *TDP-43* (GAAACACAAGTGAAAGTAA) or a control targeting firefly luciferase (CGTACGCGGAATACTTCGA) driven by the H1 promoter in the vector FUW coexpressing TagRFP both for transfection and transduction. TDP-43 overexpression lentiviral construct consists of myc-HA tagged TDP-43 that was expressed from a lentivirus plasmid driven by ubiquitin promoter. All Lentiviral particles were produced in HEK293FT cells (Life Technologies) as described previously⁵. All lentiviral constructs were a gift from Dieter Edbauer⁶.

Transfection siRNA for TDP-43 knock-down in hESCs cells was carried out in 6 well plates. Each siRNA oligo was transfected using Lipofectamine RNAi MAX (Invitrogen, 13778150) according to manufacturer's instruction. Prior to transfection, cell colonies were dissociated into single cells using accutase (Sigma, A6964) and

subsequently seeded onto Matrigel-coated 6-well plates containing mTESR1 medium supplemented with ROCKi at 100% confluency. On the next day after seeding, siRNA transfection was performed using 5 nM siRNA duplexes against *TDP-43* mixed with 5 μ l of transfection reagents in 100 μ l DMEM. A Stealth RNAi™ siRNA negative control (Invitrogen, 12935-200) was used as negative control. The mixture was incubated at room temperature for 20 min and dropwise added to the cells. The following siRNA sequence was used to target TDP-43 (Invitrogen, A-012394-14, 5'-GGCUCAAGCAUGGAUUCUA-3').

5.11 GENERATION OF INDUCIBLE TDP-43 OVER EXPRESSION LINE

mESC with inducible TDP-43-GFP overexpression were generated by targeting A2-lox ES cells⁷ with p2lox-TDP-43-GFP human construct. A2-lox-cre cell line in which genes of interest could be targeted to a specific conditionally regulated Rosa26 locus. 2 days prior to recombination, 1:1000 doxycycline of doxycycline stock (2mg/mL, Sigma D9891-1G) was added to the cells to induce Cre that is expressed under the control of tetracycline response element (TRE). Cells were transfected with p2lox-TDP-43-GFP construct using Lipofectamine 2000 (LifeTechnologies, 11668019) according to manufacturer's instructions. The cre cDNA excises itself after recombination and placement of the p2lox-TDP-43-GFP behind the TRE, resulting in highly efficient generation of TDP-43-GFP inducible overexpressing cell line. Following the transfection, the cells were allowed to recover for 24 hours in nonselective naïve mESC medium and 24 hours later selective naïve mESC medium containing G418 (250 ng/ml final concentration) was added for 7-8 days. After this period only single clones were selected and expanded to obtain inducible TDP-43-GFP overexpression mESC line. Inducible Flag-TDP-43 HEK293 flip-in cell line was used for Fig. 32 experiments¹.

5.12 FLOW CYTOMETRY ANALYSIS

Flow cytometry analysis of the *SOX2-GFP* reporter in HEK293 cells was conducted by harvesting the cells using Trypsine, 40 h after co-transfection (Lipofectamin 2000 according to manufacturer's instructions) of respective reporters (Fig. 32) together with miRZIP-21-dTomato miR-21 precursor (modified upon MZIP21-PA-1-GVO-SBI, System Biosciences. HEK293 cells are known to express undetectable levels of pri-miR-21⁸ and were therefore ideal cell system to study miR-21 dependent *SOX2* mRNA decay. FlowJo was used for data analysis, by which only cells coexpressing both plasmids (enriched dTomato/eGFP signal over the set gate; ~20%) were considered for analysis. For flow cytometry analysis coupled to antibody staining, the cells were washed once, resuspended in staining

media (2 - 4% FBS and 1mM EDTA in PBS), and analyzed by flow cytometry. miPSCs/mESCs were prepared as outlined above, and stained with a DyLight 650-conjugated SSEA-1 antibody (clone MC-480, MA1-022-D650, Life Technologies) or Alexa Fluor 647 conjugated SSEA4 Mouse Anti-Human mAb (clone MC813-70, SSEA421, Life Technologies), and analyzed using a FACS Aria III cell sorter (BD Biosciences). Cell debris were excluded by forward and side scatter gating, a 561 nm laser to detect dTomato and a 488 nm argon laser (530/30 nm bandpass filter) for GFP. FlowJo was used for data analysis.

Reprogramming of human fibroblasts was assessed by flow cytometry with anti-human TRA-1-60 antibody coupled to Alexa Fluor® 647 (560122, BD Bioscience). Cell preparation and flow cytometry analysis was performed as above.

Subcellular fractionation

5 to 10 million cells were harvested, washed with ice-cold PBS and centrifuged at 500g/4°C. The pellets were gently resuspended in 380 µl cold cytoplasmic lysis buffer (50 mM Tris-pH 6.5, 100 mM NaCl, 300 mM Sucrose, 3 mM MgCl₂, 0.15 % NP40) supplemented with 100 U RNasin Plus RNase inhibitor (PRN2615, Promega). After incubation on ice for 10 min, cells were briefly vortexed and centrifuged at 1000g/4°C for 3 min. The supernatant containing the cytoplasmic fraction was transferred, and centrifuged again at 4°C/5000 g for 2 min to remove cell debris. Immediately after centrifugation 1 ml of RNA precipitation buffer (RPS; 9.5 ml 100 % EtOH with 0.5 ml 3 M Sodiumacetate) was added and the supernatant was incubated at -20°C for 3 to 5 hours until RNA extraction (below).

The remaining pellet was washed three times with 400 µl cytoplasmic lysis buffer supplemented with an increasing concentration of Ammonium sulfate (50 mM, 200 mM and 500 mM). The rationale behind this was to disrupt the endoplasmic reticulum (ER) which is still attached to the nucleus without breaking the nuclear envelope. For each washing step, the cell suspension was centrifuged at 4°C and 5000 g for 2 min and the supernatant was processed as described below for RNA extraction

380 µl of cold Modified Wuarin-Schibler (MWS) buffer (10 mM Tris-HCl pH 7.0, 4 mM EDTA, 0.3 M NaCl, 1 M Urea, 1 % NP-40) supplemented with 100 U RNasin Plus RNase inhibitor was added to the remaining pellet. Samples were vortexed for 30 sec, incubated on ice for 5 min, then vortexed again and kept further on ice for 10 min. The suspension was centrifuged at 4°C/1000g for 3 min and the resulting supernatant representing the nucleoplasmic fraction was processed as described below for RNA extraction.

The pellet was washed three times by adding 800 µl of MWS buffer,

vortexing for 30 sec and centrifuging at 4 °C/500 g for 2min. 1 ml QIAzol® (79306, Qiagen) was added to the remaining chromatin pellet and after short vortexing, the suspension was stored at -20 °C until further usage.

The RNA fraction in RPS buffer was vortexed for 30 sec and after adding 1 µl Glycoblue (Life Technologies, AM9516) centrifuged at 4 °C/18000 g for 15 min. 1 ml QIAzol® was added to the partially air-dried pellet. 10 µl of 0.5 M EDTA was added to all samples in QIAzol® and heated up to 65 °C and incubated for 10 min to resuspend the pellet. After cooling down, 200 µl of Chlorophorm/Isoamylalcohol (24:1) was added, solution were vortexed for 30 sec and then centrifuged at 18000g/RT for 10 min. The upper aqueous phase was transferred into a new tube and the same volume of isopropanol together with 1 µl of glycoblue was added. After ON incubation at -20 °C, the suspensions were centrifuged at 18000g/RT. The pellets were washed once by adding 1 ml 70 % EtOH and centrifuging at RT/18000 g for 5 min. The pellet were air-dried for 10 min and resuspended in RNase-free water. Cytoplasmic, nucleoplasmic and chromatin fractions were analyzed by RNA-sequencing as outlined above. Random primed, strand specific cDNA libraries were prepared following Illumina stranded Tru-Seq LT protocol (above). 81 – 92 % of obtained cDNA reads after sequencing could be aligned to the human genome allowing a single mismatch.

5.13 ICLIP ANALYSIS

The iCLIP protocol was performed as described previously⁹, with the following modifications. Cells were UV irradiated once with 160 mJ/cm² in a Stratlinker 1800 at 254 nm. TDP-43 was immunoprecipitated with protein A Dynabeads (10002D, Invitrogen) conjugated to rabbit-anti TDP-43 (Sigma, SAB4200006) or GFP (Life, Technologies, A6455). The region corresponding to 55–100kDa complexes was excised from the membrane to isolate the RNA, and sequenced using Illumina HiSeq, instrument to generate 50-nt single-end reads, sequencing depth was 15- 20 Mio reads per library Analysis of reproducibility of crosslink sites, identification of the significant iCLIP crosslink clusters and z-score analysis of enriched pentamers was done as described previously¹⁰ and data was processed by iCount webserver (<http://icount.biolab.si>).

5.14 RNA-SEQ DIFFERENTIAL EXPRESSION

To identify differentially expressed genes from RNA-seq data, we mapped the reads using tophat¹¹. We then produced the per-gene count tables using ht-seq and Ensembl genome annotation v74. Finally, we applied edgeR analysis¹² to the

count tables, which resulted in the list of genes with respective p-value and fold-change values on the gene level.

5.15 APAEXPRESS ANALYSIS PLATFORM

For the analysis of alternatively polyadenylated genes, we applied the analytics implemented in ExpressRNA (expressrna.org), an online research platform with client/server architecture for processing and computational analysis of 3'-end targeted sequence data¹³. ExpressRNA combines newly developed and existing bioinformatics analytics for short read alignment, feature annotation, identifying differentially polyadenylated genes and RNA-protein binding (CLIP) profile integration via RNA-maps.

5.16 APA ANALYSIS

QuantSeq read data were aligned to the hg19 reference genome with STAR aligner¹⁴. Reads that did not align uniquely were filtered out. Identification of polyA-site with QuantSeq relies on annealing of a polyT primer to the polyA tail of mRNAs. Our first step to avoid internal priming, where the primer is annealing to a genomic polyA sequence instead to the polyA tail, was to remove A-rich regions in vicinity of identified polyA-site ([-10..10]). Additionally, only polyA-site that were more than 125 nt apart were included since cleavage is not a nucleotide exact process, and consequently we cluster closely spaced polyA-site. The clustering is done by sorting the polyA-site signals genome wide (reverse), travelling down the sorted list and attributing (clustering) signals in the region [-100, 25] nt to the considered polyA-site.

PolyA sites were categorized based on presence of a preceding polyA signal into 4 classes with polyAR¹⁵: strong, weak, lacking polyA signal and non-categorized sites. We examined the nucleotide composition, overlap with a published dataset of polyA-site¹⁶ and efficiency of cleavage (cDNA counts) for each class of sites, which confirmed that the strong and weak sites were the most reliable sites in terms of known nucleotide composition around the polyA-site.

Strong and weak polyA-site were used to define the position-dependent manner of polyA-site regulation by TDP-43. To avoid sites that result from inefficient cleavage, we employed additional filtering based on the relative site levels (summed across all experiments) within each gene as a measure of the strength of each site. Only polyA-site that exhibited >5% expression level compared to a major site for each gene were included. Since the alternative long 3' UTRs are not always fully annotated, 5k of the intergenic region downstream of each gene were included

for this analysis (however only to the middle of the downstream gene if distance < 5kb).

5.17 IDENTIFYING DIFFERENTIALLY POLYADENYLATED GENES

Genes included were those with two major polyA-site (highly expressed across both test and control). Next, polyA-site pairs were classified as same-exon, composite-exon or skipped-exon. For same-exon sites, the major site was compared to the sum of all other sites within the same exon (exon level). For composite-exon and skipped-exon pairs, we compare the major site to the expression of other sites in the gene.

To estimate the level of change in expression between control and test groups, we look at the “percent change” (pc) score, which is calculated as follows:

$$pc = \frac{\text{control}_{\text{PROXIMAL}}}{\text{control}_{\text{PROXIMAL+DISTAL}}} - \frac{\text{test}_{\text{PROXIMAL}}}{\text{test}_{\text{PROXIMAL+DISTAL}}} [-1, 1].$$

Positive values determine a higher ratio of control vs test in proximal vs distal sites, and the negative value represent the opposite trend. Fisher’s exact test was then used to determine the significance of the change. We label genes as displaying significant changes in polyadenylation between test and control for which the Fisher’s p-value < 0.1. We then further classify the significant genes by looking at pc: repressed (pc < -0.1), enhanced (pc > 0.1) and controls (abs(pc) < 0.1 and p-value > 0.1). GO-term analysis of TDP-43 regulated alternatively polyadenylated transcripts was performed using gene-ontology web interface ¹⁷.

5.18 VISUALIZING POSITION-DEPENDENT POLYA SITE REGULATION WITH RNA-MAPS

After identifying APA genes regulated by TDP-43 both in mouse and human ESCs, we took the 3 regulated sets of genes (repressed, enhanced and controls) and cumulatively plotted iCLIP data around polyA sites, marked with a red line at the center of the RNA-maps. The whole RNA-map approach is similar to plotting RNA-maps around splice sites in the alternative-splicing context ^{18,19}. We additionally plot heatmaps of the repressed and enhanced top 20 targets showing individual positional contributions to the cumulative RNA-map that are made available on the APAexpress webpage (www.apaexpress.org). These heat-maps are normalized by max target value.

5.19 RBPOME (RNA-RBP OCCUPANCY ASSAY)

The RBPome protocol was performed as in detail described previously²⁰ with the following modifications. Per replicate ~50mio human embryonic stem cells were prepared in mTeSR pluripotency medium and cells were UV irradiated once with 160 mJ/cm² in a Stratlinker 1800 at 254 nm. For mouse ontogenetic RBPome I started with ~25mio cells if not indicated otherwise. Naïve, limes and epiSCs, cells were cultured in media described in 5.1. and UV irradiated once with 160 mJ/cm² in a Stratlinker 1800 at 254 nm. PS progenitors were UV irradiated once with 160 mJ/cm² in a Stratlinker 1800 at 254 nm, followed by flow cytometry sorting of T-eGFP and FOXA2-RFP cell populations (gates were set according to naïve PSCs that don't express differentiation markers T and FOXA2).

5.20 MASS SPECTROMETRY

Before loading, the samples were centrifuged for 5 minutes at 4°C. LC-MS/MS analysis was performed on a QExactive HF mass spectrometer (Thermo Scientific) online coupled to a Ultimate 3000 nano-RSLC (Thermo Scientific). Approximately 0.5 µg of every digested sample was automatically injected and loaded onto the trap column at a flow rate of 30µl/min in 3% ACN/ 0.1% FA. After 5 min, the peptides were eluted from the trap column and separated on the C18 analytical column (75 µm i.d. x 25 cm, Acclaim PepMap100 C18, 2 µm, 100Å, Dionex) by a 90 min gradient from 5 to 25% ACN in 0.1% FA at 300 nl/min flow rate followed by a 5 min gradient from 25% to 40% ACN in 0.1% FA. Between each sample, the column was washed with 85% ACN for 5 min followed by equilibration at 3% ACN in 0.1% FA for 18 min. MS spectra were recorded at a resolution of 60000 with an AGC target of 3e6 and a maximum injection time of 50ms from 300 to 1500 m/z. From the MS scan, the 10 most abundant peptide ions were selected for fragmentation via HCD with a normalized collision energy of 27, an isolation window of 1.6 m/z and a dynamic exclusion of 30sec. MS/MS spectra were recorded at a resolution of 15000 with a AGC target of 1e5 and a maximum injection time of 50ms. Intensity threshold was set to 1e4 and unassigned charges and charges of +1 and >8 were excluded.

5.21 FASP DIGEST

Each 10µg of RBPome or total cell lysate were digested with a modified FASP procedure²¹. Briefly, the proteins were reduced and alkylated using dithiothreitol and iodoacetamide, diluted with one volume of UA buffer (8 M urea in 0.1M Tris/HCl pH 8.5) and then centrifuged through a 30 kDa cut-off filter device

(PALL, Port Washington, USA). Samples were washed thrice with UA buffer and twice with 50 mM ammonium bicarbonate prior to digest of the immobilized proteins on the filter for 2 h at room temperature using 1 μ g Lys-C (Wako Chemicals, Neuss, Germany) and for 16 h at 37°C using 2 μ g trypsin (Promega, Mannheim, Germany). Tryptic peptides were collected by centrifugation (10 minutes at 14 000g), and the samples were acidified with 0.5% TFA and stored at -20 °C.

5.22 LABEL-FREE ANALYSIS OF MRNA-INTERACTOME

The acquired spectra were loaded to the Progenesis QI software (version 2.0, Nonlinear Dynamics) for label free quantification and analyzed as described previously²². Briefly, profile data of the MS and MS/MS scans were transformed to peak lists with respective peak m/z values, intensities, abundances (areas under the peaks) and m/z width. After reference selection, the retention times of the other samples were aligned by automatic alignment to a maximal overlay of all features. After exclusion of all features with only one charge or more than seven charges, all remaining MS/MS spectra were exported as Mascot generic file and used for peptide identification with Mascot (version 2.5.1) in the Ensembl Human or Mouse protein database. Search parameters used were: 10 ppm peptide mass tolerance and 20 mmu fragment mass tolerance, one missed cleavage allowed, carbamidomethylation was set as fixed modification, methionine oxidation and asparagine or glutamine deamidation were allowed as variable modifications. A Mascot-integrated decoy database search calculated an average false discovery of 1% when searches were performed with a mascot percolator score cut-off of 13 and an appropriate significance threshold p. Peptide assignments were re-imported into the Progenesis QI software and the abundances of all peptides allocated to each protein were summed up. Resulting normalized protein abundances were used further and compared to existing mRNA-interactome datasets obtained from HeLa²³ and HEK293 cells²⁴. As the aim of this study was not to expand the repertoire of RBPs in hESC, but rather to determine the dynamics of high confidence RBPs bound to mRNAs, we used only the overlap between published and our datasets (388 high confidence RBPs) for further analysis of mRNA interactome changes upon gain of paraspeckels (NEAT1 Δ pA-site vs control hESC).

5.23 MS IDENTIFICATION OF DYNAMIC RNA BINDING PROTEINS

RBPome measurements provided us with raw RNA-occupancy measurements for individual proteins in investigated pluripotent and PS- states. To test whether observed changes in RBPome in fact only reflect changes in proteins' overall abundance during development, we in parallel measured the total cellular

proteomes from biological triplicates of naïve PSCs, limes PSCs, EpiSCs and purified T-eGFP/FOXA2-RFP progenitors. Using 2-peptide criterion for proteome data, we detected 3172 of which 519 were in common with individual state RBPomes (naïve, limes and epiblast respectively). As the RBPome is a measure of RNA occupancy and proteome is a measure of total protein levels, the log differences between states represent fold-changes in raw RNA occupancy and protein levels respectively. As both measurements are in linear relation to input levels, change in RNA occupancy reflects change in protein levels providing consistent RNA binding activity of RBP. The log differences between RBPome measurements (Δ RBPome) and proteome levels (Δ Proteome) are thus expected to remain comparable for a majority of proteins as factors governing inherent RNA binding (e.g. proteins' affinity, expression of RNA targets, changes in complex formation) will in general remain stable. Some minor perturbations are expected due to morphological changes in the cells experiencing naïve to epiblast transition – in particular enlargement of cytoplasm and corresponding increase in total and/or cytoplasmic RNA quantity. Skewed Δ Proteome/ Δ RBPome ratio across biologically relevant pluripotent state-pairs indicates with high probability towards a developmentally regulated dynamic RNA binding activity.

At the start of the analysis that was performed together with Valter Bergant we, to remove possible remaining contaminants, filtered data for proteins intersecting at least two of the naïve, limes and primed pluripotent mRNA interactomes (Fig 19). Applying generalized linear model with 1/variance as a weighting factor to the combined data generated a series of coefficients (equations for reference and m&ms). The last coefficient in the set describes discrepancy between change in mRNA occupancy (mRNA interactome) and change in proteome levels between pairs of pluripotent states – the previously mentioned Δ Proteome/ Δ RBPome ratio. The model also generated probabilities of coefficients being equal to zero, meaning change in RNA interactome was in concordance with change in Proteome. The P values were adjusted using Hommel method. Proteins that had adjusted P values <0.05 were assigned to dynamic mRNA interactome of a pair of states (N:L - 16, N:E - 77, L:E - 99, E:T - 124)

$$\begin{aligned}
 P(x)(state1) &= P(I)(state1) + k1 \\
 R(x)(state1) &= P(x)(state1) + k2 = (P(I)(state1) + k1) + k2 \\
 P(x)(state2) &= P(x)(state1) + k3 = (P(I)(state1) + k1) + k3 \\
 R(x)(state2) &= R(x)(state1) + k3 + k4 = (P(I)(state1) + k1) + k2 + k3 + k4
 \end{aligned}$$

P, R – Proteome and RBPome log2 values

x, I – protein x and protein Intercept

state1, state2 – arbitrary pluripotent states, e. g. naive and primed

k1, k2, k3, k4 – coefficients of \boxtimes lm:

$$k1 = f(\text{protein})$$

$$k2 = f(\text{dataset})(\text{protein})$$

$$k3 = f(\text{state})(\text{protein})$$

$$k4 = f(\text{dataset})(\text{state})(\text{protein})$$

$$\text{Dyn} = (P(\text{state1}) - P(\text{state2})) - (R(\text{state1}) - R(\text{state2}))$$

$$\text{Dyn} = (P(I)(\text{state1}) + k1 - (P(x)(\text{state1}) + k1 + k3)) - (P(I)(\text{state1}) + k1 + k2 - (P(\text{stateI})(1) + k1 + k2 + k3 + k4))$$

$$\text{Dyn} = (P(I)(\text{state1}) + k1 - (P(x)(\text{state1}) + k1 + k3)) - (P(I)(\text{state1}) + k1 + k2 - (P(\text{stateI})(1) + k1 + k2 + k3 + k4))$$

$$\text{Dyn} = (-k3) - (-(k3 + k4))$$

$$\text{Dyn} = \mathbf{k4}$$

Dyn - a measure of differential polyA RNA occupancy of a RBP, relative to protein levels

Materials and Methods references:

1. Budini, M., Romano, V., Quadri, Z., Buratti, E. & Baralle, F. E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum. Mol. Genet.* **24**, 9–20 (2015).
2. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162**, 412–424 (2015).
3. Raj, A., van den Bogaard, P., Rifkin, S. a., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
4. Lionnet, T. *et al.* A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8**, 165–70 (2011).
5. Fleck, D. *et al.* Dual cleavage of neuregulin 1 type III by BACE1 and ADAM17 liberates its EGF-like domain and allows paracrine signaling. *J. Neurosci.* **33**, 7856–69 (2013).
6. Schwenk, B. M. *et al.* TDP-43 loss of function inhibits endosomal trafficking and alters trophic signaling in neurons. *EMBO J.* 1–21 (2016).
7. Kyba, M., Perlingeiro, R. C. R. & Daley, G. Q. HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* **109**, 29–37 (2002).
8. Ribas, J. *et al.* A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res.* **40**, 6821–6833 (2012).
9. Huppertz, I. *et al.* iCLIP: Protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
10. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458 (2011).
11. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
12. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
13. Gregor Rot, Z. W. *et al.* expressRNA: integrative platform to unravel the regulatory principles of pre-mRNA processing. *Cell Rep.* **in press**, (2017).
14. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
15. Akhtar, M. N., Bukhari, S. A., Fazal, Z., Qamar, R. & Shahmuradov, I. A. POLYAR, a new computer program for prediction of poly(A) sites in human sequences. *BMC Genomics* **11**, 646 (2010).
16. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res.* **22**, 1173–1183 (2012).
17. Blake, J. A. *et al.* Gene ontology consortium: Going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
18. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 1–7 (2006).
19. Witten, J. T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **27**, 89–97 (2011).
20. Castello, A. *et al.* System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491–500 (2013).
21. Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Biotech* **6**, 3–7 (2009).
22. Merl, J., Ueffing, M., Hauck, S. M. & von Toerne, C. Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Müller cells. *Proteomics* **12**, 1902–1911 (2012).
23. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–406 (2012).

24. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* **46**, 674–690 (2012).

6 REFERENCES

- Adriaens, C., Standaert, L., Barra, J., Latil, M., Verfaillie, A., Kalev, P., Boeckx, B., Wijnhoven, P.W.G., Radaelli, E., Vermi, W., et al. (2016). p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* *22*, 861–868.
- Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mattick, J.S. (2008). The eukaryotic genome as an RNA machine. *Science* (80-.). *319*, 1787–1789.
- Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* *46*, 674–690.
- Balzer, E., and Moss, E.G. (2007). Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol.* *4*, 16–25.
- Bava, F.-A., Eliscovich, C., Ferreira, P.G., Miñana, B., Ben-Dov, C., Guigó, R., Valcárcel, J., and Méndez, R. (2013). CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* *495*, 121–125.
- Bedzhov, I., and Zernicka-Goetz, M. (2014). Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* *156*, 1032–1044.
- Bedzhov, I., Leung, C.Y., Bialecka, M., and Zernicka-Goetz, M. (2014). In vitro culture of mouse blastocysts beyond the implantation stages. *Nat. Protoc.* *9*, 2732–2739.
- Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., and Siu, R.K. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.* *13*, 1070–1075.
- Berkovits, B.D., and Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* *522*, 363–367.
- Betschinger, J., Nichols, J., Dietmann, S., Corrin, P.D., Paddison, P.J., and Smith, A. (2013). Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* *153*, 335–347.
- Blauwkamp, T. a., Nigam, S., Ardehali, R., Weissman, I.L., and Nusse, R. (2012). Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nat. Commun.* *3*, 1070.
- Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-gunn, P., Sun, B., Chuva, S.M., Lopes, D.S., Howlett, S.K., Clarkson, A., Ahrlund-richter, L., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *448*, 191–196.
- Budini, M., Romano, V., Quadri, Z., Buratti, E., and Baralle, F.E. (2015). TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum. Mol. Genet.* *24*, 9–20.
- Burtscher, I., and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development* *136*, 1029–1038.
- Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrao, S., Zhang, X., et al. (2015). Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* *162*, 412–424.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian

- mRNA-binding proteins. *Cell* 149, 1393–1406.
- Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. *Trends Genet.* 29, 318–327.
- Chang, H., Martinez, N.J., Thornton, J.E., Hagan, J.P., Nguyen, K.D., and Gregory, R.I. (2012). Trim71 cooperates with microRNAs to repress Cdkn1a expression and promote embryonic stem cell proliferation. *Nat. Commun.* 3, 910–923.
- Chang, H.-M., Triboulet, R., Thornton, J.E., and Gregory, R.I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature* 497, 244–248.
- Chen, L.-L., and Carmichael, G.G. (2009). Altered Nuclear Retention of mRNAs Containing Inverted Repeats in Human Embryonic Stem Cells: Functional Role of a Nuclear Noncoding RNA. *Mol. Cell* 35, 467–478.
- Chen, L.-L., DeCerbo, J.N., and Carmichael, G.G. (2008). Alu element-mediated gene silencing. *EMBO J.* 27, 1694–1705.
- Chillón, I., and Pyle, A.M. (2016). Inverted repeat *Alu* elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. *Nucleic Acids Res.*
- Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A Is a Suppressor of ER-Associated Translation in Embryonic Stem Cells. *Cell* 151, 765–777.
- Choi, H.W., Joo, J.Y., Hong, Y.J., Kim, J.S., Song, H., Lee, J.W., Wu, G., Schöler, H.R., and Do, J.T. (2016). Distinct Enhancer Activity of Oct4 in Naive and Primed Mouse Pluripotency. *Stem Cell Reports* 7, 1–16.
- Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev.* 22, 746–755.
- Cooper-Knock, J., Walsh, M.J., Higginbottom, A., Highley, J.R., Dickman, M.J., Edbauer, D., Ince, P.G., Wharton, S.B., Wilson, S.A., Kirby, J., et al. (2014). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain* 137, 2040–2051.
- Cooper-Knock, J., Higginbottom, A., Stopford, M.J., Highley, J.R., Ince, P.G., Wharton, S.B., Pickering-Brown, S., Kirby, J., Hautbergue, G.M., and Shaw, P.J. (2015). Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol.* 130, 63–75.
- Danckwardt, S., Hentze, M.W., and Kulozik, A.E. (2008). 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* 27, 482–498.
- Derti, A., Garrett-Engle, P., MacIsaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 22, 1173–1183.
- Diecke, S., Lu, J., Lee, J., Termglinchan, V., Kooreman, N.G., Burrige, P.W., Ebert, A.D., Churko, J.M., Sharma, A., Kay, M.A., et al. (2015). Novel codon-optimized mini-intronic plasmid for efficient, inexpensive, and xeno-free induction of pluripotency. *Sci. Rep.* 5, 8081.
- Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.-K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., et al. (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* 4, 403–415.
- Dodsworth, B.T., Flynn, R., and Cowley, S.A. (2015). The current state of Naive human pluripotency. *Stem Cells* 33, 3181–3186.
- Drukker, M., Tang, C., Ardehali, R., Rinkevich, Y., Seita, J., Lee, A.S., Mosley, A.R., Weissman, I.L., and Soen, Y. (2012). Isolation of primitive endoderm, mesoderm, vascular endothelial and trophoblast progenitors from human pluripotent stem cells. *Nat. Biotechnol.* 30, 531–542.
- Du, Z., Sun, T., Hacısuleyman, E., Fei, T.,

- Wang, X., Brown, M., Rinn, J.L., Lee, M.G.-S., Chen, Y., Kantoff, P.W., et al. (2016). Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat. Commun.* *7*, 10982.
- Emani, M.R., E., Stubb, A., Chakroborty, D., Viitala, M., Rokka, A., Rahkonen, N., Moulder, R., Denessiouk, K., Trokovic, R., et al. (2015). The L1TD1 protein interactome reveals the importance of post-transcriptional regulation in human pluripotency. *Stem Cell Reports* *4*, 519–528.
- Enver, T., Pera, M., Peterson, C., and Andrews, P.W. (2009). Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* *4*, 387–397.
- Eréndira Avendaño-Vázquez, S., Dhir, A., Bembich, S., Buratti, E., Proudfoot, N., and Baralle, F.E. (2012). Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.* *26*, 1679–1684.
- Bernemann C, Greber B, Ko K, Sternecker J, Han DW, Araúzo-Bravo MJ, Schöler HR. (2011). Distinct Developmental Ground States of Epiblast Stem Cell Lines Determine Different Pluripotency Features. *Stem Cells*, *11*, 1496–1503.
- Fagoonee, S., Bearzi, C., Di Cunto, F., Clohessy, J.G., Rizzi, R., Reschke, M., Tolosano, E., Provero, P., Pandolfi, P.P., Silengo, L., et al. (2013). The RNA binding protein ESRP1 fine-tunes the expression of pluripotency-related factors in mouse embryonic stem cells. *PLoS One* *8*, e72300.
- Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). *Esrrb* is a direct *Nanog* target gene that can substitute for *Nanog* function in pluripotent cells. *Cell Stem Cell* *11*, 477–490.
- Fox, A.H., Lam, Y.W., Leung, A.K.L., Lyon, C.E., Andersen, J., Mann, M., and Lamond, A.I. (2002). Paraspeckles: a novel nuclear domain. *Curr. Biol.* *12*, 13–25.
- Fu, X., Jr, M.A., Jolla, L., and Cruz, S. (2015). Context-dependent control of alternative splicing by RNA binding proteins. *Nat. Rev. Genet.* *15*, 689–701.
- Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.K., Alvarez, M., Talukder, S., Pan, Q., Mazzone, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* *147*, 132–146.
- Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature* *504*, 282–286.
- Galonska, C., Ziller, M.J., Karnik, R., and Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell* 1–9.
- Geisler S, Collier J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol.* *11*, 699-712.
- Gerstberger, S., Hafner, M., and Tuschl, T. (2013). Learning the language of post-transcriptional gene regulation. 6–10.
- Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and Consequences of Alternative Polyadenylation. *Mol. Cell* *43*, 853–866.
- Gonzales, K.A.U., and Ng, H.H. (2016). Biological networks governing the acquisition, maintenance, and dissolution of pluripotency: Insights from functional genomics approaches. *Cold Spring Harb. Symp. Quant. Biol.* *80*, 189–198.
- Gonzales, K.A.U., Liang, H., Lim, Y.S., Chan, Y.S., Yeo, J.C., Tan, C.P., Gao, B., Le, B., Tan, Z.Y., Low, K.Y., et al. (2015). Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* *162*, 564–579.
- Gonzalez, I., Munita, R., Agirre, E., Dittmer, T. a, Gysling, K., Misteli, T., and Luco, R.F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat.*

- Struct. Mol. Biol. 22, 1–10.
- Gregor Rot, Z.W., Huppertz, I., Modic, M., Lenče, T., Hallegger, M., Haberman, N., Curk, T., Mering, C. von, and Ule, J. (2017). *expressRNA*: integrative platform to unravel the regulatory principles of pre-mRNA processing. *Cell Rep. in press*.
- Hackett, J.A., and Surani, M.A. (2014). Review Regulatory Principles of Pluripotency: From the Ground State Up. *Stem Cell* 15, 416–430.
- Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* 498, 241–245.
- Hennig, S., Kong, G., Mannen, T., Sadowska, A., Kobelke, S., Blythe, A., Knott, G.J., Iyer, K.S., Ho, D., Newcombe, E.A., et al. (2015). Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* 210, 529–539.
- Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA. *Mol. Cell* 32, 276–284.
- Heyd, F., and Lynch, K.W. (2010). Phosphorylation-dependent regulation of PSF by GSK3 controls CD45 alternative splicing. *Mol. Cell* 40, 126–137.
- Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., and Li, R. (2013). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies.
- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G., and Tian, B. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods* 10, 133–139.
- Hough, S.R., Thornton, M., Mason, E., Mar, J.C., Wells, C.A., and Pera, M.F. (2014). Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. *Stem Cell Reports* 2, 881–895.
- Huang, K., Maruyama, T., and Fan, G. (2014). The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. *Cell Stem Cell* 15, 410–415.
- Ignatius, S.O., Harrich, D., F Garc A-mart Nez, L.N., and Gaynor, R.B. (1995). Cloning and Characterization of a Novel Cellular Protein, TDP-43, That Binds to Human Immunodeficiency Virus Type 1 TAR DNA Sequence Motifs. *J. Virol.* 69, 3584–3596.
- Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., Decoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. *442*, 533–538.
- Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7028–7033.
- Joo, J.Y., Choi, H.W., Kim, M.J., Zaehres, H., Tapia, N., Stehling, M., Jung, K.S., Do, J.T., and Schöler, H.R. (2014). Establishment of a primed pluripotent epiblast stem cell in FGF4-based conditions. *Sci. Rep.* 4, 7477.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.
- Kalkan, T., and Smith, A. (2014). Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369, 20130540-.
- Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Reik, W., Bertone, P., and Smith, A. (2016). Tracking the embryonic stem cell transition from ground state pluripotency.
- Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* 19, 489–501.

- Kim, D.-K., Cha, Y., Ahn, H.-J., Kim, G., and Park, K.-S. (2014). Lefty1 and lefty2 control the balance between self-renewal and pluripotent differentiation of mouse embryonic stem cells. *Stem Cells Dev.* *23*, 457–466.
- Kim, H., Wu, J., Ye, S., Tai, C.-I., Zhou, X., Yan, H., Li, P., Pera, M., and Ying, Q.-L. (2013). Modulation of β -catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* *4*, 2403.
- Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuisen, L., Karemaker, I.D., Santanach, A., Jansen, P.W.T.C., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat. Struct. Mol. Biol.* *23*, 682–690.
- Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A.F., Jones, V., Hor, A., De Alencastro, G., Logan, G.J., et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* *14*, 107–120.
- Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C.H., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühler, M., Liu, P., et al. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* *17*, 471–485.
- Kurek, D., Neagu, A., Tastemel, M., Lehmann, J., Van De Werken, H.J.G., Philipsen, S., Van Der Linden, R., Maas, A., Van Ijcken, W.F.J., et al. (2015). Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports* *4*, 114–128.
- Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* *20*, 1122–1130.
- Lanner, F., and Rossant, J. (2010). The role of FGF/Erk signaling in pluripotent cells. *Development* *137*, 3351–3360.
- Lasko, P. (2011). Posttranscriptional regulation in *Drosophila* oocytes and early embryos. *Wiley Interdiscip. Rev. RNA* *2*, 408–416.
- Lee, K.L., Lim, S.K., Orlov, Y.L., Yit, L.Y., Yang, H., Ang, L.T., Poellinger, L., and Lim, B. (2011). Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet.* *7*.
- Lee, Y.B., Chen, H.J., Peres, J.N., Gomez-Deza, J., Attig, J., Štalekar, M., Troakes, C., Nishimura, A.L., Scotter, E.L., Vance, C., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA Foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.* *5*, 1178–1186.
- Leeb, M., Dietmann, S., Paramor, M., Niwa, H., and Smith, A. (2014). Genetic Exploration of the Exit from Self-Renewal Using Haploid Embryonic Stem Cells. *Cell Stem Cell* 1–9.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* *120*, 15–20.
- Li, Y., Sun, Y., Fu, Y., Li, M., Huang, G., Zhang, C., Liang, J., Huang, S., Shen, G., Yuan, S., et al. (2012). Dynamic landscape of tandem 3' UTRs during zebrafish development. *Genome Res.* *22*, 1899–1906.
- Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* *27*, 2380–2396.
- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Wook, S., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2009). NIH Public Access. *456*, 464–469.
- Licatalosi DD, Darnell RB. (2010). RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet.* *11*, 75–87

- De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., et al. (2015). Hallmarks of pluripotency. *Nature* *525*, 469–478.
- Lu, X., Göke, J., Sachs, F., Jacques, P.-É., Liang, H., Feng, B., Bourque, G., Bubulya, P. a, and Ng, H.-H. (2013). SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat. Cell Biol.* *15*, 1141–1152.
- Lukavsky, P.J., Daujotyte, D., Tollervey, J.R., Ule, J., Stuanı, C., Buratti, E., Baralle, F.E., Damberger, F.F., and Allain, F.H.-T. (2013). Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Publ. Gr.* *20*, 1443–1449.
- Ma, C., Karwacki-Neisius, V., Tang, H., Li, W., Shi, Z., Hu, H., Xu, W., Wang, Z., Kong, L., Lv, R., et al. (2016). Nono, a Bivalent Domain Factor, Regulates Erk Signaling and Mouse Embryonic Stem Cell Pluripotency. *Cell Rep.* *17*, 997–1007.
- Mallanna, S.K., Ormsbee, B.D., Iacovino, M., Gilmore, J.M., Cox, J.L., Kyba, M., Washburn, M.P., and Rizzino, A. (2010). Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells* *28*, 1715–1727.
- Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* *13*, 95–101.
- March, Z.M., King, O.D., and Shorter, J. (2016). Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* *1647*, 1–14.
- Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012a). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* *11*, 491–504.
- Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012b). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* *11*, 491–504.
- Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* *78*, 7634–7638.
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* *9*, 625–U26.
- Mayr, and Bartel (2010). Roesser paper 2. *138*.
- Michel, M., Zacher, B., Demel, C., Tresch, A., and Gagneur, J. (2016). TT-seq maps the human transient transcriptome. *Science* *352*.
- Mitchell, S.F., and Parker, R. (2014). Principles and Properties of Eukaryotic mRNPs. *Mol. Cell* *54*, 547–588.
- Mitsui, K., Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* *113*, 631–642.
- Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., et al. (2015). MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA-DNA triplex structures. *Nat. Commun.* *6*, 7743.
- Naganuma, T., and Hirose, T. (2013). Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biol.* *10*, 456–461.
- Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* *31*, 4020–4034.
- Nakagawa, S., and Hirose, T. (2012). Paraspeckle nuclear bodies--useful

- uselessness? *Cell. Mol. Life Sci.* **69**, 3027–3036.
- Nakagawa, S., Naganuma, T., Shioi, G., and Hirose, T. (2011). Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* **193**, 31–39.
- Nakagawa, S., Shimada, M., Yanaka, K., Mito, M., Arai, T., Takahashi, E., Fujita, Y., Fujimori, T., Standaert, L., Marine, J.-C., et al. (2014). The lncRNA *Neat1* is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* **141**, 4618–4627.
- Newman, M.A., Thomson, J.M., and Hammond, S.M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *1539–1549*.
- Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391.
- Nishimoto, Y., Nakagawa, S., Hirose, T., Okano, H.J., Takao, M., Shibata, S., Suyama, S., Kuwako, K.-I., Imai, T., Murayama, S., et al. (2013). The long non-coding RNA nuclear-enriched abundant transcript 1_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. *Mol. Brain* **6**, 31.
- Pei, D. (2009). Regulation of pluripotency and reprogramming by transcription factors. *J. Biol. Chem.* **284**, 3365–3369.
- Percharde, M., Laval, F., Ng, J.H., Kumar, V., Tomaz, R.A., Martin, N., Yeo, J.C., Gil, J., Prabhakar, S., Ng, H.H., et al. (2012). Nco3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming. *Genes Dev.* **26**, 2286–2298.
- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Reyes, A.P., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos *Cell* **167**, 285.
- Piskounova, E., Polyarchou, C., Thornton, J.E., LaPierre, R.J., Pothoulakis, C., Hagan, J.P., Iliopoulos, D., and Gregory, R.I. (2011). Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147**, 1066–1079.
- Powers, J.T., Tsanov, K.M., Pearson, D.S., Roels, F., Spina, C.S., Ebright, R., Seligson, M., de Soysa, Y., Cahan, P., Theißen, J., et al. (2016). Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma. *Nature* **535**, 246–251.
- Prasanth, K. V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249–263.
- Presnyak, V., Alhusaini, N., Graveley, B.R., Collier, J., Presnyak, V., Alhusaini, N., Chen, Y., Martin, S., Morris, N., Kline, N., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Article Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**, 1111–1124.
- Rossant, J. (2008). Stem cells and early lineage development. *Cell* **132**, 527–531.
- Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Francesco, L. Di, Nencini, M., Bozzo, F., Schininà, M.E., Bagni, C., Carri, M.T., et al. (2015). Nuclear accumulation of mRNAs underlies G4C2 repeat-induced translational repression in a cellular model of C9orf72 ALS *Journal of Cell Science*.
- Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P. a, and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' UTRs and fewer microRNA target sites. *Science (80-.)*. **320**, 1643–1647.
- Schmidt, R., and Plath, K. (2012). The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. *Genome Biol.* **13**, 251.
- Schwanhaussner, B. (2011). Global

- quantification of mammalian gene expression control. *Nature* **473**, 337–342.
- Sephton, C.F., Cenik, B., Cenik, B.K., Herz, J., and Yu, G. (2012). TDP-43 in central nervous system development and function: clues to TDP-43-associated neurodegeneration. *Biol. Chem.* **393**, 589–594.
- Shelkovnikova, T.A., Robinson, H.K., Troakes, C., Ninkina, N., and Buchman, V.L. (2014). Compromised paraspeckle formation as a pathogenic factor in FUSopathies. *Hum. Mol. Genet.* **23**, 2298–2312.
- Shyh-chang, N., and Daley, G.Q. (2013). Review Lin28: Primal Regulator of Growth and Metabolism in Stem Cells. *Stem Cell* **12**, 395–406.
- Singh, S.K., Marisetty, A., Sathyan, P., Kagalwala, M., Zhao, Z., and Majumder, S. (2015). REST-miR-21-SOX2 axis maintains pluripotency in E14Tg2a.4 embryonic stem cells. *Stem Cell Res.* **15**, 305–311.
- Smith, a G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690.
- Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. *Nat. Rev. Mol. Cell Biol.* **17**, 139–154.
- Stefani, G., Chen, X., Zhao, H., and Slack, F.J. (2015). A novel mechanism of LIN-28 regulation of let-7 microRNA expression revealed by in vivo HITS-CLIP in *C. elegans*. *RNA* **21**, 985–996.
- Suh, M.-R., Lee, Y., Kim, J.Y., Kim, S.-K., Moon, S.-H., Lee, J.Y., Cha, K.-Y., Chung, H.M., Yoon, H.S., Moon, S.Y., et al. (2004). Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* **270**, 488–498.
- Sysoev, V.O., Fischer, B., Frese, C.K., Gupta, I., Krijgsveld, J., Hentze, M.W., Castello, A., and Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*. *Nat. Commun.* **7**, 12128.
- Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *448*, 2–8.
- Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487.
- Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Zupunski, V., et al. (2011a). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458.
- Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Zupunski, V., et al. (2011b). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458.
- Tsanov, K.M., Pearson, D.S., Wu, Z., Han, A., Triboulet, R., Seligson, M.T., Powers, J.T., Osborne, J.K., Kane, S., Gygi, S.P., et al. (2016). LIN28 phosphorylation by MAPK/ERK couples signalling to the post-transcriptional control of pluripotency. *Nat. Cell Biol.* **1**.
- Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W.B., Cho, C.H.-H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* **136**, 1339–1349.
- Vogt, E.J., Meglicki, M., Hartung, K.I., Borsuk, E., and Behr, R. (2012). Importance of the pluripotency factor LIN28 in the mammalian nucleolus during early embryonic development. *Development* **139**, 4514–4523.
- Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., et al. (2014). Primate-specific endogenous

- retrovirus-driven transcription defines naive-like stem cells. *Nature* *516*, 405–409.
- Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* *39*, 380–385.
- Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* 30676.
- West, J.A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R.E., Hirose, T., Bond, C., et al. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* jcb.201601071.
- West, J. a, Viswanathan, S.R., Yabuuchi, A., Cunniff, K., Takeuchi, A., Park, I.-H., Sero, J.E., Zhu, H., Perez-Atayde, A., Frazier, a L., et al. (2009). A role for Lin28 in primordial germ cell development and germ cell malignancy. *Nature* *460*, 909–913.
- Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol. Cell* *48*, 195–206.
- Wright, J.E., and Ciosk, R. (2013). RNA-based regulation of pluripotency. *Trends Genet.* *29*, 99–107.
- Wu, J., Okamura, D., Li, M., Suzuki, K., Luo, C., Ma, L., He, Y., Li, Z., Benner, C., Tamura, I., et al. (2015). An alternative pluripotent state confers interspecies chimaeric competency. *Nature* *521*, 316–321.
- Ye, J., and Blelloch, R. (2014). Regulation of Pluripotency by RNA Binding Proteins. *Cell Stem Cell* *15*, 271–280.
- Yeo, J.-C., and Ng, H.-H. (2013). The transcriptional regulation of pluripotency. *Cell Res.* *23*, 20–32.
- Ying, Q.-L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* *453*, 519–523.
- You, K.T., Park, J., and Kim, V.N. (2015). Role of the Small Subunit Processome in the Maintenance of Pluripotent Stem Cells. 1–22.
- Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *6*, 3–7.

7 ACKNOWLEDGMENT

First and foremost I'd like to thank my supervisor **Micha Drukker** for giving me the possibility to go through experience of postgraduate research work in his lab, allowing me such a high degree of research freedom and supporting me during these early steps of my academic career, thereby enabling me to grow as a scientist and as a person.

My deepest gratitude goes to **Jernej Ule** for his continuous support, motivation and friendly mentoring. With your enthusiasm for science and all creative ideas you are a strong motivating force and inspiration to me and am grateful for every moment I had the privilege to work with you.

I would like to express my gratitude to my university supervisor **Magdalena Götz** for her support throughout my PhD, very valuable advice on ongoing projects but also future steps in science. Your wide scientific knowledge, enthusiasm and strive for scientific excellence gave me an example to look forward to. I'd also like to thank my other TAC-committee member, **Andreas Pichlmair**, for his time, valuable advice on ongoing projects during all very pleasant discussions.

Next, I would like to thank people who contributed* to this work:

Special thanks go to **Markus Grosch, Gregor Rot, Valter Bergant** and **Christopher Mulholland**. Working closely with you was the ideal experience of scientific collaboration. You are amazing colleagues and dear friends.

Thanks to **Dmitry Shaposhnikov** for providing nascent RNAseq data, invaluable knowledge you shared with me and being an indispensable office mate everybody would wish for. Thanks to **Anna Pertek** and **Ejona Rusha** for friendly supportive environment in the lab and transfer of yearlong practical knowledge. Special thanks go to **Derk ten Berge** and **Silvia Engert** for enticing me with beautiful world of embryogenesis, being exemplary scientists and paving part of my research interest for future years.

Thank you to several external collaborators for the outstanding close collaboration over the last few years. I learned so much from you, enjoyed working with you and am hoping for many fruitful future scientific moments that we'll share: **Sebastian Bultmann, Boris Rogelj, Davide Cacchiarelli, Alfredo Castello, Derk ten Berge, Kaloyan Tsanov, Jianlong Wang, Bettina Schmidt, Stefanie Hauck, Miriam Esgleas, Tomaz Bratkovic, ...**

My appreciation also goes to all colleagues at the Institute of Stem Cell Research at Helmholtz Zentrum, who made it such a beautiful place and the time being a graduate student so unforgettable. Special thanks to the Drukker lab folks (Markus, Chris, Dima, Karen, Ejona, Anna, Valentina, Freddy, Fatma, Maria, Haidi, Lena, Roman, ...) for not only excellent scientific

discussions but also precious moments. You have set up a high criteria and it will be very hard to find such great colleagues in the future.

Very warm thanks also to master students: Sabina Kolar, Tajda Klobučar, Isabel Weisheit and abovementioned Markus and Valter. You guys rock and it was a real pleasure to work with you.

Thanks to the IMPRS and HELENA graduate schools and their coordinators.

A big thanking goes, especially, to my family for always believing in me and supporting me, unconditionally! The most important person in my life I would like to thank is **Tjasa Lepko**. Thank you for being by my side for nearly a decade. You are my point of reference and I look forward to our future life together.

***Contributions to data presented in PhD thesis:**

- smFISH experiments were carried out and analyzed by Markus Grosch
- Gregor Rot developed RNAexpress pipeline enabling analysis of 3'end RNAseq data used in this study
- *Lin28a-eGFP* mPSC line was generated by Christopher Mulholland,
- NEAT1 Δ PAS hPSC line was produced by Markus Grosch. Sabina Kolar helped with generation of human *LIN28A-eGFP* PSC line.
- Juliane Merl-Pham performed label free analysis of RBPome
- Valter Bergant was major driving force behind integration of RBPome and proteome datasets.
- Igor ruiz de los Mozos Aliaga imported iCLIP data to iCount.
- Derk ten Berge and Alex Neagu derived and characterized limes PSCs.

