UNIVERSITY OF NAPLES FEDERICO II

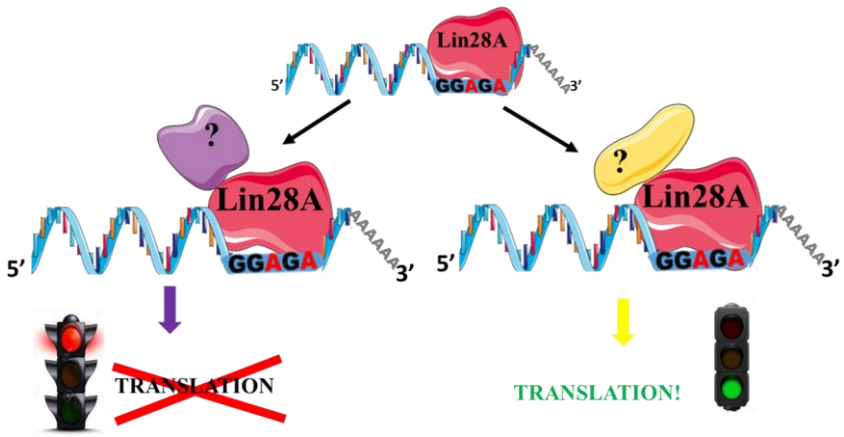
DOCTORATE IN

MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXII CYCLE

Daniela Castaldo

Identification and characterization of Lin28A molecular complexes regulating mRNA recognition and translation in Epiblast Stem Cells



Year 2020

UNIVERSITY OF NAPLES FEDERICO II

DOCTORATE IN

MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXII CYCLE

Identification and characterization of Lin28A molecular complexes regulating mRNA recognition and translation in Epiblast Stem Cells

Tutor

Prof. Tommaso Russo

Co-tutor

Prof. Silvia Parisi

Candidate

Daniela Castaldo

COORDINATOR

Prof. Vittorio Enrico Avvedimento

Year 2020

INDEX

ABBREVIATIONS

ESC - embryonic stem cell

iPSCs - induced pluripotent stem cells

ICM - inner cell mass

EpiLC - epiblastlike stem cell

GFP - green fluorescent protein

IP - immunoprecipitation

miRNA - microRNA

qPCR - quantitative PCR

shRNA - short hairpin RNA

ABSTRACT

RNA binding protein Lin28A is involved in the first steps of mammalian development and in the reprogramming of adult cells into iPS cells. This protein is known to regulate let-7 miRNA family and several mRNAs. It is able to positively regulate some targets, such as Igf2 and Oct4 mRNA, or negatively, as in the case of Hmga2 mRNA. and let-7 miRNA. Since we observed Lin28A accumulation during the establishment of the embryonic stem cell (ESC) differentiated phenotype, we looked at new mRNA targets, coding for proteins involved in the differentiation of ESCs into epiblast stem cell and bearing Lin28 consensus sequence. We found Dnmt3A mRNA as new Lin28A target and its expression resulted to be positively regulated by Lin28A. Furthermore, we were wondering which are the molecular mechanisms underlying the post transcriptional regulation mediated by Lin28A. To do that, we developed an experimental setting to identify Lin28A partners, able to affect its function, based on size-exclusion chromatography, Lin28A co-immunoprecipitation and identification of co-immunoprecipitated proteins through mass spectrometry analysis. We preliminary evaluated the contribution of these putative interactors on the effect of Lin28A overexpression on Dnmt3A mRNA translation, through their silencing. We found proteins having a positive effect on Dnmt3A mRNA translation in a Lin28A-dependent manner and proteins having a negative effect on Dnmt3A mRNA translation that is counteracted by Lin28A. In conclusion, we assumed that some interactors could have an important role in Lin28A-mediated post transcriptional regulation of Dnmt3a expression. These preliminary evaluations provide the proof of concept that Lin28A is part of a molecular machinery, involving several proteins, that regulates the translation of Lin28A target mRNAs.

1. INTRODUCTION

1.1 Stem cells

Stem cells are undifferentiated cells present in embryonic, fetal and adult tissues. They are cells able to self-renew, in a process where, through mitosis, they generate and ensure an undifferentiated stem cell pool, able to differentiate into other types of cells.

Stem cells differ in terms of cell potency, that means they have different abilities to differentiate into specialized cell types. On this bases, it is possible to define:

-totipotent stem cells, as the cells that generate all embryonic and extra-embryonic tissues (zygote and early pre-morula embryo)

-pluripotent stem cells, as the cells that generate all three primary germ layers: endoderm, mesoderm and ectoderm (embryonic stem cells)

-multipotent stem cells, as the cells that generate a limited number of cell types in a specific lineage

-oligopotent stem cells, as the cells that generate few cell types (progenitor cells)

-unipotent stem cells, as the cells that generate only one cell type.

Since embryonic stem cells can differentiate into all kinds of body cells, they represent a biological resource for regenerative medicine. However, there are two main reasons why the use of human embryonic stem cells is limited; first of all, the process of isolating these cells requires the destruction of human embryos and so all of ethical problems about this issue. Moreover, the immune system of the patient recognizes donor

embryonic stem cell-derived cells or tissues as 'non-self', resulting in an immune rejection of the graft.

It has been possible to overcome these issues since when, in 2012, Shinya Yamanaka received his Nobel Prize in Physiology or Medicine discovering that mature cells can be reprogrammed to become pluripotent stem cell, so called induced pluripotent stem cells (iPSCs). Indeed, according to Takahashi (2006), the transduction of four transcription factors, Oct4, Sox2, Klf4 and c-Myc, converts mouse fibroblasts into pluripotent stem cells, they named, induced pluripotent stem cells (iPSCs). These cells were almost indistinguishable from mouse embryonic stem cells in terms of morphology and potency; it means that these cells were able to differentiate into all three germ layers. Afterwards, various iPS cells have been derived from a number of different species, including human, rat, monkey, pig, rabbit and sheep, by transducing fibroblasts, or other adult cells, with the Yamanaka cocktail, Oct4, Sox2, Klf4 and c-Myc, as shown in Figure 1, or with other cocktails, such as for example that containing, further than the above mentioned four factors, Lin28 and Nanog too (Yu 2007).

iPSCs show pluripotent characteristics, such as pluripotency-related gene reactivation, tissue-specific gene inactivation, differentiation capacity to all three germ layers-derived cells, and an epigenetic pattern corresponding to the pluripotent stem cells. Therefore, these cells represent, nowadays, a biological resource for treatment of disease without ethical issues.

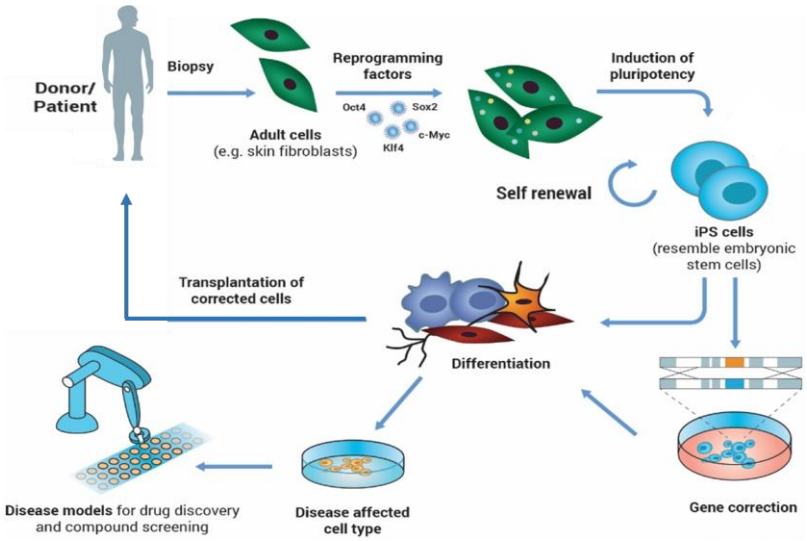


Figure 1: iPSCs origin and application. *iPS cells can be made from adult cells by introducing a cocktail of specific genes. The cells are thereby reprogrammed so becoming pluripotent and similar to the embryonic stem cells. iPS cells can be differentiated to any cell type and then used for transplantation in cell replacement therapy. Similarly, iPSCs can be generated from patients and used to develop disease models precious for exploring disease models and/or to identify new drugs (adapted from Rossbach 2010).*

1.2 Embryonic stem cells and Epiblast stem cells

Embryonic stem cells are pluripotent stem cells that derive from the inner cell mass (ICM) of blastocyst, an early-stage pre-implantation embryo (Evans 1981). Fertilization of the oocyte generates a first single cell, called zygote, that gives rise, through cell divisions, to the morula, a 16 cell pool. Further cell divisions and the formation of a cavity in the morula lead to the blastocyst generation.

Blastocyst is a cluster of approximately 100 cells, composed of the blastocoel, a central cavity filled with fluid, and of two different cell layers:

- trophoectoderm, the external layer of blastocyst that will give rise to all extraembryonic tissues such as placenta, critical for fetus development
- inner cell mass (ICM), the internal layer that generates the embryo

As shown in Figure 2, when the blastocyst moves along the uterine horn into the uterus and implants in the uterine wall, ICM differentiates in:

- epiblast, or primitive ectoderm, that takes place at ICM distal pole, and during gastrulation gives rise to all three embryonic germ layers (ectoderm, mesoderm and definitive endoderm)
- ipoblast, or primitive endoderm, that is a single layer of pseudostratified epithelia, borders the blastocoel and differentiates in yolk sac, one of extraembryonic tissues.

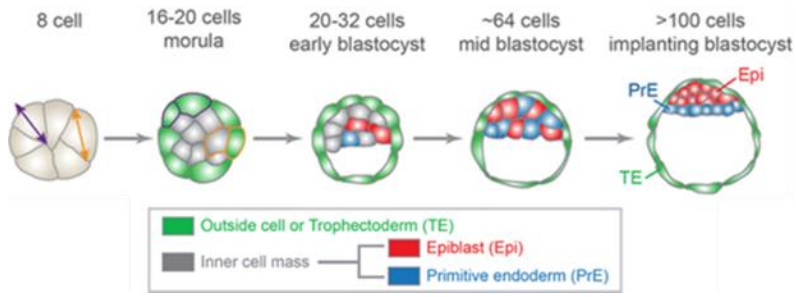


Fig 2: Mouse embryonic development. *Embryonic development starts from the 8 cell stage (2.5 days post-fertilization) to the implanting blastocyst (4.5 days post-fertilization). Starting from the 8 cell stage, the orientation of the plane of cleavage will be important in the generation of inside ICM and outside trophectoderm cells. While symmetric division (purple) will generate two outside cells, asymmetric division (orange) will generate one inside and one outside cell. Subsequently, ICM cells will differentiate into the Epi and PrE (Artus 2014).*

During gastrulation, a process where the embryo is reorganized into a multilayered structure known as gastrula, ICM-derived epiblast differentiates in:

- ectoderm, that generates skin and neuronal cells
- endoderm, that generates lung, thyroid and stomach cells
- mesoderm, that generates cardiac muscle, skeletal muscle, kidney, red blood and smooth muscle cells
- primordial germ cells (PGCs), that generate the germline cells, sperm and eggs (see Figure 3).

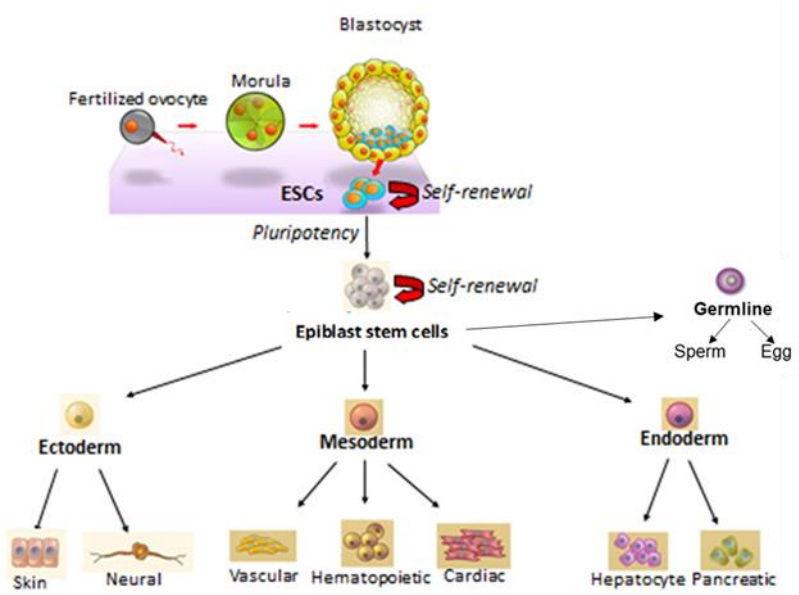


Fig 3: ESC differentiation. *ESCs, naïve pluripotent stem cells, derived from ICM of blastocyst, are able to self-renew and to differentiate in Epiblast stem cells. Epiblast stem cells (primed pluripotent stem cells) are able to self-renew and differentiate into 3 germ layers: ectoderm, mesoderm and endoderm and into primordial germ cells (adapted from Murry 2008).*

At the end of gastrulation, the embryo differentiation started, the body axes, such as dorsal-ventral and antero-posterior axes, are set up and organogenesis takes place.

Mouse embryonic stem cells and epiblast stem cells belong to two different pluripotent states:

- naïve or ground pluripotent state, an '*in vitro'* embryonic stem cell state that resembles the preimplantation embryo state *in vivo*
- primed pluripotent state of epiblast stem cells.

They differ for several hallmarks:

- cell derivation
- self-renewal pathway
- colony morphology
- growth factor requirement for maintenance of pluripotent state
- MHC class I antigen expression
- generating energy pathway
- X inactivation status
- transcription factor expression pattern
- DNA methylation

Mouse ESCs, that are derived from the ICM of preimplantation blastocyst, are able to differentiate into all kinds of body cells and have an unlimited self-renewal capacity when grown *'in vitro'* in the presence of leukemia inhibitor factor (LIF) (Evans 1981). LIF, a multi-functional cytokine which belongs to the IL-6 family (Hilton 1991), is the best-characterized effector of self-renewal in mESCs. LIF acts via a specific cell membrane receptor, a heterodimeric complex composed of a glycoprotein subunit gp130 and the receptor subunit LIFR (also called LIFRβ) (Ernst 2004). The interaction between LIF and its receptor induces the dimerization with gp130 subunit, resulting

in the formation of an activated receptor complex that switches on the constitutively bound tyrosine kinase Janus kinase (JAK). As JAK is activated, it is able to activate, by phosphorylation, the signal transducers and activators of transcription STAT1 and STAT3 (Boulton 1995). In this way, the activated STAT homodimers or heterodimers move from the cytosol to the nucleus and, as transcription factors, promote Oct4, Sox2, c-Myc and Nanog expression (Auernhammer 2000), as shown in Figure 4.

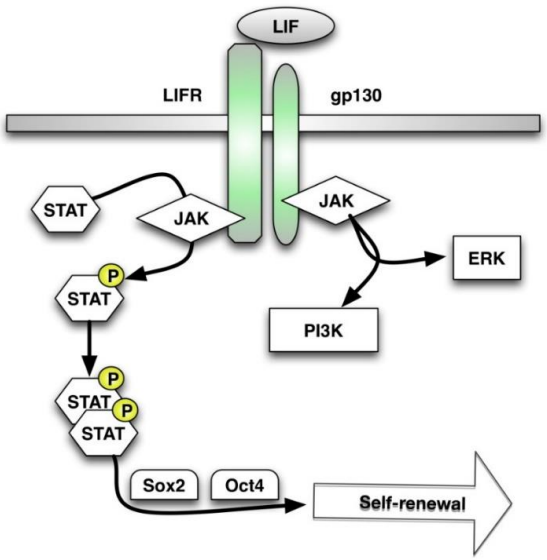


Figure 4: LIF signaling pathways*. LIF binds the membrane receptor complex composed of a glycoprotein subunit gp130 and the receptor subunit LIFR. The interaction between LIF and its receptor induces the dimerization with gp130 subunit, resulting in an activated receptor complex. Only at this point, the receptor-bound tyrosine kinase Janus kinase (JAK) is activated and phosphorylates STAT1 and STAT3. In this way, the activated STAT proteins, as homodimers or heterodimers, move to the nucleus to act as transcription factors, promoting the expression of self-renewal transcription factors Oct4, Sox2, c-Myc and Nanog expression. Adapted from Niwa 2009.*

By light microscopy, it is possible to see that ESCs grow up as uniform, compact, and small dome-shaped refractile colonies. For their metabolism, they are able to use both oxidative phosphorylation (mitochondrial respiration) and glycolysis (Teslaa 2015). Mouse ESCs and EpiSCs share multiple key transcriptional players, including Oct4, Sox2, and Nanog (Boiani 2005, Tesar 2007). They are called the core transcription factors (Silva 2008) because they recognize and bind to specific DNA sequences, such as promoter-proximal DNA elements and to more distal regions (Alan 2011) either activating or preventing transcription. The initial specification of pluripotent cells *in vivo* requires Oct4 expression (Nichols 1998); in fact, the loss of Oct4 expression leads to trophoectoderm differentiation (Niwa 2000).

Oct4 is a critical factor for the maintenance of naïve pluripotency and for the exit from this state (Karwacki-Neisius 2013, Radzisheuskaya 2014). Moreover, Oct4 functions by forming heterodimer with Sox2 in ESCs; in fact, Sox2 binds to DNA sequences adjacent to the Oct4 binding sites (Avilion 2003). Another critical transcription factor, Nanog, is essential for pluripotent cell specification during ICM development (Silva 2008) and induction of pluripotency to finalize somatic cell reprogramming during induction of pluripotency. Nanog promotes a stable undifferentiated ESC state and it is needed for pluripotency to develop in ICM cells (Silva 2008). It acts as a molecular switch to turn on pluripotent program in mammalian cells (Theunissen 2011).

Instead, mEpiSCs, that are derived from epiblast of the post-implantation embryo, typify the primed pluripotent state, together with human ES and iPS cells (Nichols 2011), and represent a more differentiated state than mESCs. These cells maintain an unlimited potential to self-renew, when grown '*in vitro'* with Fibroblast Growth Factor 2 (FGF2) and Activin A (Brons 2007, Tesar 2007). Activin A is one of the TGF-beta superfamily members and interacts with two types of cell surface transmembrane receptors. These receptors are

serine/threonine kinases receptor and, when activated, interact with and phosphorylates two proteins of SMAD protein family that act as transduction factors of TGF-beta superfamily, SMAD2 and SMAD3 proteins. Upon SMAD4 binding to phosphorylated SMAD2 and SMAD3, this complex is able to translocate to the nucleus and, as active transcriptional complex, activates the transcription of different genes, such as Nanog, Sox2 and Oct4, as shown in Figure 5.

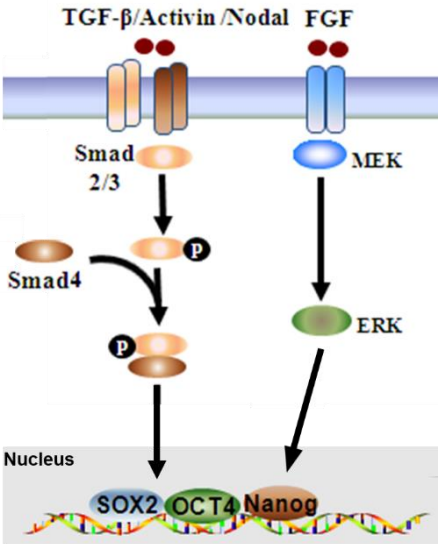


Figure 5: Exogenous growth factor pathways. *Activin A and FGF2 are two growth factors that promote epiblast cell self-renewal and pluripotency. Activin A signaling occurs upon its binding to serine/threonine kinases receptor, resulting in the phosphorylation of SMAD2/3 proteins. Once phosphorylated, these SMADs interact with the SMAD4 and translocate to the nucleus as active transcriptional complexes. Instead, the growth factor FGF2 can signal via transmembrane RTKs down the RAS/RAF/MEK/ERK pathway. Upon stimulation with FGF2, this signaling cascade leads to phosphorylation of ERK, enabling its nuclear translocation. In the nucleus, SMAD and ERK proteins promote the transcription of Nanog, Sox2 and Oct4. Adapted from Molecular Mechanisms of Embryonic Stem Cell Pluripotency, chapter 13.*

Instead, the growth factor FGF2 recognizes and binds tyrosine kinase receptors (RTKs) that leads to the activation of MAPK cascade. As a result, ERK is phosphorylated and then can translocate into the nucleus. Here, SMAD and ERK proteins promote the expression of Nanog, Sox2 and Oct4 genes.

In addition, mEpiSCs differentiate into three germ layers and into PGCs. PGCs are cells capable to start both spermatogenesis and oogenesis; this was demonstrated *in vitro* without using cytokines, but overexpressing a single transcription factor called Prdm14 (Nakaki 2013). By light microscopy, mEpiSCs are larger than mESCs and grow as a flattened monolayer (Tesar 2007). Unlike naïve stem cells, primed pluripotent cells express MHC class I antigen and preferentially generate energy through the glycolytic pathway. Moreover, these cells show an already established random X chromosome inactivation (XCI) in female cells while female mESCs retain two active X chromosomes, as well as female human ESCs. For this reason, the XCI status can be used to identify the optimal culture conditions for maintenance of the naïve stem cells state and for monitoring the epigenetic stability human pluripotent stem cells (Nichols 2011).

In mEpiSCs, Oct4 undergoes a relocalization mediated, at least in part, by a cell-state specific cooperation with other transcription factors, such as Otx2. Although Otx2 is required for transcription of a limited subset of post-implantation epiblast genes, its overexpression in the naïve state is sufficient to recapitulate a large portion of the primed pluripotency gene expression program and to induce widespread Oct4 relocalization to epiblast enhancers (Buecker 2014).

1.3. DNA methylation

DNA methylation, one of epigenetic marks (Chen 2007), plays an important role in stabilizing cell identity and sets up many processes such as genomic imprinting and X chromosome inactivation during the early embryogenesis (Okano 1999) and malignant behavior in tumor cells (Fernandez 2012).

In 1999, studies performed by Okano et al. showed that DNA methyltransferases were dispensable for ESC self-renewal, but that DNA methylation-deficient ESCs were unable to differentiate. These data indicate that naïve pluripotent state is associated with global DNA hypomethylation (Leitch 2013) and that the accumulation of DNA methylation is required for the exit from naïve pluripotent state (Jackson 2004). Indeed, soon after implantation, the embryo undergoes a *de novo* methylation that establishes a genome-wide hypermethylation pattern during ESC differentiation, which reflects a stable ‘‘lock’’ on lineage commitment until cell terminal differentiation (Mohn 2008).

The most frequent DNA methylation occurs when a methyl group is covalently added at the 5-carbon of the cytosine present into the major groove of DNA, inhibiting the transcription. In mammals, it happens almost exclusively in the symmetric CG context and is estimated to occur at ~70-80% of CG dinucleotides throughout the genome (Ehrlich 1982). Since CpG is a simple palindrome sequence, it pairs with another CpG sequence present on the DNA complementary strand; when one of these is methylated, the other one methylated too (Robertson 2000). This pattern allows DNA methylation to be transmitted through cell division thanks to a DNA methyltransferase family (Cheng 2008). This family is composed of:

- maintenance DNA methyltransferases, such as Dnmt1
- de novo DNA methyltransferases, such as Dnmt3

Whereas Dnmt1 is an enzyme that takes care of maintenance DNA methylation to ensure the fidelity of replication of inherited epigenetic pattern, Dnmt3 is a *de novo* DNA methyltransferase family that plays an important role in establishment of DNA methylation patterns. This family, composed of Dnmt3a and Dnmt3b proteins, is strongly expressed in ES cells and developing germ cells, but is expressed at low level in differentiated somatic cells (Chen 2003).

1.4 Lin28 proteins

Lin28 (Cell Lineage Abnormal 28), a highly conserved RNA-binding protein (RBP), is a master post-transcriptional regulator of cell fate that controls embryonic development from the nematode *C. Elegans* to mammals. It is involved in many biological events, such as pluripotency, differentiation, glucose metabolism, reprogramming (Hanna 2009), and oncogenesis (Hamano 2012). Lin28 was first identified and described in *C. Elegans* as an heterochronic gene, an important regulator of developmental timing (Ambros 1984, Moss 1997), and loss of Lin28 function leads to precocious terminal differentiation of skin cells in C. Elegans (Ambros 1989).

In all vertebrates, Lin28 family is a RNA-binding protein family composed of two small proteins (<30KDa): Lin28A and, its paralog, Lin28B. The well-known function of Lin28A and Lin28B is to act as negative regulators of let-7 miRNA biogenesis. Let-7 is one of the first miRNAs to be discovered in *C. Elegans* and its function was first described in the same species because of the opposing expression pattern of Lin28 and let-7 miRNA during larval development (Moss 1997). At an early stage of larval development both Lin28 and pri-let-7 are present; however, no pre-let-7 or mature let-7 can be detected,

indicating a regulation at a post-transcriptional level (Priscilla 2011).

In mammalian cells, during embryonic development, as well as in cancerous tissue, this inverse relationship between Lin28 proteins and let-7 miRNA is retained, as shown in Figure 6.

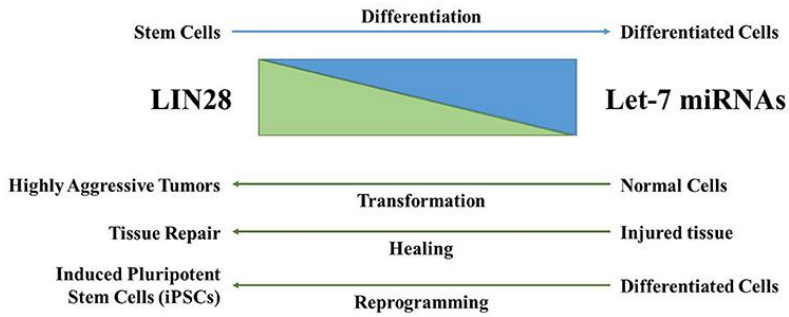


Figure 6: The relationship between LIN28 and let-7 has significant biological functions. *LIN28 levels are high in undifferentiated cells where it blocks the biogenesis of the let-7 microRNA. During the differentiation, LIN28 expression is reduced, resulting in the increase of mature let-7 miRNAs. Lin28 is involved in several mechanisms, such as pluripotency, self-renewal, de-differentiation, and/or cellular transformation during cellular transformation, wound healing (in particular in young mammals), and iPSC generation events (Balzeau 2017).*

Indeed, Lin28 proteins are ubiquitously expressed in embryonic stem cells, in early embryogenesis and in cancer cells, but their expression is downregulated in several tissues in late embryogenesis and in adult life (Yang 2003). For this reason, Lin28 can be considered a marker of stemness (Richards 2004). Once development is completed, Lin28 expression is restricted to only some adult cells such as Henle loop epithelial cells, kidney collecting duct, cardiac and skeletal muscle (Yang 2003) and erythrocytes (De Vasconcellos 2014). Instead, mature let-7 miRNA is undetectable in undifferentiated cells, detected upon

embryonic stem cells differentiation and tissue development (Ribak 2008) and downregulated in tumors (Büssing 2008).

However, this mechanism occurs in a Lin28/let-7 double-negative feedback loop. In particular, on the one hand, let-7 binds to the 3′ UTR of Lin28 mRNA to regulate negatively its expression in worms to mammals (Reinhart 2000). On the other hand, Lin28 binds the consensus sequence in terminal loop of pri-let-7 in the nucleus or nucleolus and pre-let-7 in the cytosol and blocks their processing by Drosha (Newman 2008) and by Dicer (Heo 2008), respectively. In the cytosol, the interaction between Lin28 and pre-let-7 induces the recruitment of polymerase TUT4/7 (Hagan 2009). TUT4 and TUT7 are terminal uridyl trasferase (TUTase) that add uridines to the 3' end of pre-miRNAs to increase their decay. As a result, the poliuridylation of pre-miRNA marks them for degradation mediated by the ribonuclease Dis3L2 (Shyh-Chang 2013) (see Figure 7).

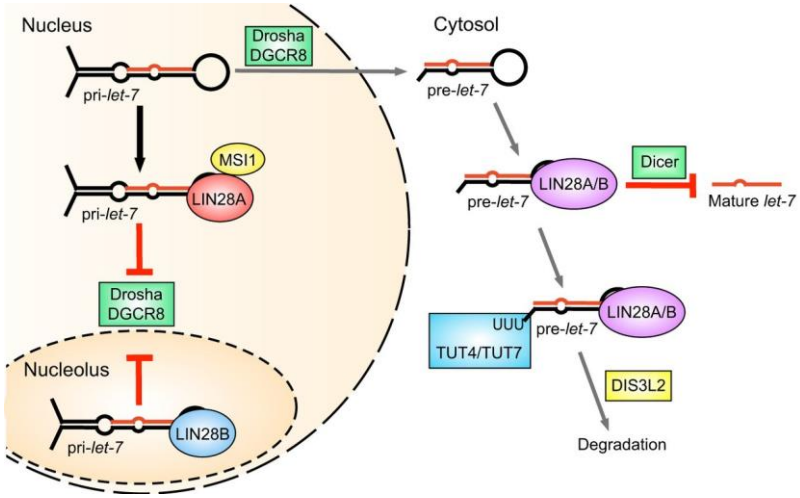


Figure 7: Inhibition of let-7 processing mediated by Lin28. *In the nucleus, LIN28 co-transcriptionally binds pri-let-7 and blocks its processing. The RNA-binding protein MSI1 enhances the localization of LIN28A to the nucleus and, synergistically with LIN28A, blocks the cropping of pri-let-7 (Kawahara 2011). By contrast, LIN28B predominantly localizes to the nucleolus, where it sequesters pri-let-7 to block further processing. In the cytoplasm, LIN28 binds pre-let-7 to block its processing by Dicer and instead induce its oligo-uridylation. In mammals, TUT4 and, to a lesser extent, TUT7 catalyze the oligo-uridylation. Once oligo-uridylated, pre-let-7 is more rapidly degraded than unmodified pre-let-7; in mouse ESCs, the 3′-5′ exonuclease DIS3L12 catalyzes this decay (Shyh-Chang 2013).*

1.4.1. Lin28: structure and functions

All Lin28 proteins contain two RNA-binding domains (RBDs): a N-terminal cold-shock domain (CSD) and a C-terminal CCHC-type zinc fingers domain (ZKD) comprised of two tandemly arranged retroviral-type cysteine cysteine histidine cysteine (CCHC) Zn knuckles (ZnKs). Lin28A and Lin28B share an overall sequence identity of 65% that is so much high in these regions, but differ in some aspects. For example,

Lin28B harbors an extended low-complexity tail region at the C-terminus of the protein (Lee 2014) which contains lysine/arginine (K/R)-rich stretches sequences, such as a nuclear localization signals (NLS) and a nucleolar localization sequences (NoLS) (Piskounova 2008), as shown in Figure 8.

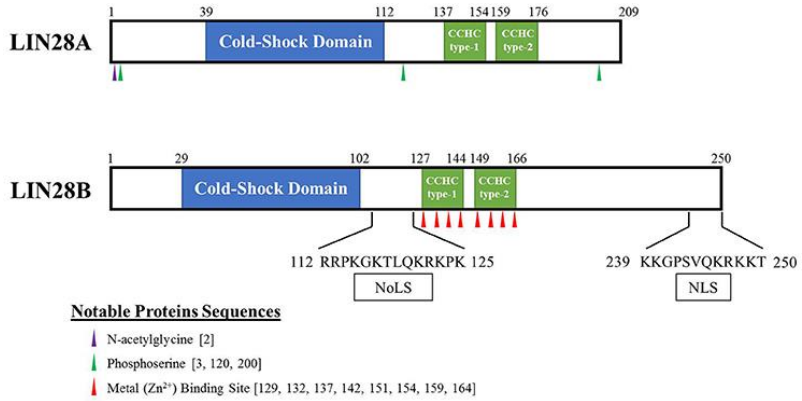


Figure 8: LIN28A and LIN28B protein structures. *Lin28A and Lin28B contain two RNA-binding domains (RBDs): a N-terminal cold-shock domain (CSD) and a C-terminal Zn-knuckle domain (ZKD) comprised of two retroviral type CCHC Zn knuckles (ZnK). Additionally, Lin28B harbors a nuclear localization signals (NLS) and putative nucleolar localization sequences (NoLS), explaining its nucleolar localization. Amino acids belonging to CSD or ZKD are showed in blue or green, respectively (Balzeau 2017).*

Lin28A and Lin28B can localize to both cytosol and nucleus, detected in association with ribosomes, P-bodies and stress granules (Balzeau 2017); however, Lin28A is predominantly located in the cytoplasm of cells whereas Lin28B is predominantly found in the nucleus, specifically in the nucleolus (Piskounova 2011). Crystallography and biochemical studies demonstrate that the CSD and ZKD bind GNGAY and GGAG motifs in the pre-let-7 terminal loop, respectively (Nam

2011). These two domains are separated by a flexible region that allows the binding to distinct pre-let-7 loops, where the GNGAY and GGAG motifs are separated by spacers of different lengths. Both RBDs are necessary to bind let-7 precursors and block their processing.

However, it was demonstrated that Lin28's CSD, through the establishment of interaction with a less conserved terminal hairpin loop, is the first domain to bind to single-stranded RNA sequence and is able to melt a double-stranded stem loops in order to generate an optimal binding interface for the proper recognition of conserved GGAG consensus sequence by Lin28's ZKD (Mayr 2012). Moreover, upon RNA binding, Lin28's ZKD undergoes a conformational change due to a rotation of Pro158 angle present within the Pro-rich linker region between the two ZnK motifs (Loughlin 2011). This allows the interaction with the single-stranded GGAG consensus sequence, where G1 and G4 are closed in a hydrophobic pocket formed by two conserved His140 and Tyr140 amino acid residues in the first Zn knuckle and two conserved His162 and Met170 amino acid residues in the second Zn knuckle, as shown in Figure 9.

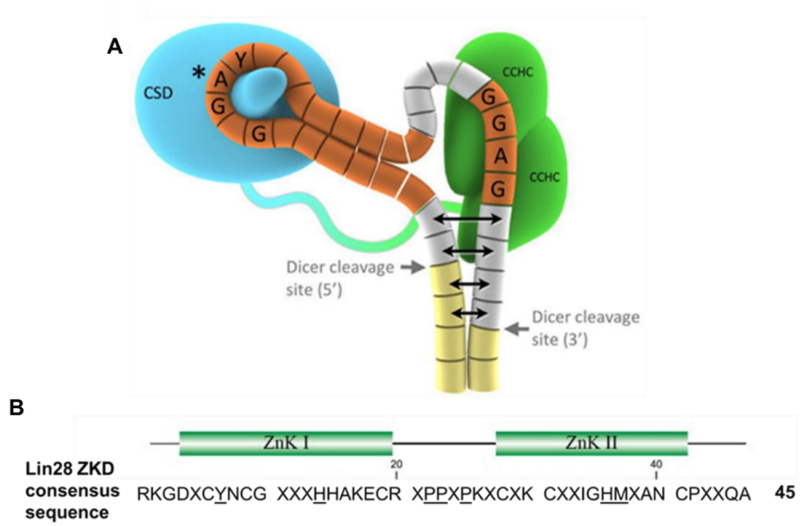


Figure 9: Interaction between LIN28 and let-7. *A) Schematic Model for Interaction between Lin28 and Prelet7. Lin28 proteins contains two RNA binding sites: CSD (blue) and CCHCx2 (green). They bind to two different regions of prelet7. B) ZKD consensus sequence of Lin28 proteins specifically recognizes and binds the single-stranded GGAG consensus sequence in a hydrophobic pocket formed by His140, His162, Tyr140 and Met170 (underlined residues) after ZKD conformational change caused by a rotation of the Pro158 angle within the Pro-rich linker region upon RNA binding (adapted from Mayr 2013).*

This sequence specificity is due to the hydrogen bonding mediated by backbone carbonyl and amide groups of residues that are located within the rigid parts of the CCHC Zn knuckle.

Furthermore, different studies have shown that post-translational modifications of Lin28 proteins, such as acetylation or deacetylation of lysine residues or phosphorylation of serine residues, might affect Lin28 activities. In fact, it is known that when Lin28A is acetylated by PCAF, a well-characterized acetyltransferase that catalyzes the acetylation of proteins (Yao 2001), it induces an alteration of

the Lin28 activity and stability and, as a result, a destabilization of Lin28-dependent let-7 miRNA regulation. However, Lin28 is a substrate not only for PCAF-mediated acetylation, but also for SIRT-mediated deacetylation. SIRT1, a class III deacetylase, is an important regulator involved in metabolism, differentiation, cancer, stress responses and cell senescence. So, the balance between the both activities of PCAF and SIRT1 proteins determines the level of Lin28 acetylation (Wang 2014). Instead, LIN28 phosphorylation occurs through MAPK and ERK kinases in serine 200 residue. This phosphorylation enhances its function and promotes the transition from naïve to primed pluripotent state (Tsanov 2017) and, for this reason, represents a molecular link between cell signaling and cell fate control.

1.4.2 Lin28: direct mRNA targets

Nowadays, the regulation of let-7 miRNAs is the best-studied mechanism by which LIN28 proteins control the gene regulatory networks. In fact, Lin28 expression causes the downregulation of let-7 and the subsequent activation of oncogenes, such as K-RAS, C-MYC, HMGA2, Cyclin D1/D2, CDC25A, CDK6 (Bussing 2008) and diabetes resistance-related genes, such as IGF1R, INSR, and IRS2 (Zhu 2011), indicating its implication in metabolism and cell cycle events.

Over the previous twelve years, it has been shown that LIN28 proteins have also let7-independent functions, exerted through the direct binding to specific mRNAs thus resulting in the post-transcriptional regulation of their expression. This binding occurs at a GGAGA(U) consensus sequence present in an unpaired secondary structure mRNA targets. This sequence is predominantly located within coding exons or in the 3′-UTR of mRNA targets (Wilbert 2012). One of the first mRNAs discovered to be bound by Lin28A (Polesskaya 2007) was the insulin-like growth factor 2 (Igf2). In skeletal myoblasts, one of

few adult tissue where Lin28 expression is still retained, Lin28A regulates positively Igf2 expression by recruiting Igf2 mRNA to complexes composed of mRNA molecules and polysomes. The latter are complexes composed of two or more ribosomes that act to translate one mRNA molecule at the same time through interactions with translation initiation proteins. In fact, Lin28 was called 'translational enhancer' because it is able to drive specific mRNAs to polysomes and thus promote Igf2 protein synthesis.

According to Qiu (2010), in human embryonic stem cells, it was demonstrated that Lin28A acts as a post-transcriptional regulator for the transcription factor Oct4. Oct4, also called Pou5f1 or Oct3/4, was first discovered as an ESC-specific and germline-specific transcription factor (Rosner 1990) and it is one of the four factors in Yamanaka protocol for iPSC reprogramming, as described in paragraph 1. However, as shown in Figure 10, Lin28A promotes Oct4 expression by direct binding to the coding region of its mRNA (Qiu 2009) via interaction with two proteins: RNA helicase A (RHA), that is able to unwind double-stranded RNA structure in order to improve mRNA translation efficiency (Hartman 2006) and eIF3-β, one of RNA-binding protein members of the eukaryotic translation initiation factor 3 (eIF-3) complex, required for several steps in the initiation of protein synthesis.

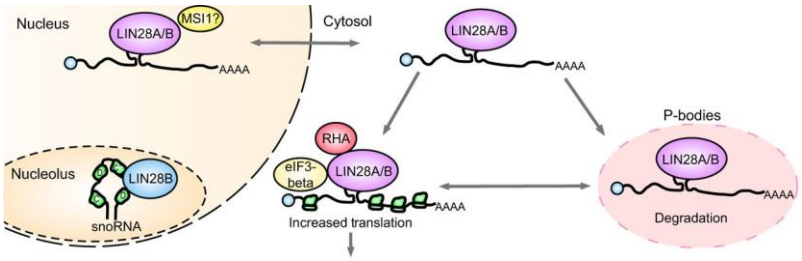


Figure 10: LIN28 promotes the translation of mRNA targets*. The effect of LIN28 on translation is not clearly understood. It could be related to its cellular localization, as it has been found to localize to cellular processing bodies (P-bodies) and stress granules (Balzer 2007), and be mediated through the action of co-factors. The interaction between LIN28 and RHA seems to be required to reach the polysomes (Hartman 2006, Jin 2011, Qiu 2009). Additionally, LIN28A can bind the translation initiation factor eIF3β, although how this interaction impacts LIN28 function is not known (Polesskaya 2007, Qiu 2009). In addition, both forms of LIN28), consistent with reports that LIN28B targets a subset of box C/D small nucleolar RNAs (Piskounova 2011). Taken from Tsialikas 2015.*

However, Lin28 proteins are able to translocate into the nucleus, by an active import mechanism, and assemble into mRNP complexes through their RNA-binding domains. Then, they are exported to the cytoplasm as part of the mRNPs, and remain bound to these complexes until the mRNA targets are translated or sequestered in cytoplasmic RNA granules (Balzer 2007). These include the P-bodies, that contain proteins involved in the translational repression and degradation of mRNAs, and stress granules, which contain mRNA molecules targeted both to translation and degradation machinery; these protein localizations confirm that Lin28 is a regulator of mRNA translation or stability.

In mouse ESCs, other studies have shown that Lin28A regulates some more mRNA targets directly binding their coding region or 3'UTR and increasing their expression (Xu 2009); they include: histone H2a, cyclin A, cyclin B and Cdk4.

According to Parisi (2017), Lin28 expression is increased in mouse epiblast-like stem cells (Epi-LCs), compared to undifferentiated ESCs confirming that Lin28 accumulation starts during the exit of ESC from ground pluripotency state to primed pluripotency state. However, soon after the induction of ESC differentiation, the chromatin-associated non histone protein, Hmga2, is accumulated and controls the ESC differentiation into EpiLCs (Navarra 2016). In particular, Hmga2 expression is high during embryogenesis and oncogenesis and becomes low or undetectable in adults suggesting that Hmga2 has an important role in embryonic development (Fusco 2007, Zhou 1995). Indeed, it is demonstrated that Hmga2 suppression blocks the exit of ESCs from the pluripotent ground state and, together with the transcription factor Otx2, it promotes the activation of new enhancers that govern the EpiLC-specific gene expression program (Navarra 2016), such as Lin28a and Lin28b gene loci enhancer (Parisi 2017).

Additionally, Hmga2 mRNA is known to be one of let-7 miRNA targets (Lee 2007) and, as told before, Lin28 inhibits let-7 biogenesis. So, according to the Lin28/let-7/Hmga2 axis, Lin28 downregulates mature let-7 miRNA levels promoting the subsequent Hmga2 upregulation. Actually, Parisi S. et al. demonstrated that Lin28 proteins directly bind to a GGAGAU sequence in the Hmga2 mRNA 3'UTR (Figure 11).

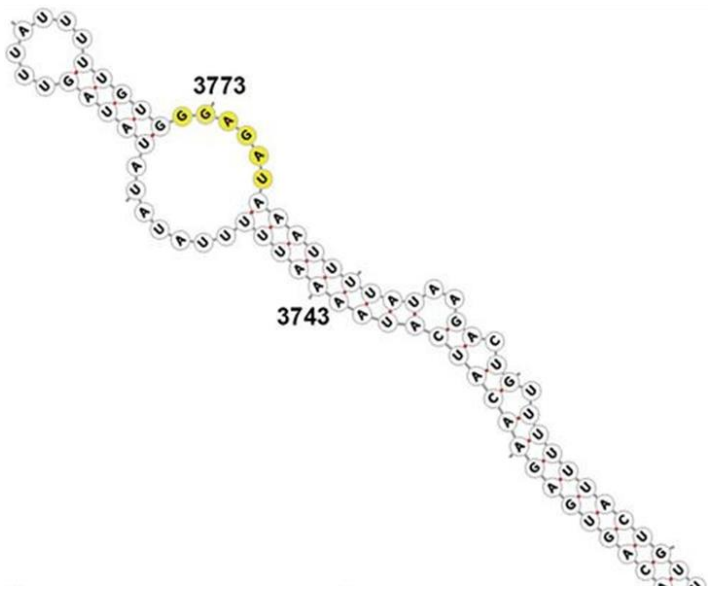


Figure 11: Hypothetical hairpin structure of a region of the 3'UTR of the Hmga2 mRNA involved in interaction with Lin28 proteins. *Lin28-binding site sequence is highlighted in yellow (Parisi 2017).*

Considering that in EpiLCs, let-7 miRNA levels are undetectable or very low, the downregulation of Hmga2 by Lin28 is independent from let-7. In fact, as shown in Figure 12, Lin28 knockdown induces an increased expression of Hmga2 protein; instead, the overexpression of Lin28a results in the downregulation of Hmga2 protein levels. These results are confirmed by the increase of luciferase activity in cells transfected with the luciferase-3'UTR Hmga2 mRNA construct in Lin28 silencing conditions.

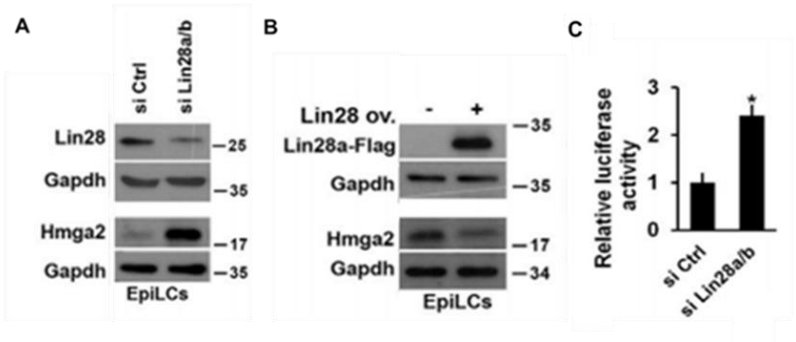


Figure 12: Lin28 regulates Hmga2 expression in let-7-independent fashion. *A) ESCs transfected with Lin28a/b siRNAs and then induced to differentiate into EpiLCs show a downregulation of Lin28a/b protein levels and an increase of Hmga2 protein expression. B) ESCs transfected with a vector driving expression of Lin28a-Flag and then induced to differentiate into EpiLCs show an overexpression of Lin28a protein and a downregulation of Hmga2. C) ESCs were transfected with luciferase-3'UTR Hmga2 mRNA construct and siRNAs targeting Lin28a/b mRNAs or with control siRNAs. Taken from Parisi 2017.*

In 2018, Hui Sang et al. confirmed the critical role of Lin28 proteins during the embryonic stem cells differentiation. In particular, they demonstrated that Lin28A knockdown in ESCs provoked a delayed conversion from naïve to prime pluripotent state; moreover, Lin28A overexpression seems to facilitate the naïve-primed transition and increase the expression of primed related genes, such as Gata6, Fgf5, Dnmt3a and Dnmt3b. This function could be performed thought Lin28A-mediated regulation of Dppa3 gene expression; this is defined as a naïve related gene first observed in preimplantation embryos. This regulation occurs thought Lin28A/let-7/Dnmt3a/b axis; in particular, Lin28-mediated inhibition of let-7 miRNA induces an overexpression of DNA methyltrasferase Dnmt3A/B, one of let-7 targets (Sang 2019). As a results, Dnmt3a/b induces DNA methylation of Dppa3 and its suppression promotes the conversion from naïve to primed pluripotent state. So, Lin28A

expression has an inverse relationship with Dppa3 expression, as shown in Figure 13.

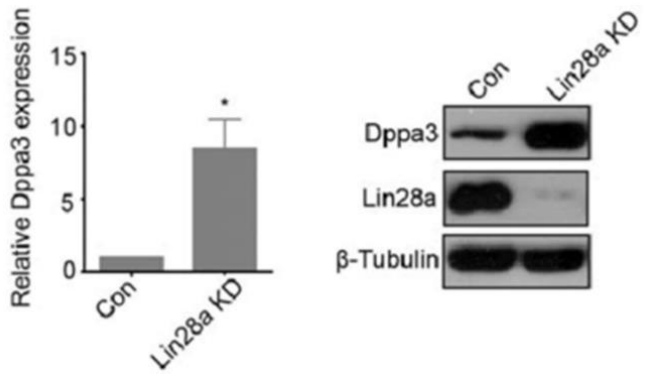


Figure 13: Lin28a negatively regulates Dppa3 expression. *Relative expression of Dppa3 in Lin28a KD ES cells compared with control ES cells was detected by real-time PCR on the left. Western blot analysis of Dppa3 and Lin28a protein levels in Lin28a KD ES cells and control ES cell on the right. Taken from Sang 2019.*

2. AIMS

Regulation of gene expression requires many steps that include transcription, mRNA splicing, mRNA turnover, translation and post-translational events (Orphanides 2002). Over the last decade, the relevance of post-transcriptional mechanisms is increasing and becoming multifaceted; this is due to the involvement of these mechanisms in the regulation of gene activity, time and place (Hershey 2012).

In fact, the mechanisms underlying post-transcriptional regulation are critical for translation, localization and decay of mRNA molecules and, moreover, they are controlled by RNA-binding proteins (RBPs) or non-coding RNAs (Halbeisen 2008). RBPs are proteins that, through one or more globular RNA-binding domains (RBDs), bind and regulate mRNA targets changing their fate or function (Hentze 2018). Since 1989, when Lin28 was discovered in C. Elegans, the role of this protein in stem cell development has become increasingly important enough to be defined as a promoter of pluripotency. According to this definition, Lin28 is considered one of reprogramming factors that can be used to induce mammalian somatic cells to become pluripotent stem cells, called iPSCs (Yu 2007, Hanna 2009).

In all vertebrates, Lin28A and Lin28B are RNA-binding proteins that can regulate positively or negatively their mRNA targets translation. Considering the overexpression of Lin28 proteins during the exit from naïve pluripotent state to primed pluripotent state (Parisi 2017), this PhD thesis aims at investigating the mechanisms underlying regulation of the translation of Lin28-bound mRNAs in mouse epiblast-like stem cells (mEpiLCs). The high expression levels of Lin28 proteins in these cells make mEpiLCs one of the best biological system to analyze Lin28 functions.

Experimental evidence shows that Lin28 proteins act by regulating some of its mRNA targets through its association with polysomes (Jin 2011) and localize in cytoplasmic P-body and stress granules (Balzer 2007), where mRNAs are decapped or sorted to different cellular districts, respectively. Indeed, since Lin28A is predominantly located in the cytoplasm of the cells (Piskounova 2011), we decided to focus our attention on the regulatory events mediated by Lin28A. In particular, we wanted to address the role of Lin28A in the regulation of specific mRNA translation.

As told before, on one hand, Lin28A binds many transcripts, such as Igf2 and Oct4 mRNAs, upregulating their translation; on the other hand, it binds Hmga2 mRNAs, downregulating its translation. Indeed, since Lin28 is able to regulate positively and negatively its targets translation, we have been wondering whether opposite effects of Lin28A on different mRNAs could be due to possible Lin28 partners. To do that, we have developed an experimental setting to identify Lin28A partners in Hela cell lines and in mEpiLCs. This setting is based on the following steps:

- fractionation of cell extracts by size-exclusion chromatography in native conditions
- immunoprecipitation of fractions containing Lin28A
- identification of putative Lin28A interactors thought mass spectrometry fingerprinting
- screening of the putative interactors by exploring the effects of their suppression on Lin28A activity
- preliminary evaluation of the activity of Lin28 partners

This strategy has been used to obtain a proof of concept that Lin28A doesn't work alone when regulating the translation of its targets and that its interactors are critical for the translational fate of mRNAs targeted by Lin28A.

3. MATERIALS AND METHODS

3.1 Cell culture and differentiation

HeLa cell line were maintained in DMEM medium (Gibco) supplemented with 2mM glutamine and 10% FBS (Euroclone) at 37°C in 5% CO2 condition.

E14Tg2a (BayGenomics) mouse ESCs were maintained on feeder-free gelatine-coated plates in the following medium (ESC medium): GMEM (Sigma) supplemented with 2mM glutamine (Invitrogen), 100U/ml penicillin/streptomycin (Invitrogen), 1mM sodium pyruvate (Invitrogen), 1x non-essential amino acids (Invitrogen), 0.1 mM -mercaptoethanol (Sigma), 10% FBS (Hyclone) and 103 U/ml leukemia inhibitory factor (LIF) (Euroclone) at 37°C in 7% CO2 condition.

Mouse EpiLC differentiation is performed plating a single-cell suspension of E14 mESCs on matrigel-coated plates in medium composed of 50% DMEM-F12 medium and 50% Neurobasal medium supplemented with: 1X Pen-Strep, 1X Glutamax (Gibco), 0,33% BSA, 50 μM β-mercaptoetanol (Gibco), 0,5% N2 supplement (Gibco Cat No:17502048), 1% B27 supplement (Gibco), 12 ng/mL basic FGF (Peprothech) and 20ng/mL Activin A (Gibco) at 37°C in 5% CO2 condition.

3.2 Trasfection

For LIN28A-Flag profile analysis, RNA immunoprecipitation and luciferase assays, it was used lipofectamin 2000 (Invitrogen) and vector driving the expression of Lin28A-Flag protein, or with empty vector; in particular HeLa cell line was transfected with DMEM medium (Gibco) and mouse 2day EpiLCs were transfected with Optimem medium (Gibco).

3.3 RNA extraction and Reverse transcription (RT)

Total RNA was extracted by TriSure (Bioline) and RT was performed synthesizing the first-strand cDNA with Mu-MLV RT kit (New England BioLabs), 10 mM dNTPs (Euroclone) and 50µM random hexamers (Bioline). Both methods were performed according to the manufacturer's instructions.

3.4 Quantitative real time PCR (qPCR) analysis

qPCR was carried out with the QuantStudio 7 Flex (Thermo Fisher Scientific) using Fast SYBR Green PCR Master Mix (Thermo Fisher Scientific). The housekeeping GAPDH mRNA was used as an internal standard for normalization. Gene-specific primers used for amplification are listed in Table A1 Appendix. qPCR data are presented as fold changes relative to the indicated reference sample.

3.5 RNA immunoprecipitation

3day mEpiLCs transfected with a plasmid expressing Lin28a-Flag or empty vector were lysed in 1X Polysome Lysis Buffer containing 50% 2X Polysome Lysis Buffer supplemented with: 1 mM Na3VO4, Protease inhibitor cocktail (Sigma-Aldrich), 50mg/mL β-gliceroP, 50mM NaF, 1mM DTT, 400 µM RVC and 100U RnaseOut (Invitrogen).

2X Polysome Lysis Buffer is composed of 20mM pH7 Hepes, 200 mM KCl, 10 mM MgCl2 and 1% NP40.

10% cell estract was stored to be used as input of RNA immunoprecipitation; the rest is immunoprecipited through anti-Flag M2 affinity resin 2h incubation at 4°C on wheel using 50% 2X NT2 buffer supplemented with 1 mM Na3VO4, protease inhibitor cocktail (Sigma-Aldrich), 50mg/mL β-gliceroP, 50mM NaF, 1mM DTT, 400 µM RVC and 100U RnaseOut.

2X NT2 buffer is composed of 100 mM pH7,5 Tris-HCl, 300 mM NaCl, 2 mM $MgCl_2$ and 0,1% NP40.

After the incubation, the immunoprecipitation samples were washed in not supplemented 50% 2X NT2 buffer and stored in Trisure (Bioline) to be used for RNA extraction.

3.6 Protein extraction followed by Western blot

HeLa, ESCs and EpiLCs were lysed in a buffer containing 1 mM EDTA, 50 mM Tris-HCl (pH 7.5), 70 mM NaCl, 1% Triton, and protease inhibitor cocktail (Sigma-Aldrich), and analyzed by Western blot.

3.7 Protein extraction followed by Size-exclusion chromatography

HeLa and 3day mEpiLCs were lysed in 1X Polysome Lysis Buffer. 20 γ RNase A (Sigma-Aldrich) or 100U RnaseOut (Invitrogen) were added to 1X Polysome Lysis Buffer in order to obtain RNAse A-treated or not protein extracts. The RNAse A treatment is performed at R.T. for 30 minutes.

3.8 Ultracentrifugation and Size-exclusion chromatography

After protein quantification, the same amount of protein extract in each sample was ultracentrifuged using Beckman TLA 120.2 120K rpm rotor at 18.000rpm for 30 minutes.

Only the supernatant was taken to be loaded in Superose 6 Increase 10/300 GL column (GE Healthcare) and, afterwards, to be eluted in column buffer composed of 100 mM KCl, 20 mM TrisCl pH 7.8 and 5 mM MgCl2. The column parameters to keep checked were 0,5 ml/min flow rate and 5 MPa column pressure limit. All column parameters are viewed in Table A2 Appendix.

3.9 Protein immunoprecipitation

The protein fractions are immunoprecipited through anti-Flag M2 affinity resin (Sigma-Aldrich) 2h incubation at 4°C on wheel using the supplemented 50% 2X NT2 buffer. After the incubation, the immunoprecipitation samples were washed in 50% 2X NT2 buffer not supplemented and stored in 2X Leammli buffer to be analyzed by mass spectrometry (MS).

3.10 Antibodies and Western blot analysis

The following primary antibodies were used: rabbit anti-Lin28 (1:1000; Abcam), mouse anti-Gapdh (1:1000; Santa Cruz Biotechnology), rabbit anti-RPL3 (1:2000; Abnova), anti Dnmt3a (1:1000; Novus) and mouse anti-Flag (1:2000; Sigma-Aldrich). Gapdh was used as an internal control. Western blots were developed with an ECL system (Advansta) using the following horseradish peroxidase–conjugated antibodies: anti-rabbit IgG (1:10000; Amersham Pharmacia Biotech) and anti-mouse IgG (1:5.000; Amersham Pharmacia Biotech). Western blot bands were quantitated by ImageJ software (Image Processing and Analysis in Java) following user guide instructions. The results were expressed as intensity of each protein band relative to its own Gapdh.

3.11 Mass spectrometry analysis

The mass spectrometry analysis was not performed in our laboratory but in ISPAAM Institute in collaboration with Professor Andrea Scaloni and Dr. Chiara D'Ambrosio.

3.12 Luciferase assay

For luciferase assay, the shRNAs and the plasmid carrying the Dnmt3a 3′ UTR and pRL-TK vector (Promega) expressing *Renilla* luciferase were cotransfected in 2day EpiLCs. ESCs were plated at 15.000 cells per well in a 96-well plate and differentiated in EpiLCs. After 24 h from transfection,

at 3day EpiLCs, the cells were lysed, and firefly and *Renilla* luciferase activities (as an internal control) were measured with the Dual-Luciferase® Reporter Assay System (Promega) by Sirius Luminometer (Berthold Detection Systems, Huntsville, AL, USA). The data were expressed as fold change mean; each value is relative to the same cells transfected with shGFP and obtained after normalization to *Renilla* luciferase reading.

3.13 String analysis

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a biological database and web research for Known and predicted protein-protein interactions. This database contains information obtained from several sources such as experimental data, computational predictive methods and collections of public texts. The protein-protein interactions found can be both physical and functional in more than 1100 fully sequenced organisms. It is freely accessible and regularly updated. It is freely accessible and regularly updated. Through http://string-db.org website, we conducted researches about the interaction study between Lin28A and its putative interactors.

3.14 Statistical analysis

The number of biological replicates of each experiment is indicated in the figure captions. The means of at least 2 independent experiments were used to calculate sd to perform statistical analysis. All *P* values were calculated by Student's *t* test using a 2-tailed test and paired samples.

4. RESULTS

4.1 Increased expression level of Lin28A in mEpiLCs

Recent studies showed that Lin28 expression increases during the transition from naïve pluripotent state to primed pluripotent state. My research group previously demonstrated that LIN28 proteins start a transient accumulation during the first steps of ESC differentiation (Parisi 2017). So, first of all, we investigated the expression level of the specific paralog of LIN28 proteins, LIN28A, during the transition from naïve to primed stem cell state, using a time course of epiblast stem cell differentiation, as shown in Figure 14.

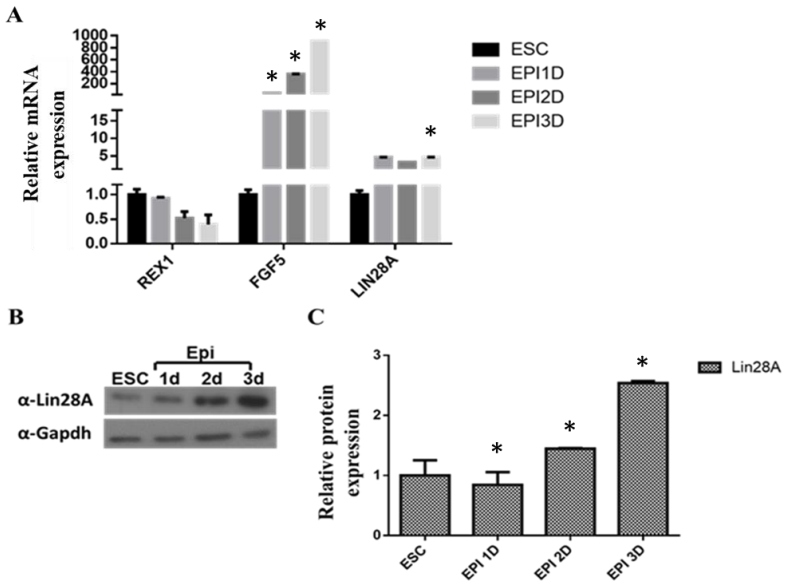


Figure 14: Lin28A expression level increases during the transition from naïve to primed stem cell state. *A) Real-time PCR of Rex1, Fgf5 and Lin28A mRNAs, compared to Gapdh mRNA expression levels. B) and C) Western Blot analysis of Lin28A and relative protein expression, compared to GAPDH protein expression level. The asterisks indicate statistically significant differences (* $p \leq 0,02$).*

This time course was performed on mouse embryonic stem cells from the undifferentiated state (ESC), and at the first three days of differentiation into epiblast-like stem cells (EPI 1D, 2D and 3D). As shown in figure 14A, Lin28A mRNA increased during the epiblast stem cell differentiation. To assess the differentiation of ESC, we explored the expression of Rex1 mRNA, a pluripotency marker known to be highly expressed in the preimplantation-stage embryos, in the placenta and during spermatogenesis (Rogers 1991), the resulted decreased, and the expression of FGF5 mRNA, a widely used marker of the epiblast in the post-implantation embryo and strongly expressed in mouse epiblast stem cells (Brons 2007), that accumulated. Moreover, we analyzed the Lin28A protein expression, using an antibody able to recognize only the C-terminal sequence of this specific paralog; as shown in Figure 14 B and C, we observed a gradual accumulation of the protein, thus confirming the increased expression of Lin28A during the establishment of the epiblast phenotype.

4.2 Dnmt3A mRNA as new Lin28A target

As told before, Lin28A is a RNA binding protein able to bind a specific consensus sequence, GGAGA, present in an unpaired secondary structure located within coding exons or in the 3′-UTR of mRNA targets (Wilbert 2012). This direct binding results in positive or negative regulation of these targets expression mediated by Lin28A. In particular, pri- or pre-let-7 miRNA, first discovered Lin28 target, and Hmga2 mRNA undergo a Lin28A-mediated negative regulation (Piskounova 2011, Parisi 2017), whereas Oct4 and Igf2 mRNAs undergo positive regulation (Polesskaya 2007, Qiu 2010). As already mentioned, my research group previously demonstrated that Lin28 directly binds the consensus sequence GGAGA present in 3'UTR of Hmga2 mRNA, regulating its post-transcriptional expression. However, Hmga2 mRNA level are not detected in naïve embryonic stem cells and the induction of Hmga2

expression is necessary to allow the transition from the naïve pluripotent state of ESCs into the primed pluripotent state of EpiLCs (Navarra 2016, Parisi 2017).

To find new possible mRNAs targets of Lin28A, we investigated several proteins involved in the embryonic stem cell differentiation and playing a significant role in epiblast stem cell establishment.
One of the relevant differences between naïve and primed pluripotent state is the setting of DNA methylation patterns due to the mechanisms underlying the expression and function of *de novo* DNA methyltransferase proteins, DNMT3A and DNMT3B (Chen 2003, Jackson 2004, Leitch 2013, Mohn 2008). Since DNMT3B protein is already expressed in naïve embryonic stem cells (Veillard 2014), we decided to focus our attention on DNMT3A protein.
So, first of all, to investigate Dnmt3A mRNA as new candidate Lin28 target we observed the presence of a putative Lin28 consensus sequence in Dnmt3A mRNA 3'UTR, as shown in Figure 15.

Figure 15: Lin28 consensus sequence in Dnmt3A mRNA 3'UTR. *The putative Lin28 consensus sequence GGAGA is present between 8736 and 8740 nucleotides of Dnmt3A mRNA 3'UTR.*

Then, we analyzed the DNMT3A expression during the transition from naïve to primed stem cell state, performing a time course of epiblast stem cell differentiation. As shown in Figure 16, the Dnmt3A transcript (A) and both DNMT3A1/2 proteins (B and C) are increased during the transition from naïve to primed pluripotent state.
Moreover, to investigate whether Lin28A protein is able to recognize and directly bind the GGAGA consensus sequence located in 3'UTR of Dnmt3A1/2, we transfected EpiLCs with a

vector driving the expression of Lin28A-Flag protein, or with empty vector. Afterwards, we immunoprecipitated Lin28A with an anti-Flag antibody. The mRNAs co-immunoprecipitated with Lin28A were retrotranscribed and measured by qPCR. Figure 17A shows that while the Sox1 mRNA, which accumulated upon the induction of the differentiation and therefore used as control, was not present in the immunoprecipitate from cells expression Lin28a-Flag, DNMT3A mRNA was abundantly present in the immunoprecipitate. Thereafter, in order to assess the effects of Lin28A on Dnmt3A mRNA translation, we used a construct that drives, under the control of the chicken β-actin gene promoter, the expression of a luciferase reporter bearing the Dnmt3a 3′ UTR containing GGAGAU Lin28A consensus motif. The co-transfection of this construct together with that driving the expressing of Lin28a-Flag resulted in a robust induction of the luciferase expression, compared to that observed in cells transfected with and empty Flag vector (Figure 17B).

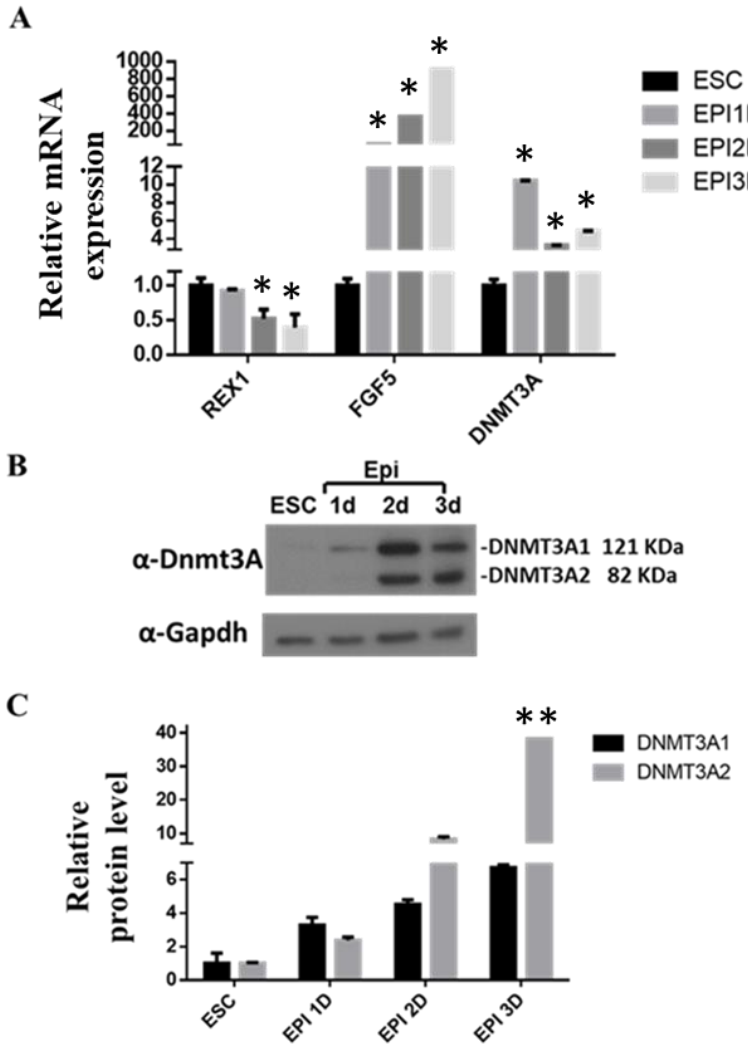


Figure 16: DNMT3A expression level increase during the transition from naïve to primed stem cell state. *A) Real-time PCR revealed Rex1, Fgf5 and Dnmt3A mRNA expression levels, compared to Gapdh mRNA expression levels. B) and C) Increased DNMT3A1 (121 KDa) and DNMT3A2 (82 KDa) protein levels shown in Western Blot analysis and relative protein expression, compared to GAPDH protein expression level. The asterisks indicate statistically significant differences (*p≤0,01, ** p≤0,02).*

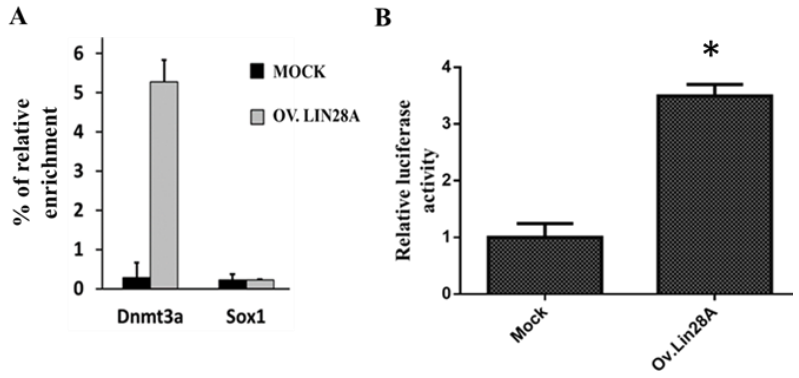


Figure 17: Dnmt3A mRNA is a Lin28A target. *A) EpiLCs, transfected with a plasmid expressing Lin28a-Flag or empty vector, are used to perform RNA-IP analysis. Binding of Lin28A to Dnmt3a mRNA was measured by qPCR on reverse-transcribed IP RNAs. Sox1 mRNA was used as negative control. Data are represented as percentages of relative enrichment of immunoprecipitated RNA relative to input RNA. B) EpiLCs are transfected with a plasmid expressing Lin28a-Flag or empty vector, and with a construct containing Dnmt3A 3'UTR downstream of the luciferase coding sequence. Cell extracts are used to measure the luciferase activity, normalized to Renilla luciferase, transfected as an internal control. The asterisks indicate statistically significant differences (*p≤0,01).*

4.3. Development of a protocol for the purification of Lin28A-containing complexes in HeLa cells

The direct binding of Lin28A to its RNA targets results in positive or negative regulation of their expression. However, the molecular mechanisms underlying this Lin28A-dependent regulation of mRNAs are still not completely understood. Whatever the mechanism that could involve translation or mRNA turnover/trafficking (or both), the dual activity of Lin28A (expression induction or suppression) suggests the existence of regulatory events upstream of Lin28A. Among the many possible regulatory events, we decided to explore the possibility that different Lin28A binding partners could have a crucial role in the effects of Lin28A on its mRNA targets. To

address this possibility, the purification of Lin28A-containing complexes, if any, could allow us to identify the Lin28A partners affecting its activity.

To this aim, the first step of our work was to set up an experimental setting for the purification of Lin28A-containing complexes.
Figure 18 illustrates the workflow of this setting, that we decided to develop in HeLa cells, considering that it is obviously less expensive and easier to grow large amounts of HeLa cells than of EpiSCs.

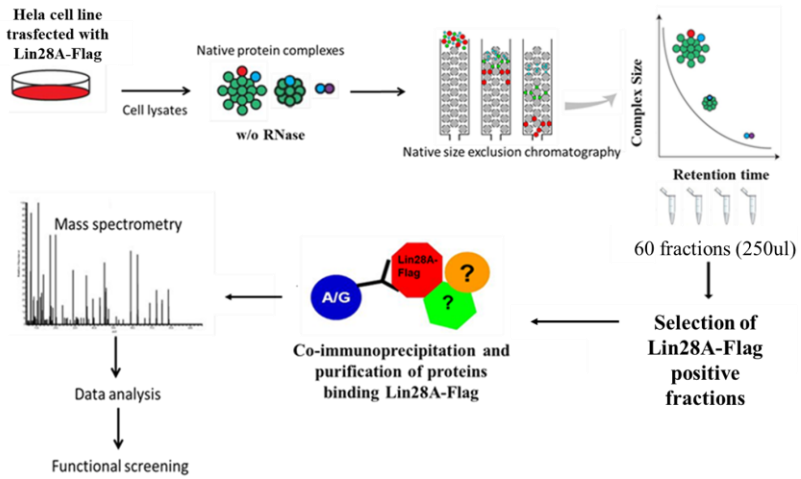


Figure 18: Workflow. *The above schematic representation is the workflow which setting is developed in HeLa cells transfected with Lin28A-Flag vector.*

HeLa cell line were transfected with a vector encoding Lin28A in frame with a Flag-tag or with an empty vector. After 24h, we harvested, lysed and treated each sample with or without RNase A to degrade the RNA; this treatment was done with 20 µg of RNase A for 30 minutes at room temperature. Then, protein mixtures were separated by native size-exclusion chromatography (SEC) using a Superose 6 Increase 10/300 GL

column to obtain fractions of different molecular sizes. Apoferritin (≈450kDa) was used as molecular size marker. Protein profiles of all cell lysates and apoferritin marker are shown in Figure 19A. Western blot analysis shows the different Lin28A-Flag elution profile in the 60 fractions of 250µL each without or with the RNase treatment. In particular, as shown in figure 19B, in RNase A-untreated cells, a strong Lin28A-Flag enrichment was observed in three peaks:

- from 6,5 mL to 8 mL (fractions 4-9)
- from 10 mL to 12 mL (fractions 18-24)
- from 14 mL to 15 mL (fractions 33-36).

On the contrary, in RNase A-treated cell Lin28A profile is observed to shift in a unique peak from 13 mL to 15 mL corresponding to fractions 30-36.

In order to find out an explanation about the presence of Lin28A in the first fractions that contain high molecular size complexes (>450kDa), we studied the elution profile of ribosomal protein L3 (RPL3), a member of the large subunit of cytoplasmic ribosomes, in these fractions, to investigate the potential assembly of a complex between Lin28A and ribosome-containing high molecular weight complexes. As shown in Figure 19C, we observed that the RPL3 elution profile, ranging from fraction 2 to 9, is very similar in HeLa cell lysates exposed or not to RNase. Lin28A molecules are distributed from fraction 4 to fraction 9 in RNase-untreated cells, but absent in the latter fractions in RNase-treated cells. Considering that rRNA molecules present in the ribosome are protected from RNase, the results of this experiment strongly suggests that, in the high molecular weight fractions, Lin28A co-elutes with ribosomes through its interaction with mRNAs, which is lost as a consequence of mRNA degradation by RNase treatment.

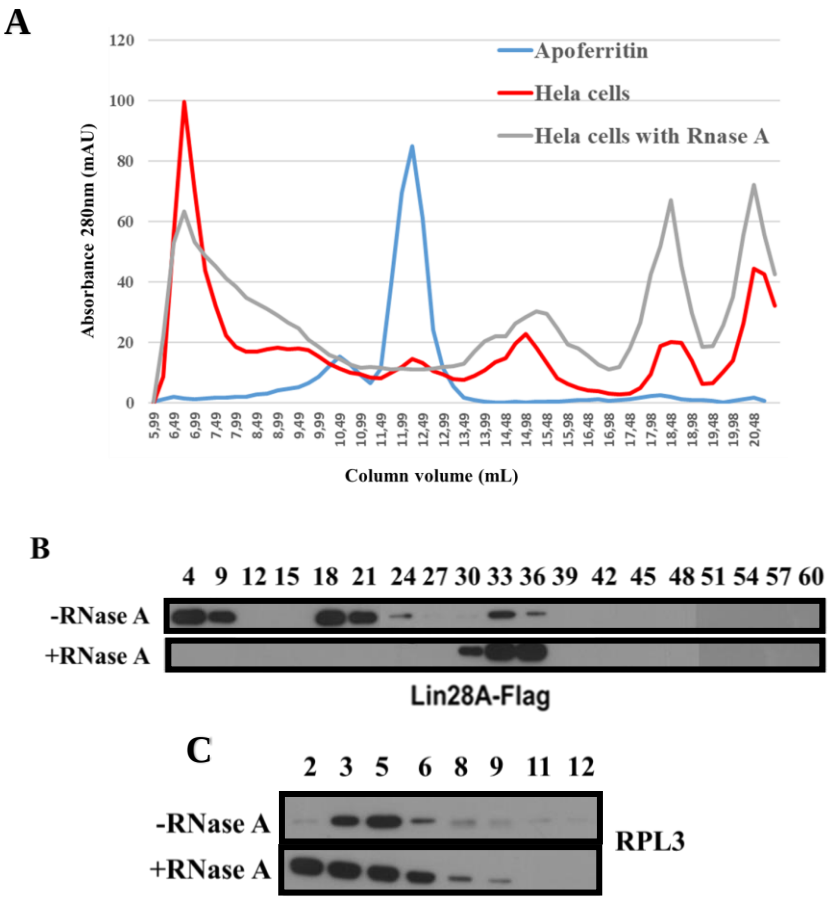


Figure 19: *A) Elution profiles of protein extracts derived from HeLa cell line transfected with Lin28A-Flag vector treated with and without RNase A and apoferritin (≈450 KDa) are shown. B) and C) Western Blot analyses show the Lin28A-Flag and RPL3 expressions in the indicated fractions eluted from size-exclusion chromatography in HeLa cell line treated or not with RNase A.*

4.4 Immunoprecipitation and mass spectrometry data in Hela cell line

In order to set up the Lin28A-Flag immunoprecipitation experiment, we used HeLa extracts derived from transfection with or without Lin28A-Flag vector and treated or not with

RNase A. We used 500µg of protein lysates for size-exclusion chromatography (SEC) and, following SEC, we considered, in each sample, all the fractions included in all three Lin28A-Flag-containing peaks, already observed in Western Blot analysis (Figure 19B) derived from RNase A-untreated cells. Subsequently, we pre-cleared all these fractions to reduce non-specific binding and background and we immunoprecipitated them with anti-Flag M2 affinity resin for 2h incubation at 4°C on wheel. Then, all the immunoprecipitated fractions were washed, stored in 2X Laemmli buffer and sent to ISPAAM Institute laboratories (Dr. Andrea Scalone and Dr. Chiara D'Ambrosio) to be analyzed by mass spectrometry.

MS data showed that 1404 proteins are present in the extracts from Lin28-Flag transfected cells and completely absent from the extracts from mock transfected cells. Moreover, many proteins, specifically immunoprecipitated from the Lin28-Flag transfected cells, are absent from the high-molecular weight fractions of the RNase treated extracts. These results indicated the robustness of the purification procedure and its suitability to be used to isolate candidate Lin28 partners.

In conclusion, the protocol we developed in HeLa cells is based on: cell transfection with and without Lin28A-Flag vector, treatment with RNase A (or mock treatment), size-exclusion chromatography and immunoprecipitation of Lin28A-Flag complexes followed by mass spectrometry analysis.

On the basis of this protocol we assessed that:

- 500 µg of protein in total cell extracts is a sufficient amount to obtain readable mass spectrometry data.
- RNase A treatment is efficient, as shown in Figure 19B, since two out of three peaks, likely representing Lin28A bound to mRNA-containing complexes, disappear (or become barely detectable) in RNase A treated-Lin28A elution profile.

To adapt the protocol developed in HeLa cells to EpiLCs we decided:

- To elute fractions of 500µL, rather than 250 µL, because small fractions did not give substantial advantage.
- To add an ultracentrifugation step, before loading the samples on the column, to remove all lipids and particulate matter from the protein extracts, to prevent column obstruction.

4.5. Identification of proteins present in Lin28A-containing complexes in EpiLCs

To identify candidate partners of Lin28a during the transition from naïve ESCs into EpiLCs, we used 3day mEpiSCs because, as mentioned above, the highest level of Lin28A was observed at this stage, thus high concentration of interaction partners is expected.
Therefore, mEpiLCs were transfected at 2nd day of differentiation with the Lin28A-Flag vector; then, we collected the cells at day 3 of differentiation. We treated two aliquots of the cell lysate with or without RNase A and, after ultracentrifugation at 18.000rpm for 30min, we applied the protein mixtures on the Superose 6 Increase 10/300 GL column; 30 fractions of 500µL each were collected.
In order to evaluate the elution profile of Lin28A free protein, with a molecular weight close to 30kDa, we first used the purified carbonic anhydrase protein as reference marker: we observed that the elution of ≈30KDa protein molecules occurs around 18mL, when fraction 25 is eluted (Figure 20A, Table A3 Appendix). Then, we used thyroglobulin (≈660kDa), apoferritin (≈450kDa) and albumin (≈70kDa) as molecular size markers to have an idea of the molecular sizes of complexes including Lin28A.
Upon size-exclusion chromatography we achieved the protein profiles of all cell lysates and all molecular size range markers

mentioned-above, as shown in Figure 20B and in Table A4 Appendix; in particular, thyroglobulin, apoferritin and albumin result to be eluted in 13mL, 14,5mL and 16,5mL respectively.

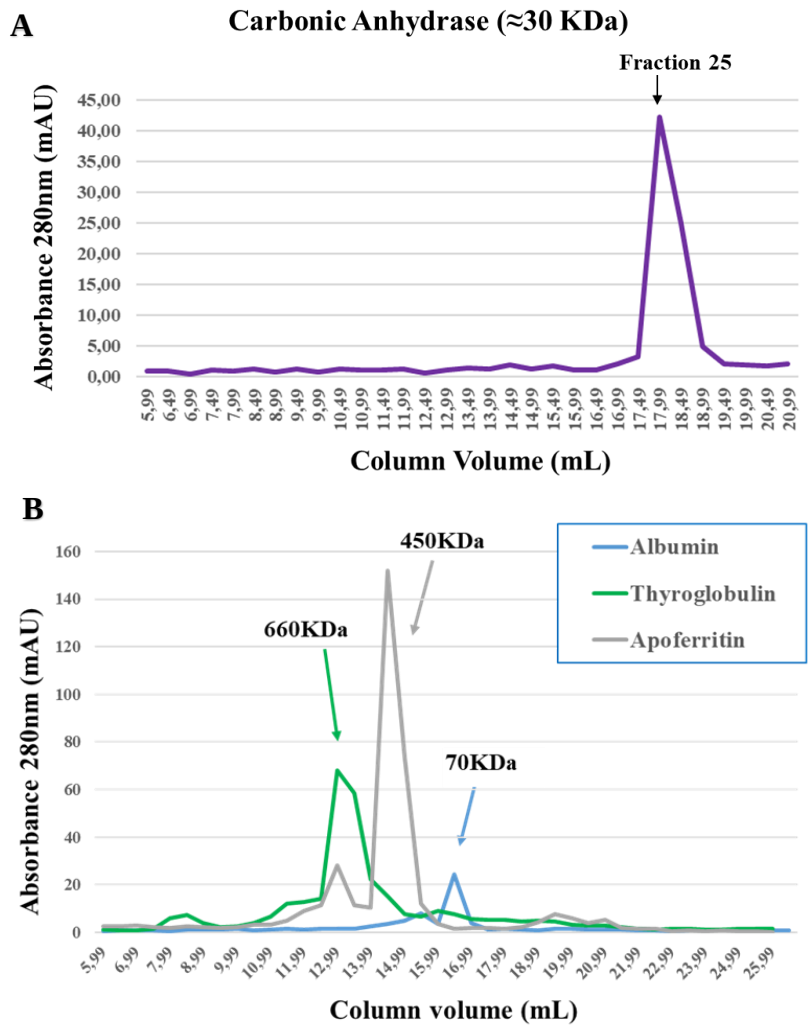


Figure 20: *A) and B) Size exclusion profiles of carbonic anhydrase, as Lin28A Flag reference size marker, and thyroglobulin (≈660 KDa), apoferritin (≈450 KDa) and albumin (≈70 KDa), on Superose 6 increase 10/300 GL column.*

Then, we achieved the protein profiles of all cell lysates upon size-exclusion chromatography and we used the purified apoferritin protein as molecular size marker (≈450 KDa) as performed in Hela cell line. As shown in Figure 21A and Table A5 Appendix, apoferritin results to be eluted in mL 14,5 (fraction 17). As a result, we obtained several fractions, differing in their molecular sizes, that we used to investigate the presence of Lin28A-Flag. This analysis, performed by western blot, has highlighted the strong Lin28A-Flag enrichment in fraction 7 in RNase-untreated EpiLCs, whereas its shifting in fraction 10-11 in EpiLCs treated with RNase (Figure 21B).

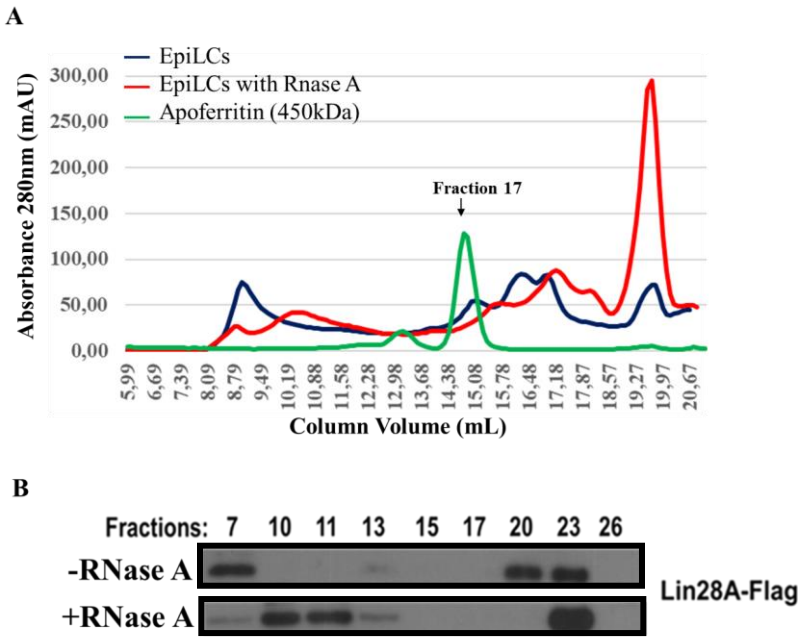


Figure 21: *A) The chart shows the protein profiles of cell lysates upon size-exclusion chromatography (Superose TM 6 Increase 10/300 GL) in EpiLCs treated or not with RNase A, compared to apoferritin size marker. B) Western Blot analysis shows Lin28A-Flag present in the fractions eluted from size-exclusion chromatography in EpiLCs treated or not with RNase.*

4.6 Immunoprecipitation and mass spectrometry steps in mEpiLCs

Since the immunoprecipitation protocol setting in Hela cell line resulted to be efficient, the same method has been applied in mEpiLCs. Hence, we used mEpiLC extracts derived from transfection with and without Lin28A-Flag vector and treatment or not with RNAse A. We pre-cleared the Lin28A-containing fractions, already observed in Western Blot analysis (Figure 21B), to reduce non-specific binding and background and, afterwards, we immunoprecipited them with anti-Flag M2 affinity resin.

In particular, in mEpiLCs under assessment there are:

•fractions 6-7-8-20-21-22-23 in Mock and ov.Lin28A no RNase A-treated samples

•fractions 8-9-10-22-23-24 in Mock and ov.Lin28A RNase A-treated samples

Afterwards, the immunoprecipitated fractions are washed, stored in 2X Leammli buffer and sent to ISPAAM Institute laboratories to be analyzed by mass spectrometry (MS), as performed in Hela cell line.

Subsequently, we received a protein identification report containing a list of proteins previously cleaved by trypsin and subsequently resulted in a list of the accurate peptides analyzed by MS. The identified proteins found in no RNase A-treated samples report were 487, whereas the identified proteins found in RNase A-treated samples report were 493; both numbers result to be very lower than those found in Hela cell line reports.

4.6.1 Protein identification report description

In addition to the protein list, the mass spectrometry report contains, for each detected protein, several parameters, most of which we are going to take into account, such as:

-q-Value, the minimal false discovery rate that indicates the expected proportion of false-positive findings among all the findings; this value enables us to verify how well the protein identification can be considered correct;

-Mascot Score, a summed value of statistical, logarithmic scores of each individual peptide indicating how well the experimental data match the known protein amino acid sequences using the Mascot database search software; this parameter enables us to judge whether a result is significant or not;

-number of amino acids (AAs); molecular weight (MW); coverage, the percentage of the protein sequence covered by identified peptides; theoretical isoelectric point (pI); type of detected protein post-translational modification; all these parameters give us more information about the detected protein;

-peptides, the total number of distinct peptide sequences identified in the protein group; unique peptides, number of peptide sequences that are unique to a protein group and do not occur in the proteins of any other group; number of peptide spectrum matches (PSM's), also called spectral counting, the total number of identified peptide spectra matched for the protein; posterior error probability (PEP) score, a statistical score for the probability that the observed number of peptide spectrum matches is incorrect; all these parameters give us more information about the protein identification;

-Exponentially Modified Protein Abundance Index (emPAI), an approximate quantitation of the protein in a mixture based on protein coverage by the peptide matches in a database search

result; this parameter gives us information about the relative quantification of the detected protein.

4.6.2 Protein identification report elaboration

In order to analyze the mass spectrometry data in mEpiLCs, we decided to perform a data elaboration firstly collecting all proteins found in the list in three lanes for each samples; in particular:

-in no RNase A-treated Mock and ov.Lin28A samples lane 1 corresponds to 6-7-8 fractions, lane 2 corresponds to 20-21 fractions and lane 3 corresponds to 22-23 fractions;

-in RNase A-treated Mock and ov.Lin28A samples lane 1 corresponds to 8-9-10 fractions, lane 2 corresponds to 22-a part of 23 fractions and lane 3 corresponds to a part of 23-24 fractions.

Secondly, focusing our attention on all the information obtained from the protein identification report, in order to carry out a first but solid selection of Lin28A putative interactors, we decided to consider the following criteria:

-observing solely the proteins absent in the empty vector transfected-mEpiLCs fractions;

-observing the percentage of enrichment, the value representing the improvement of protein mascot score in OvLin28A sample compared to that one in Mock sample in the same lane (all proteins having ≤100% enrichment will be excluded);

-observing the average of the spectral count in these proteomic studies compared to the maximum spectral count, as a practical measurement of detected protein abundance;

-paying attention on Contaminant Repository for Affinity Purification (CRAPome) value.

CRAPome value gives us information on how much the protein can be considered a non-specific interactor and/or background contaminant. We performed this evaluation using CRAPome database and a dataset of 411 experiments; the higher this value is, the higher is the probability that the detected protein can be considered a background contaminant.

4.6.3 Global analysis of mass spectrometry data elaboration in mEpiLCs

Through all these instructions about the data elaboration, a global analysis can be performed. As mentioned above, Lin28A-overexpressed no RNase A-treated samples MS data have reported 487 candidate interactors, whereas the identified proteins found in RNase A-treated samples report were 493. Since the implementation of the filters previously described, we were able to select 201 proteins in Lin28A-overexpressed no RNase A-treated samples and 169 proteins in Lin28A-overexpressed RNase A-treated samples as Lin28A candidate interactors. Focusing our attention on global overview of these proteins, we observed that protein distribution is quite different from each other through lane 1, 2 and 3; this is easily in particular, observing the separation profile of ribosomal proteins in Table 6. This distribution suggests that the native protein separation by size-exclusion chromatography worked. Anyway, in no RNase A-treated sample, more than 60% of proteins found in lane 1 were lost in lane 1 in RNase A-treated sample; among these proteins, we observed that 22% of these proteins was shifted in lane 2 and a very small amount was shifted in lane 3 in RNase A treatment condition. Moreover, exactly the half of proteins found in lane 2 was completely lost in lane 2 in RNase A-treated sample and more than 80% proteins found in lane 3 were lost in lane 3 in RNase A-treated sample. In RNase A-treated sample, more than 80% proteins found in lane 1 are common among lane 1 proteins in no RNase

A-treated. Only less than 30% found in lane 2 and less than 20% found in lane 3 are common among lane 2 and lane 3 proteins in no RNase A-treated samples. In Table 7 it is shown all protein groups divided for the different size peaks obtained by immunoprecipitation, representing the different molecular weight protein complexes.

However, more than 75% proteins, found among the selected candidates in no Rnase A-treated data elaboration, was lost described to be sensitive on RNase A treatment, suggesting that most of candidates found may be Lin28A putative interactors through RNA-dependent bond.

	S	Lane 1	Lane 2	Lane 3
no RNAse-treated	**40S**	Rps15a	Rps13	absent
		Rps12	Rps5	
		Rps16		
		Rps13		
		Rps4x		
		Rps7		
		Rps5		
		Rps11		
		Rps19		
		Rps20		
	60S	Rpl13		
		Rpl10a		
		Rplp2		
		Rpl27		
		Rplp0		
		Rpl22l1		
		Rpl22		
		Rpl38		
	28S	Mrps22	Mrps28	
	39S	Mrpl21		
		Mrpl39		
Rnase-treated	**60S**	Rpl22l1	absent	absent
		Rpl13		
	28S	Mrps22		
		Mrps28		
	39S	Mrpl21		
		Mrpl39		

Table 6: Size-based separation profile of all ribosomal proteins found in mass spectrometry data reports.

Table 8 shows the comparison of our results with ribonucleoproteins found in the literature.

Moreover, an overview of gene ontology annotations about mass spectrometry data elaboration in mEpiLCs provides some information about their associated biological processes (BP), molecular functions (MF), and cellular components (CC). We focused our attention on the false discovery rate (FDR) of each categories and we considered several groups, including intracellular ribonucleoprotein complex, CRD-mediated mRNA stabilization, translation, poly(A) RNA binding, RNA binding, showing a significant FDR, as reported in Table 9 and 10.

Lanes	Elements
all	Ndufa6,Rfc2
1,1R,2,2R,3	Prpf8,Rfc4,Dcakd,Copa
1,1R,2,3,3R	Hnrnpu,Eif3l
1,2,2R,3,3R	Xpo1
1,1R,2,2R,3R	Slc25a13,Vars,Pfkl,Ndufs1
1,1R,2,3	Flnc
1,2,2R,3	Ddx17,Ddx5
1,2,3,3R	Psmd14
1,1R,2,2R	Galk1,Lrp2,Copb2,Smc2,Psmd3,Arid1a,Ptk7, Psmc6,Gcn1,Eif3f,Tomm70a,Smc4,Nomo1
1,2,2R,3R	Iqgap1,Msh2
1,1R,3,3R	Slc16a3
2,2R,3,3R	Ddx3x
1R,2,2R,3R	Mcm7,Cse1l
1,2,3	Copb1,Flnb,Myh10,Actl6a
1,1R,2	Ctbp2,Ruvbl2,Aimp2,Gm45927,Atp5c1, Sf3b3,Aimp1,Eif3e,Lars,Eif3i,Psmd2
1,2,2R	Mybbp1a,Eif3a,Rpl27,Dync1h1,Rpl13,Hspa9,Hnrnpm, Col18a1,Afdn,Pcna,Rdh11,Igf2bp1,Rps11,Rps4x
1,2,3R	Nmral1,Qars,Eif4a3l2,Mars,Hnrnpc,Rps20
1,2R,3	Rps13

Table 7.1

Lanes	Elements
1,1R,2R	Sf3b1,Lin28A,Eif3c,Slc3a2,Mrps22,Mrpl39
1,2R,3R	Atp2a2
1R,2,3	Eif3h
1R,2,2R	NARS1,Rpl22l1,Prmt1
2,2R,3R	Mcm5,Ogdhl
1R,2R,3	Mrps28,Dhx9
2R,3,3R	Myh4
1,2	Rps5,Rpl22,Eef1d,Iars,Eif3k,Ipo5,Psmc4,Rars,Trim28, Rps19,Cad,Rplp0,Pdhb,Tardbp,Flna,Psmd11, Mthfd1,Hnrnpf,Rps7,Rps12,Gm9392,Dars,Eprs
1,3	Myl12b
1,1R	Pcbp2,Suclg2,Atp1a1
1,2R	Eif4g1,Mrpl21,Eef1e1
1,3R	Rpn1,Atp5pd,Slc25a4
2,3	Psmb2,Gnb2
2,2R	Trap1,Kif5b,Ctnnd1,Prmt5,Smarca4,TF,Dsg4,Eif3b
2,3R	Pfkm,Aldh9a1,Napa,Pfkp,Hk2,Hat1
2R,3	Eif3m
3,3R	Prep,Plod3,Mettl16,Fbxl18,Myh7,Tnpo1,Etfa
1R,2R	Pzp,Cope
2R,3R	Eef1b
1	mKIAA0034,Myh9,Rps18ps5,Hsp90b1,Slc25a3, Pcbp1,Gnai2,Myl6,Rpsa,Vdac2,Slc25a5,Rpl38, Ddost,Atp5po,Rpl10a,Rps15a,Rack1, Eef1g,Rps16,Gnai3,Gm11214,Rplp2,Rpl12
2	Tmod3,Txnl1,GSN,Pa2g4,Ddx39b,Cdk1,Gnb1, Otub1,Glud1,Eif3g,Mthfd1l,Hnrnpk,Uba1, Krt12,Ddx39,Cdk4,Mat2a,Ttr,Idh2
3	Hnrnph1,Fubp3,Adk,Mob2,Lrrc40, Syncrip,Rheb,Uck2,Pfas
1R	Phb,Lin28a
2R	Ldha,Ldhb,Mrpl58,Ltf
3R	Map2k2,Capzb,Prpsap2,Acadm,Prpsap1, Upp1,Cnn3,Aars,Sucla2,Nsun2,Gfm1, Mat2b,H2ac13,Map2k4,Dus3l,Gcat, Tufm,H2A,Sae1,Nub1,Lmna,Impa1,Ahcy, Pof1b,Rrps14,Mtx1,Tomm34,Mtmr14,H2az2

Table 7.2

Tables 7: The proteins found in mass spectrometry data elaboration divided in several groups belonging to lane 1, 2 and 3. The letter R indicates the RNase-treated sample lanes.

mouse ESC RNP	**human cell RNP**	**human mRNA binding proteins**	**mouse ESC RNP**	**human cell RNP**	**human mRNA binding proteins**
DDX17	DDX17		IPO5	IPO5	
DDX39	DDX39		METTL16	METTL16	
DDX39B	DDX39B		MYBBP1A	MYBBP1A	
DDX3X	DDX3X	DDX3X	NOMO1		
DDX5	DDX5		NSUN2	NSUN2	
DHX9	DHX9	DHX9	PA2G4		
EEF1G			PCBP1	PCBP1	
EIF3A	EIF3A		PCBP2	PCBP2	
EIF3C	EIF3C		RPL10A	RPL10A	
EIF3E	EIF3E		RPL13		
EIF3G	EIF3G		RPL22	RPL22	
EIF4G1	EIF4G1		RPL22L1		
FLNA	FLNA		RPL27	RPL27	
FUBP3	FUBP3		RPS11		
GCN1L1	GCN1L1		RPS16	RPS16	
HK2			RPS19	RPS19	
HNRNPA1	HNRNPA1	HNRNPA1	RPS20	RPS20	
HNRNPA3	HNRNPA3	HNRNPA3	RPS4X	RPS4X	
HNRNPC	HNRNPC		RPS5	RPS5	
HNRNPF	HNRNPF	HNRNPF	RPS7	RPS7	
HNRNPH1	HNRNPH1	HNRNPH1	SYNCRIP	SYNCRIP	SYNCRIP
HNRNPK	HNRNPK	HNRNPK	TARDBP	TARDBP	
HNRNPL	HNRNPL	HNRNPL	TNPO1	TNPO1	
HNRNPM	HNRNPM	HNRNPM	XPO1		
IGF2BP1	IGF2BP1	IGF2BP1			

Table 8: Overlap of our data with known proteomes

Category	Term	Count	P-Value	Genes	FDR
GOTERM_CC_DIRECT	GO:0030529~intracellular ribonucleoprotein complex	42	1,15E-18	XPO1, RPL13, IGF2BP1, RPLP2, SYNCRIP, RPL22L1, RPL38, IQGAP1, HNRNPA3, HNRNPL, HNRNPM, HNRNPK, PRPF8, HNRNPF, RPLP0, PCBP1, PCBP2, RPL10A, HNRNPC, RPS20, RPL12, MRPL39, DHX9, MRPS28, MRPS22, RPL27, EPRS, RPS4X, DDX5, RPS5, HNRNPA1, HNRNPU, RPS7, PA2G4, MRPL21, RPS19, RPS16, RPL22, RPS13, RUVBL2, RPS11, HNRNPH1	0
GOTERM_BP_DIRECT	GO:0070934~CRD-mediated mRNA stabilization	4	1,12E+11	DHX9, SYNCRIP, IGF2BP1, HNRNPU	0.017
GOTERM_BP_DIRECT	GO:0075525~viral translational termination-reinitiation	4	1,12E+11	EIF3A, EIF3B, EIF3G, EIF3L	0.017
GOTERM_BP_DIRECT	GO:0006412~translation	47	1,05E-18	NARS, RPL13, RPS15A, RPL22L1, RPL38, VARS, IARS, EIF3C, EIF3A, EIF3B, EIF3G, EIF3H, EIF3E, EIF3F, EIF3K, SLC25A3, LARS, EIF3L, EIF3I, RPL10A, RPS20, RPL12, MRPL39, EIF3M, MARS, DARS, AIMP1, SLC25A4, AIMP2, RPL27, EPRS, RPS4X, RPS5, RPS7, EIF4G1, MRPL21, RPS19, SLC25A13, RPS16, RPL22, EEF1E1, RARS, RPS12, RPS13, EEF1G, RPS11, EEF1D	0
GOTERM_MF_DIRECT	GO:0044822~poly(A) RNA binding	67	4,48E-16	RPL13, SYNCRIP, DDX17, FUBP3, PRMT1, RPLP0, TARDBP, PCBP1, PCBP2, MRPL39, RPL12, DYNC1H1, DARS, DDX39B, RPS4X, MYH9, FLNB, HNRNPU, FLNA, EIF4G1, DDX39, TRAP1, PA2G4, RPS19, RPS16, IPO5, ATP5C1, RPS12, RPS13, RPS11, MYBBP1A, IGF2BP1, RPS15A, GCN1L1, EIF3C, HNRNPL, SF3B1, HNRNPM, EIF3A, HNRNPK, EIF3G, DDX3X, EIF3H, EIF3E, PRPF8, HNRNPF, EIF3L, RPL10A, RPS20, HNRNPC, TNPO1, NSUN2, HSPA9, DHX9, MRPS28, TRIM28, RPL27, DDX5, RPS5, HNRNPA1, RPS7, SLC16A3, MRPL21, UBA1, RPL22, HNRNPH1, METTL16	0
GOTERM_MF_DIRECT	GO:0003723~RNA binding	46	2,46E-05	XPO1, RPL13, IGF2BP1, SYNCRIP, QARS, RPL38, SF3B3, EIF3C, HNRNPA3, HNRNPL, HNRNPM, DDX17, EIF3A, EIF3B, HNRNPK, DDX3X, EIF3G, PRPF8, HNRNPF, PCBP1, TARDBP, PCBP2, EIF3K, RPL10A, HNRNPC, MRPL39, RPS20, RPL12, NSUN2, MARS, DHX9, AIMP1, DDX39B, EPRS, RPS4X, DDX5, RPS5, HNRNPA1, HNRNPU, EIF4G1, PA2G4, MRPL21, RPS16, RPL22, RPS11, HNRNPH1	0,03

Table 9: Gene ontology of proteins found in samples not treated with RNase

Category	Term	Count	P-Value	Genes	FDR
GOTERM_BP_DIRECT	GO:0075525~viral translational termination-reinitiation	4	6,22E+09	EIF3A, EIF3B, EIF3G, EIF3L	0.009
GOTERM_BP_DIRECT	GO:0070934~CRD-mediated mRNA stabilization	4	6,22E+09	DHX9, SYNCRIP, IGF2BP1, HNRNPU	0.009
GOTERM_BP_DIRECT	GO:0075522~IRES-dependent viral translational initiation	4	2,15E+10	EIF3A, EIF3B, PCBP2, EIF3F	0.033
GOTERM_MF_DIRECT	GO:0031369~translation initiation factor binding	6	2,97E+10	EIF3C, EIF3B, DDX3X, EIF3F, LIN28A, EIF3M	0.004
GOTERM_MF_DIRECT	GO:0000287~magnesium ion binding	12	2,98E+10	PKM, IMPA1, STK38, MSH2, NUDT5, PPM1A, IDH2, PRPS1L3, IDH1, PRPS2, IDH3A, PRPSAP2	0.004
GOTERM_MF_DIRECT	GO:0004386~helicase activity	9	2,74E+11	DHX9, DDX17, MCM7, DDX3X, RUVBL2, GM5580, DDX5, MCM5, SMARCA4	0.038

Table 10: Gene ontology of proteins found in samples treated with RNase

4.7 Putative Lin28A interactors screening, validations and analyses

As mentioned above, the overexpression of Lin28A results in the increase of the luciferase expression, driven by a construct where the 3'UTR of the Dnmt3a mRNA was cloned downstream of the luciferase coding sequence (Figure 17B). We considered the possibility that the effects of Lin28A over expression could be the consequence of the function of molecular complexes containing Lin28A that we have purified. Of course, the proteins identified by MS, are candidates to be partners of Lin28 in these molecular complexes, thus we reasoned that the downregulation of these Lin28A candidate partners should modify (increase or decrease) the effects of Lin28A overexpression on the 3'UTR of Dnmt3a mRNA. Taking advantage of the availability of an shRNA library targeting a large number of mouse genes, we designed a small-scale screening of the shRNAs which target the mRNAs coding for the proteins identified by our purification/MS experiments. In particular, we chose the candidates obtained from mass spectrometry data elaboration of no RNase A-treated Lin28A overexpression fractions. Moreover, among these candidates, we used 113 shRNAs only on the following grounds:

-Lin28A interactors shRNAs availability in the library

-some Lin28A interactors could be, in this context, more interesting than others.

Moreover, we decided to study these effects at 3day of epiblast stem cell differentiation, exactly when we observed the accumulation of Lin28A protein during the establishment of the epiblast phenotype.

For this purpose, we used 113 shRNAs each targeting one of the putative Lin28A partners, were co-transfected, in EpiLCs at day 2, with Lin28A-Flag or with the empty vector and the plasmid

containing the Dnmt3A 3'UTR downstream of the luciferase coding sequence. The shRNA that targets the GFP was used as a control non-silencing shRNA. After 24h from the transfection, the luciferase activity was measured using Renilla as an internal control. The results of the screening are shown in Figure 22 and are reported as the fold change observed in mock transfected cells exposed to control shRNA vs the shRNA targeting the indicated proteins (black columns) and the fold change observed in Lin28A transfected cells exposed to control shRNA vs the shRNA targeting the indicated proteins. The redline indicates the value 1, that is the change observed when the control shRNA was transfected (GFP). All the experiments were run in biological independent triplicates.

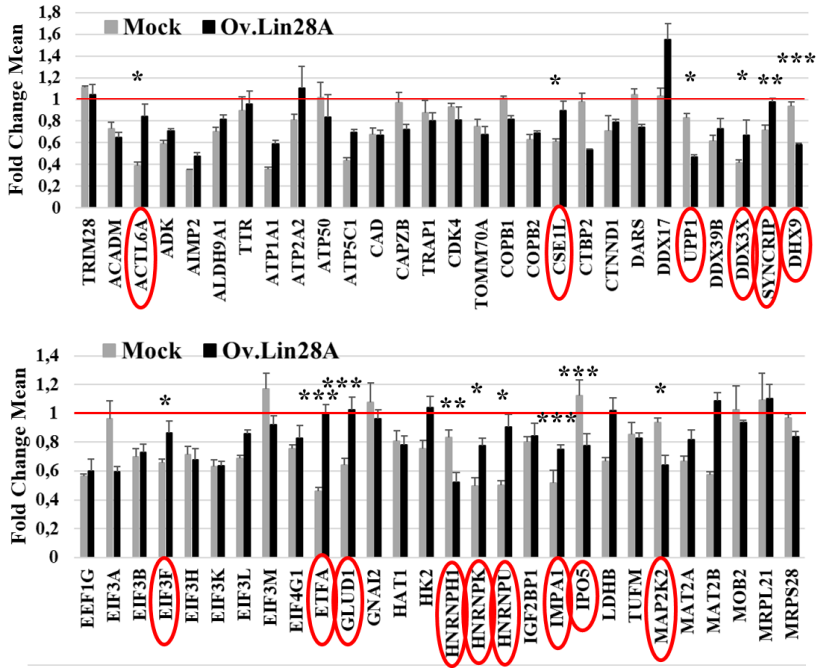


Figure 22.1

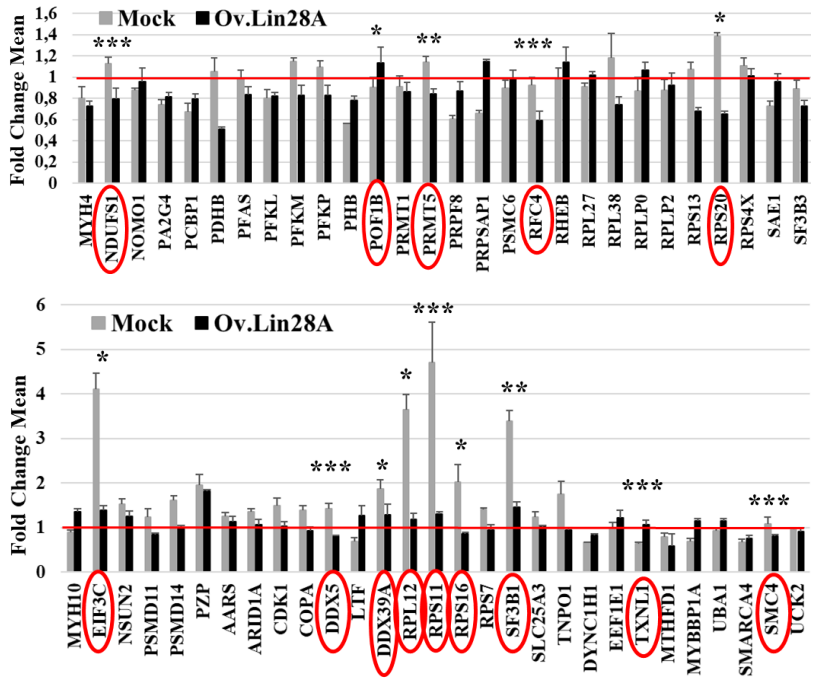


Figure 22.2

Figures 22.1 and 22.2: Effects of Lin28A putative interactor on Dnmt3A translation. *EpiLCs are transfected with plasmid expressing Lin28a-Flag or empty vector, with a construct containing Dnmt3A 3'UTR downstream of the luciferase coding sequence and shRNAs against mRNAs encoding Lin28A putative interactors. Cell extracts are used to measure the luciferase activity, normalized to Renilla luciferase. The asterisks indicate statistically significant differences (*p≤0,05, ** p≤0,02 and *** p≤0,001).*

The proteins whose silencing affects the accumulation of the luciferase driven by Luc-Dnmt3a construct were divided into four groups:

1) proteins having a positive effect on Luc-Dnmt3A translation independent from the effects of Lin28A overexpression; indeed, the silencing of these proteins

reduces the luciferase expression in mock transfected cells and Lin28A overexpression counteracts this reduction, thus maintaining its effects. These proteins are: ACTL6A, CSE1L, EIF3F, ETFA, GLUD1, HNRNRPU, POF1B, SYNCRIP and TXNL1.

2) proteins having a positive effect on Dnmt3A translation regardless of Lin28A function; indeed, their silencing reduces the luciferase expression in mock and in Lin28A transfected cells. These proteins are: DDX3X, HNRNPK and IMPA1.
3) proteins having a positive effect on Luc-Dnmt3A translation in a Lin28A-dependent manner; indeed, their silencing has no effects in mock transfected cells, while it significantly reduces the luciferase expression in Lin28A overexpressing cells. The proteins belonging to this group are: DHX9, RFC4, UPP1, HNRNPH1 and MAP2K2.
4) proteins having a negative effect on Luc-Dnmt3A translation that is counteracted by Lin28A; in particular, their silencing increases the luciferase expression in mock transfected cells but not in Lin28A overexpressing cells. These proteins are: DDX39A, DDX5, EIF3C, IPO5, NDUFS1, PRMT5, RPL12, RPS11, RPS16, RPS20, SF3B1.

4.8 Protein-protein association analyses

To further analyze the relationships between the selected candidate interactors and Lin28A protein, based on the information obtained from luciferase assays, we decided to get further information through a functional enrichment analysis based on known and predicted protein–protein interactions performed by using the String database. In particular, we

focused on some proteins belonging to the third group mentioned above, that is the protein group having a positive effect on Luc-Dnmt3A translation in a Lin28A-dependent manner. The results shown in Figure 23 A and B are obtained carrying out a protein interaction research between Lin28A and Map2k2 and between Lin28A, Hnrnph1 and Dhx9 respectively. As shown in Figure 23A, the association between Lin28A and Map2k2, also called Mek2, a member of MAP kinase kinase family, could take place through the already known interaction, experimentally determined, of both proteins with a Ras oncogene family member, H-Ras. Afterwards we investigated the association between Lin28A, Hnrnph1 and Dhx9 and, as shown in Figure 23B, the interaction of Dhx9, a DEAH-containing family of RNA helicases, with Lin28A is already experimentally determined as well as that one with Hnrnph1, a component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complexes. This result suggests that Lin28A and Hnrnph1 could be associated through the already known interaction with Dhx9.

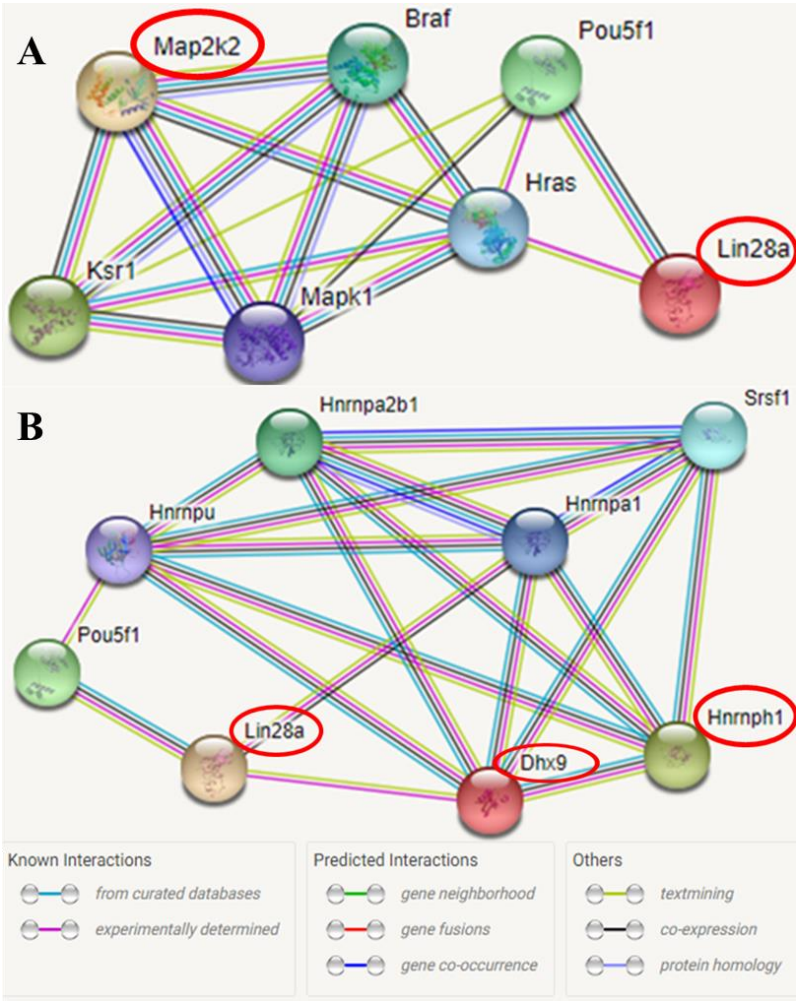


Figure 23: Map2k2-Lin28A association (A) and Hnrnph1-Dhx9-Lin28A association (B) by String analysis.

5. DISCUSSION

In this study, we have examined the Lin28A protein, a RNA binding protein, that gradually accumulates during the establishment of the epiblast phenotype, in particular at day 3 of epiblast stem cell differentiation. In order to find new mRNAs targets of Lin28A, in addition to those already known, like Hmga2, Oct4 and Igf2, we investigated several proteins involved in the embryonic stem cell differentiation and playing a significant role in epiblast stem cell establishment. Furthermore, we were seeking of putative targets containing in their own transcript the Lin28 consensus sequence GGAG, as observed for all Lin28 mRNA targets. Therefore, we found this consensus sequence in 3'UTR mRNA of Dnmt3A, a gene encoding a *de novo* DNA methyltransferase involved in DNA methylation pattern setting of primed pluripotent state.

Firstly, we observed that the expression of Dnmt3A transcript and both two different splicing products, DNMT3A1 and DNMT3A2, is increased during the transition from naïve to primed pluripotent state. Then, we investigated whether Lin28A protein is able to recognize and directly bind the consensus sequence Dnmt3A1/2 and we observed an abundant presence of DNMT3A mRNA in Lin28A immunoprecipitation. Thereafter, in order to assess the effects of Lin28A on Dnmt3A mRNA translation, we used a construct that drives the expression of a luciferase reporter bearing the Dnmt3a mRNA 3′ UTR containing Lin28A consensus motif. The co-transfection of this construct together with the plasmid driving the expression of Lin28A-Flag resulted in a robust induction of the luciferase signal.

At this point, since Lin28 is known to controls negatively Hmga2 translation and positively the one of Oct4 and Igf2, we were wondering what kind of regulation Lin28A is able to perform on Dnmt3A mRNA and, more generally, which are the

molecular mechanisms underlying the post transcriptional regulation mediated by Lin28A. In particular, we were wondering whether Lin28A partners exist, able to affect its function.

For this reason, we developed an experimental setting to identify Lin28A partners in Hela cell lines and we optimized it in mEpiLCs. The experimental setting steps are:

-the fractionation of cell extracts by size-exclusion chromatography in native conditions in order to obtain a size-based molecular complexes segregation

-the immunoprecipitation of fractions containing Lin28A in order to isolate the proteins able to bind directly or indirectly Lin28A

-the mass spectrometry analysis and elaboration in order to identify putative Lin28A interactors

-finally a screening of the putative interactors by exploring the effects of their suppression on DNMT3A mRNA translation mediated by Lin28A, in order to select proteins involved in Lin28A activity.

We actually found some protein able to do this job. Firstly, we found proteins having a positive effect on Dnmt3A translation in a Lin28A-dependent manner, such as: DHX9, RFC4, UPP1, HNRNPH1 and MAP2K2. We hypothesized that these proteins could be a part of Lin28A-containing molecular complexes or, at least, proteins necessary for their function; even so, we can certainly argue that Lin28A requires the proteins mentioned above in order to play its role on DNMT3A expression regulation. Further protein-protein interaction analyses performed by String database have shown the association between Lin28A and Map2k2; both proteins interact with H-Ras. Map2k2, also called Mek2, is a kinase that phosphorylates and thus activates ERK2 and ERK3. As already mentioned, Lin28A phosphorylation occurs through ERK kinases in serine

200 residue enhancing its function and promoting the transition from naïve to primed pluripotent state (Tsanov 2017). As shown in Table 7, we found Map2k2 as putative interactor in lane 3R, corresponding to RNase A-treated fraction 23-24 and mL 17, exactly where albumin (70KDa) is eluted. This means it is possible that in these fractions it exists a molecular size complex composed of Lin28A (30KDA) and Map2k2 (44KDa). Moreover, this information suggests that this interaction could:

-occur when Lin28A doesn't bind any RNA molecules

-promote Lin28A phosphorylation in serine 200 residue through ERKs

-be critical for DNMT3A expression regulation mediated by Lin28A.

Afterwards we found, through String database, a protein-protein association between Lin28A, Hnrnph1 and Dhx9. In particular, the interaction between Lin28A C-terminal portion and Dhx9 is already known and it occurs in a RNA-independent manner. This protein could stimulate LIN28A-dependent mRNA translation probably by facilitating ribonucleoprotein remodeling during the translation event. In accordance with that, we found Dhx9 as putative interactor in RNase A-treated lane 1R, 2R and in RNase A-no treated lane 3, as shown in Table 7. Moreover, the results of gene ontology research, performed with all Lin28A putative interactors found in RNase A-treated samples, classify Dhx9 as one of proteins belonging to CRD-mediated mRNA stability complex, as shown in Table 10. DHX9 binds to and promotes the stabilization of all mRNA molecules containing the coding region instability determinant (CRD). All this information, taken together, suggests that DHX9 bound to Lin28A could be critical for the Dnmt3A mRNA stabilization.

Moreover, we found another groups of proteins having a negative effect on Dnmt3A translation that is counteracted by Lin28A, such as DDX39A, DDX5, EIF3C, IPO5, NDUFS1,

PRMT5, RPL12, RPS11, RPS16, RPS20, SF3B1. In particular, it seems that Lin28A can substitute these proteins because they have the same effect on Dnmt3A mRNA translation. Another option could be that, more generally, Lin28A supports their function. Indeed, we can hypothesize that the expression of these proteins could be affected by itself Lin28A overexpression that balances their silencing.

These preliminary evaluations provide the proof of concept that Lin28A doesn't work alone when regulating the translation of its targets and that some interactor could be critical for the translational fate of mRNAs targeted by Lin28A.

6. CONCLUSIONS

The relevant role of the RNA binding protein Lin28A in the first steps of mammalian development and in the reprogramming of adult cells into iPS cells is already known. Here, we focused our attention on role of Lin28A during the epiblast stem cell differentiation. In particular, we found a new Lin28A mRNA target, Dnmt3a, and we observed a positive post-transcriptional regulation of its expression, in contrast to the Lin28 negative effect on Hmga2 mRNA. Furthermore, we explored the mechanisms through which Lin28A exerts this function on its targets assuming that Lin28A doesn't work alone but instead its interactors are critical for the translational fate of Lin28A target mRNAs. Through the development of an experimental setting for size-based protein segregation and immunoprecipitation to identify Lin28A partners, we preliminary evaluated the contribution of these putative interactors on Lin28A function. We assumed, in conclusion, that Map2k2 and Dhx9 proteins could be critical interactors on Lin28A-mediated post transcriptional regulation of Dnmt3a expression.

7. APPENDICES

Table A1. Primer list

Fgf5: Fw: TCCATGCAAGTGCCAAATTTACGGA
Rv: TTCTGTGGATCGCGGACGCA

Gapdh: Fw: GTATGACTCCACTCACGGCAAA
Rv: TTCCCATTCTCGGCCTTG

Lin28a: Fw: GTTCGGCTTCCTGTCTATGACC
Rv: CTTCCATGTGCAGCTTGCTCT

Sox1: Fw: ACCGCTACGACATGGGC
Rv: GCTCATGTAGCCCTGAGAGT

Dnmt3a: Fw: CACCCCTGAGCCAGTAGGAG
Rv: GGCCATCCTCATACTCAGGC

Table A2. Parameters of Superose 6 Increase 10/300 GL column

Superose 6 Increase 10/300 GL	
Matrix	Composite of cross-linked agarose
Average particle size	8.6 µm
Exclusion limit (Mr)	Approx. $4x10^7$
Optimum separation range for globular proteins	5000 to 5×10^6
Approximate bed volume (ml)	24
Column hardware pressure limit	5 MPa
Recommended flow rate (ml/min)	0.5

Table A3. Size exclusion profile data of Carbonic Anydrase

mAU	ml	Fractions
0,94	5,99	1
0,97	6,49	2
0,48	6,99	3
1,12	7,49	4
0,89	7,99	5
1,24	8,49	6
0,79	8,99	7
1,22	9,49	8
0,85	9,99	9
1,23	10,49	10
1,09	10,99	11
1,14	11,49	12
1,21	11,99	13
0,64	12,49	14
1,12	12,99	15
1,36	13,49	16
1,36	13,99	17
1,89	14,49	18
1,34	14,99	19
1,74	15,49	20
1,12	15,99	21
1,17	16,49	22
2,06	16,99	23
3,24	17,49	24
42,33	17,99	25
24,95	18,49	26
4,98	18,99	27
2,17	19,49	28
1,91	19,99	29
1,78	20,49	30
2,17	20,99	Waste

Table A4. Size-exclusion profile data of thyroglobulin, apoferritin and albumin as molecular size markers.

Thyroglobulin			Apoferritin			Albumin		
mAU	mL	fractions	mAU	mL	fractions	mAU	mL	fractions
1,01	5,99	1	2,40	6,00	1	0,54	5,99	1
0,99	6,49	2	2,41	6,50	2	0,91	6,49	2
0,67	6,99	3	2,75	7,00	3	0,89	6,99	3
1,40	7,49	4	2,23	7,50	4	0,98	7,49	4
5,98	7,99	5	2,02	8,00	5	0,45	7,99	5
7,20	8,49	6	2,43	8,50	6	1,13	8,49	6
3,93	8,99	7	2,19	9,00	7	1,02	8,99	7
2,25	9,49	8	2,02	9,50	8	1,15	9,49	8
2,48	9,99	9	2,23	10,00	9	1,51	9,99	9
3,96	10,49	10	3,16	10,50	10	0,70	10,49	10
6,54	10,99	11	3,36	11,00	11	1,19	10,99	11
11,99	11,49	12	5,05	11,50	12	1,47	11,49	12
12,83	11,99	13	8,91	12,00	13	1,31	11,99	13
14,02	12,49	14	11,54	12,50	14	1,64	12,49	14
67,97	12,99	15	28,02	13,00	15	1,47	12,99	15
58,34	13,49	16	11,52	13,50	16	1,47	13,49	16
22,35	13,99	17	10,29	14,00	17	2,43	13,99	17
15,07	14,49	18	151,88	14,50	18	3,60	14,49	18
7,76	14,99	19	73,65	15,00	19	5,03	14,99	19
6,73	15,49	20	11,93	15,50	20	7,87	15,49	20
8,85	15,99	21	3,40	16,00	21	3,61	15,99	21
7,57	16,49	22	1,67	16,50	22	24,22	16,49	22
5,72	16,99	23	1,87	17,00	23	4,02	16,99	23
5,32	17,49	24	1,86	17,50	24	1,29	17,49	24
5,43	17,99	25	1,48	18,00	25	1,66	17,99	25
4,74	18,48	26	2,29	18,50	26	1,27	18,49	26
5,00	18,98	27	4,38	19,00	27	0,96	18,99	27
4,50	19,48	28	7,73	19,49	28	1,50	19,49	28
3,12	19,98	29	6,08	19,99	29	1,49	19,99	29
2,82	20,48	30	4,04	20,49	30	1,08	20,49	30
2,95	20,98	Waste	5,18	20,99	Waste	1,16	20,99	Waste

Table A5. Size-exclusion profile data of EpiLCs w/o RNase A protein extracts and apoferritin.

EpiLCs			EpiLCs with Rnase A			Apoferritin		
mAU	ml	Fractions	mAU	ml	Fractions	mAU	ml	Fractions
0,69	5,99	1	1,71	5,99	1	4,46	5,99	1
0,86	6,49	2	1,16	6,49	2	3,79	6,49	2
0,74	6,99	3	0,84	6,99	3	3,61	6,99	3
0,84	7,49	4	0,81	7,49	4	3,34	7,49	4
1,08	7,99	5	1,10	7,99	5	3,24	7,99	5
16,91	8,49	6	15,36	8,49	6	2,95	8,49	6
71,32	8,99	7	24,19	8,99	7	2,31	8,99	7
51,79	9,49	8	21,36	9,49	8	2,38	9,49	8
36,12	9,99	9	35,27	9,99	9	2,88	9,99	9
29,12	10,49	10	42,13	10,49	10	2,50	10,49	10
24,80	10,99	11	36,03	10,99	11	3,58	10,99	11
23,61	11,49	12	29,68	11,49	12	3,90	11,49	12
21,56	11,99	13	25,76	11,99	13	6,90	11,99	13
19,70	12,49	14	20,63	12,49	14	6,47	12,49	14
18,93	12,99	15	18,28	12,99	15	17,87	12,99	15
19,61	13,49	16	18,40	13,49	16	12,80	13,49	16
23,72	13,99	17	22,17	13,99	17	2,94	13,99	17
27,17	14,49	18	21,89	14,49	18	43,53	14,49	18
42,65	14,99	19	28,79	14,99	19	98,42	14,99	19
51,84	15,49	20	44,22	15,49	20	6,73	15,49	20
55,09	15,99	21	51,00	15,99	21	2,06	15,99	21
83,69	16,49	22	53,72	16,49	22	1,46	16,49	22
77,37	16,99	23	73,93	16,99	23	1,63	16,99	23
62,78	17,49	24	81,68	17,49	24	1,57	17,49	24
34,92	17,99	25	63,68	17,99	25	1,81	17,99	25
29,82	18,49	26	51,58	18,49	26	1,73	18,49	26
27,12	18,99	27	53,15	18,99	27	2,61	18,99	27
34,59	19,49	28	178,39	19,49	28	4,81	19,49	28
71,58	19,99	29	183,06	19,99	29	2,93	19,99	29
39,11	20,49	30	48,67	20,49	30	1,73	20,49	30
45,30	20,99	Waste	48,23	20,99	Waste	2,97	20,99	Waste

8. REFERENCES

Ambros V, Horvitz HR. Heterochronic mutants of the nematode Caenorhabditis elegans. Science. 1984;226(4673):409-16.

Ambros V. A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans. Cell. 1989 Apr 7;57(1):49-57.

Artus J, Chazaud C. A close look at the mammalian blastocyst: epiblast and primitive endoderm formation. Cell Mol Life Sci. 2014;71(17):3327-38.

Auernhammer CJ, Melmed S. Leukemia-inhibitory factor-neuroimmune modulator of endocrine function. Endocr Rev. 2000;21(3):313-45.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 2003;17(1):126-40.

Balzeau J, Menezes MR, Cao S, Hagan JP. The LIN28/let-7 Pathway in Cancer. Front Genet. 2017;8:31.

Balzer E, Moss EG. Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. RNA Biol. 2007;4(1):16-25.

Boiani M, Schöler HR. Regulatory networks in embryo-derived pluripotent stem cells. Nat Rev Mol Cell Biol. 2005;6(11):872-84.

Boulton TG, Zhong Z, Wen Z, Darnell JE Jr, Stahl N, Yancopoulos GD. STAT3 activation by cytokines utilizing gp130 and related transducers involves a secondary modification requiring an H7-sensitive kinase. Proc Natl Acad Sci U S A. 1995; 92(15):6915–19.

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-

Richter L, Pedersen RA, Vallier L. Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature. 2007;448(7150):191-5.

M. Rossbach, Factsheet EuroStemCell 5.10.2010

Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, Wysocka J. Reorganization of enhancer patterns in transition from naïve to primed pluripotency. Cell Stem Cell. 2014; 14(6): 838–853.

Büssing I, Slack FJ, Grosshans H. let-7 microRNAs in development, stem cells and cancer. Trends Mol Med. 2008;14(9):400-9.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet. 2007;39(3):391-6.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 2003;23(16):5594-605.

Cheng X, Blumenthal RM. Mammalian DNA methyltransferases: a structural perspective. Structure. 2008;16(3):341-50.

De Vasconcellos JF, Fasano RM, Lee YT, Kaushal M, Byrnes C, Meier ER, Anderson M, Rabel A, Braylan R, Stroncek DF, Miller JL. LIN28A expression reduces sickling of cultured human erythrocytes. PLoS One. 2014;9(9):e106924.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 1982;10(8):2709-21.

Ernst M, Jenkins BJ. Acquiring signalling specificity from the cytokine receptor gp130. Trends Genet. 2004;20(1):23-32.

Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981;292(5819):154-6

Fernandez AF, Huidobro C, Fraga MF. De novo DNA methyltransferases: oncogenes, tumor suppressors, or both? Trends Genet. 2012;28(10):474-9.

Fusco A, Fedele M. Roles of HMGA proteins in cancer. Nat Rev Cancer. 2007;7(12):899-910.

Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol. 2009;16(10):1021-5.

Halbeisen RE, Galgano A, Scherrer T, Gerber AP. Post-transcriptional gene regulation: from genome-wide studies to principles. Cell Mol Life Sci. 2008;65(5):798-813.

Hamano R, Miyata H, Yamasaki M, Sugimura K, Tanaka K, Kurokawa Y, Nakajima K, Takiguchi S, Fujiwara Y, Mori M, Doki Y. High expression of Lin28 is associated with tumour aggressiveness and poor prognosis of patients in oesophagus cancer. Br J Cancer. 2012; 106(8): 1415–1423.

Hanna J, Saha K, Pando B, Van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature. 2009;462(7273):595-601.

Hartman TR, Qian S, Bolinger C, Fernandez S, Schoenberg DR, Boris-Lawrie K. RNA helicase A is necessary for translation of selected messenger RNAs. Nat Struct Mol Biol. 2006;13(6):509-16.

Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol. 2018;19(5):327-341.

Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell. 2008;32(2):276-84.

Hershey JW, Sonenberg N, Mathews MB. Principles of Translational Control: An Overview. Cold Spring Harb Perspect Biol. 2012;4(12): a011528.

Hilton DJ, Gough NM. Leukemia inhibitory factor: a biological perspective. J Cell Biochem. 1991;46(1):21-6.

Jackson M, Krassowska A, Gilbert N, Chevassut T, Forrester L, Ansell J, Ramsahoye B. Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol Cell Biol. 2004;24(20):8862-71.

Jin J, Jing W, Lei XX, Feng C, Peng S, Boris-Lawrie K, Huang Y. Evidence that Lin28 stimulates translation by recruiting RNA helicase A to polysomes. Nucleic Acids Res. 2011; 39(9): 3724–3734.

Karwacki-Neisius V, Göke J, Osorno R, Halbritter F, Ng JH, Weiße AY, Wong FC, Gagliardi A, Mullin NP, Festuccia N, Colby D, Tomlinson SR, Ng HH, Chambers I. Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell. 2013;12(5):531-45.

Kawahara H, Okada Y, Imai T, Iwanami A, Mischel PS, Okano H. Musashi1 cooperates in abnormal cell lineage protein 28 (Lin28)-mediated let-7 family microRNA biogenesis in early neural differentiation. J Biol Chem. 2011;286(18):16121-30.

Lee SH, Cho S, Kim MS, Choi K, Cho JY, Gwak HS, Kim YJ, Yoo H, Lee SH, Park JB, Kim JH. The ubiquitin ligase human TRIM71 regulates let-7 microRNA biogenesis via modulation of Lin28B protein. Biochim Biophys Acta. 2014;1839(5):374-86

Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev. 2007;21(9):1025-30.

Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, Mansfield W, Nashun B, Knezovich JG, Smith A, Surani MA, Hajkova P. Naïve pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol. 2013;20(3):311-6.

Loughlin FE, Gebert LF, Towbin H, Brunschweiger A, Hall J, Allain FH. Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. Nat Struct Mol Biol. 2011;19(1):84-9

Mayr F, Heinemann U. Mechanisms of Lin28-mediated miRNA and mRNA regulation--a structural and functional perspective. Int J Mol Sci. 2013;14(8):16532-53.

Mayr F, Schütz A, Döge N, Heinemann U. The Lin28 cold-shock domain remodels pre-let-7 microRNA. Nucleic Acids Res. 2012;40(15):7492-506.

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübeler D. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell. 2008;30(6):755-66.

Moss EG, Lee RC, Ambros V. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell. 1997;88(5):637-46.

Mullen AC, Orlando DA, Newman JJ, Lovén J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. Master Transcription Factors determine Cell-Type-Specific Responses to TGF-b Signaling. Cell. 2011;147(3):565–576.

Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell. 2008;132(4):661-80.

Nakaki F, Hayashi K, Ohta H, Kurimoto K, Yabuta Y, Saitou M. Induction of mouse germ-cell fate by transcription factors in vitro. Nature. 2013;501(7466):222-6.

Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P. Molecular basis for interaction of let-7 microRNAs with Lin28. Cell. 2011;147(5):1080-91.

Navarra A, Musto A, Gargiulo A, Petrosino G, Pierantoni GM, Fusco A4, Russo T, Parisi S. Hmga2 is necessary for Otx2-dependent exit of embryonic stem cells from the pluripotent ground state. BMC Biol. 2016;14:24.

Newman MA, Thomson JM, Hammond SM. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. RNA. 2008;14(8):1539-49.

Nichols J, Smith A. The origin and identity of embryonic stem cells. Development. 2011;138(1):3-8.

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H, Smith A. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. 1998;95(3):379-91.

Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet. 2000;24(4):372-6.

Niwa H, Ogawa K, Shimosato D, Adachi K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature. 2009;460(7251):118-22.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247-57.

Orphanides G, Reinberg D. A unified theory of gene expression. Cell. 2002;108(4):439-51.

Parisi S, Passaro F, Russo L, Musto A, Navarra A, Romano S, Petrosino G, Russo T. Lin28 is induced in primed embryonic stem cells and regulates let-7-independent events. FASEB J. 2017;31(3):1046-1058.

Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, Iliopoulos D, Gregory RI. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. Cell. 2011;147(5):1066-79.

Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, Gregory RI. Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. J Biol Chem. 2008;283(31):21310-4.

Polesskaya A, Cuvellier S, Naguibneva I, Duquet A, Moss EG, Harel-Bellan A. Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. Genes Dev. 2007;21(9):1125-38.

Priscilla M. Van Wynsberghe, Zoya S. Kai, Katlin B. Massirer, Victoria H. Burton, Gene W. Yeo and Amy E. Pasquinelli. LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in C. elegans. Nat Struct Mol Biol. 2011; 18(3): 302–308.

Qiu C, Ma Y, Wang J, Peng S, Huang Y. Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. Nucleic Acids Res. 2010; 38(4): 1240–1248.

Radzisheuskaya A, Silva JC. Do all roads lead to Oct4? the emerging concepts of induced pluripotency. Trends Cell Biol. 2014;24(5):275-84.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901-6.

Richards M, Tan SP, Tan JH, Chan WK, Bongso A. The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells. 2004;22(1):51-64.

Robertson KD, Wolffe AP. DNA methylation in health and disease. Nat Rev Genet. 2000;1(1):11-9.

Rogers MB, Hosler BA, Gudas LJ. Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. Development. 1991 Nov;113(3):815-24.

Rosner MH, Vigano MA, Ozato K, Timmons PM, Poirier F, Rigby PW, Staudt LM. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature. 1990;345(6277):686-92.

Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, Wulczyn FG. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat Cell Biol. 2008;10(8):987-93.

Sang H, Wang D, Zhao S, Zhang J, Zhang Y, Xu J, Chen X, Nie Y, Zhang K, Zhang S, Wang Y, Wang N, Ma F, Shuai L, Li Z, Liu N. Dppa3 is critical for Lin28a-regulated ES cells naïve–primed state conversion. J Mol Cell Biol. 2019;11(6):474-488.

Shyh-Chang N, Zhu H, Yvanka de Soysa T, Shinoda G, Seligson MT, Tsanov KM, Nguyen L, Asara JM, Cantley LC, Daley GQ. Lin28 enhances tissue repair by reprogramming cellular metabolism. Cell. 2013;155(4):778–792.

Silva J, Smith A. Capturing pluripotency. Cell. 2008;132(4):532-6.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76.

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD. New cell lines from mouse

epiblast share defining features with human embryonic stem cells. Nature. 2007;448(7150):196-9.

Teslaa T, Teitell MA. Pluripotent stem cell energy metabolism: an update. EMBO J. 2015;34(2):138-53.

Theunissen TW, Silva JC. Switching on pluripotency: a perspective on the biological requirement of Nanog. Philos Trans R Soc Lond B Biol Sci. 2011;366(1575):2222-9.

Tsanov KM, Pearson DS, Wu Z, Han A, Triboulet R, Seligson MT, Powers JT,Osborne JK, Kane S, Gygi SP, Gregory RI, Daley GQ. LIN28 phosphorylation by MAPK/ERK couples signalling to the post-transcriptional control of pluripotency. Nat Cell Biol. 2017;19(1):60-67.

Tsialikas J, Romer-Seibert J. LIN28: roles and regulation in development and beyond. Development. 2015;142(14):2397-2404

Veillard AC, Marks H, Bernardo AS, Jouneau L, Laloë D, Boulanger L, Kaan A, Brochard V, Tosolini M, Pedersen R, Stunnenberg H, Jouneau A. Stable Methylation at Promoters Distinguishes Epiblast Stem Cells from Embryonic Stem Cells and the in Vivo Epiblasts Stem Cells Dev. 2014;23(17):2014-29.

Wang LX, Wang J, Qu TT, Zhang Y, Shen YF. Reversible acetylation of Lin28 mediated by PCAF and SIRT1. Biochim Biophys Acta. 2014;1843(6):1188-95

Wilbert ML, Huelga SC, Kapeli K, Stark TJ, Liang TY, Chen SX, Yan BY, Nathanson JL, Hutt KR, Lovci MT, Kazan H, Vu AQ, Massirer KB, Morris Q, Hoon S, Yeo GW. LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol Cell. 2012;48(2):195-206.

Xu B, Zhang K, Huang Y. Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. RNA. 2009;15(3):357-61.

Yang DH, Moss EG. Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. Gene Expr Patterns. 2003;3(6):719-26.

Yao YL, Yang WM, Seto E. Regulation of transcription factor YY1 by acetylation and deacetylation. Mol Cell Biol. 2001;21(17):5979-91.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007;318(5858):1917–20.

Zhou X, Benson KF, Ashar HR, Chada K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature. 1995;376(6543):771-4.

Zhu H, Shyh-Chang N, Segrè AV, Shinoda G, Shah SP, Einhorn WS, Takeuchi A, Engreitz JM, Hagan JP, Kharas MG, Urbach A, Thornton JE, Triboulet R, Gregory RI; The Lin28/let-7 axis regulates glucose metabolism. Cell. 2011;147(1):81-94.

9. LIST OF PUBBLICATIONS

Castaldo D. et al. Molecular mechanisms of Lin28A-dependent regulation of translation. Manuscript in preparation.

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 "Dottorati Innovativi con caratterizzazione Industriale"