UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXI CICLO DEL DOTTORATO DI RICERCA IN BIOMEDICINA MOLECOLARE

Oct4 pseudogene lncRNA: dissecting its role in embryonic stem cell differentiation and epigenetic silencing of its ancestral gene *in trans*

Settore scientifico disciplinare: **BIO/11**

DOTTORANDO
MASSIMO ROSSO

COORDINATORE
PROF. GERMANA MERONI

SUPERVISORE DI TESI
PROF. STEFAN SCHOEFTNER

CO-SUPERVISORE DI TESI
DOTT.SSA ROBERTA BENETTI

ANNO ACCADEMICO: 2017/2018

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXI CICLO DEL DOTTORATO DI RICERCA IN BIOMEDICINA MOLECOLARE

Oct4 pseudogene lncRNA: dissecting its role in embryonic stem cell differentiation and epigenetic silencing of its ancestral gene *in trans*

Settore scientifico disciplinare: **BIO/11**

DOTTORANDO
MASSIMO ROSSO

COORDINATORE
PROF. GERMANA MERONI

SUPERVISORE DI TESI
PROF. STEFAN SCHOEFTNER

CO-SUPERVISORE DI TESI
DOTT.SSA ROBERTA BENETTI

ANNO ACCADEMICO: 2017/2018

Table of contents

Abstract 1

1. Introduction 2

1.1. Pseudogenes 3

1.1.1 Types of pseudogenes 3

1.1.2 Pseudogene number and expression 6

1.1.3 Pseudogene function and biological role 8

1.1.4 The boundary between functional pseudogenes and genes 17

1.2. Oct4 (Octamer-binding transcription factor 4) 18

1.2.1 Oct4 and cancer 21

1.2.2 OCT4 and ovarian cancer 22

1.2.3 Oct4 pseudogenes 24

1.3. FUS RNA binding protein 27

1.3.1 FUS as Transcription regulator 28

1.3.2 FUS as regulator of pre-mRNA splicing 28

1.3.3 FUS modulates mRNA stability, transport and translation 29

1.3.4 FUS involvement in noncoding RNA processing 29

1.3.5 FUS and DNA damage repair 29

2. Aim of the thesis 31

3. Results 33

3.1 Murine Oct4 Pseudogene 4 is essential for mESC differentiation 33

3.2 hOCT4P3 is the functional homolog of mOct4P4 36

3.3 mOct4P4 deletion studies to identify functionally relevant lncRNA regions 40

3.4 Functional characterization of crucial mOct4P4 lncRNA region 43

3.5 The role of SUV39H1 role mOct4P4 for minimal functional region silencing activity 46

3.6 The long non coding RNA mOct4P4 meets FUS (Fused in Sarcoma) 47

3.7 FUS role in mOct4P4 minimal functional region 50

3.8 FUS role conservation in hOCT4P3 function 52

3.9 The role of the FUS/SUV39H1/mOct4P4 interaction 53

3.10 Final model 55

4. Conclusion and discussion 57

5. Material and methods 63

5.1 Cell culture and generation of Embryoid Bodies (EBs) 63

5.2 Generation of plasmid 63

5.3 Viral transduction and generation of stable cell lines 65

5.4 Transient transfection 66

5.5 Mass spectrometry analysis 66

5.6 RNA immunoprecipitation 68

5.7 Western blotting and antibodies 68

5.8 RNA extraction and analysis 69

5.9 Chromatin immunoprecipitation 71

5.10 Co-Immunoprecipitation 72

5.11 Statistical analysis 72

Bibliography 73

Abstract

Pseudogenes are defined as non-functional genomic sequences originally derived from functional genes to which they are similar but with important defects, such as mutations, insertions and deletions, that make them unable to produce a functional protein. For a long time, pseudogenes have been considered as non-functional RNA relics, however, recently, some pseudogenes-derived lncRNA have been shown to be functional and play an important role in various physiological and pathological conditions. We recently discovered a murine Oct4 pseudogene (mOct4P4) which encodes a nuclear restricted lncRNA able to silence the transcription of the ancestral Oct4 gene. mOct4P4 is upregulated during mESC differentiation and recruits SUV39H1 to the Oct4 promoter to impose repressive heterochromatin *in trans*, leading to Oct4 gene silencing to promote cell differentiation. In this study, we obtained novel insights into the molecular mechanism of mOct4P4 pseudogene-dependent gene silencing and its conservation in human cells. CRISPR/Cas9 mediated mOct4P4 loss-of-function experiments showed an essential role for mOct4P4 in orchestrating an effective mESC differentiation. In addition, mOct4P4 deletion analysis identified the minimal functional region in mOct4P4 lncRNA (base pairs 984-1183) required for SUV39H1 binding and Oct4 ancestral gene targeting. Moreover, mass spectrometry analysis on mOct4P4 lncRNA pull-down identified a novel mOct4P4 interactor, the RNA binding protein FUS, which is required for mOct4P4 dependent silencing of Oct4 expression, suggesting that a mOct4P4 lncRNA/FUS/SUV39H1 complex has an important role in the silencing of Oct4 during mESC differentiation. Finally, we characterized the human homolog of mOct4P4 (hOCT4P3) and demonstrated the functional conservation between the two pseudogene-derived lncRNAs. In fact, we show that hOCT4P3 is able to form a complex with SUV39H1 and FUS that is able to impose repressive heterochromatin marks on ancestral gene promoter leading to silencing of OCT4 expression in ovarian cancer cells.

In conclusion, we provided novel insights into the epigenetic mechanism that the murine Oct4P4 lncRNA uses to control ancestral Oct4 expression *in trans* to regulate self-renewal and differentiation and further demonstrated that this mechanism is also conserved in humans.

1. Introduction

Gene expression profiles are far more complex than previously estimated. In fact, it has been assessed that only <2% of the genomic DNA is translated into protein, however transcriptome analysis revealed that almost two-third of the human genome is transcribed (Djebali et al., 2012; Maeda et al., 2006). Interestingly, the portion of untranslated transcriptome increases with the complexity of the organism. In other words, the degree of organismal complexity among species correlates with the proportion of each genome being transcribed into non-coding RNA (ncRNAs) (Fatica and Bozzoni, 2014; Taft et al., 2007). Moreover, the importance of non-coding genomic regions is reflected by the fact that approximately 90% of single-nucleotide polymorphisms in the human genome linked to various diseases are found outside of protein-coding regions (Maher, 2012).

In general, these ncRNAs are mainly responsible for the regulation of the expression of many genes at mRNA level (mRNA transcription, splicing and stability) but they have also showed to be involved in many different processes such as protein biosynthesis, scaffolding, ribozymes, telomere maintenance and many others. NcRNA genes number exceeds the number of protein coding genes and includes different ncRNA classes with different structure and functions such as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), spliceosomal RNAs (snRNA), small interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). In particular, lncRNAs are RNA transcripts that are longer than 200 nucleotides, often polyadenylated but devoid of evident open reading frames (ORFs), thus lacking coding capacity (Hu et al., 2018) which are transcribed by RNA polymerase II and are subjected to RNA splicing. The nucleic acid nature of lncRNAs confers them the ability to bind proteins and to mediate base-pairing to guide lncRNA complexes to RNA and DNA sites of action. This ability is conserved through all ncRNAs species, however, unlike small ncRNAs, lncRNAs can fold in complex secondary and higher order structures giving them the ability to act as flexible and modular scaffolds in order to tether protein factors that would not interact or functionally cooperate if they only relied on protein-protein interaction (Guttman and Rinn, 2012; Ng et al., 2012; Rinn and Chang, 2012). This complex nature suggests an essential role for lncRNAs as transcriptional or post-transcriptional modulators of protein-coding genes

as well as non-coding genes. For instance, one of the first examples of lncRNA acting as epigenetic transcription regulators include X-inactive specific transcript (XIST) (Lee and Bartolomei, 2013; Schoeftner et al., 2006; Wutz et al., 2002) and HOX transcript antisense RNA (HOTAIR) (Rinn and Chang, 2012). By contrast, examples of lncRNA acting as post-transcriptional modulators include competitive endogenous RNAs (ceRNAs), which compete with mRNA for miRNAs (Salmena et al., 2011), and lncRNAs that can modulate RNA translation and stability, such as lincRNA-p21 and BACE1-AS (Faghihi et al., 2008; Huarte, 2013; Yoon et al., 2012). As a matter of fact, dysregulation of lncRNAs correlates with a number of human diseases, such as cancer (Prensner and Chinnaiyan, 2011; Wahlestedt, 2013). A vast number of lncRNAs have been reported so far and have been classified in diverse groups, such as transcribed ultraconserved regions (T-UCR), natural antisense transcripts (NATs), enhancer RNAs, long intergenic RNAs (lincRNAs) and many others. Among these, lncRNAs derived from the so-called pseudogenes are gaining importance.

1.1. Pseudogenes

1.1.1 Types of pseudogenes

Pseudogenes are known as non-functional sequences of genomic DNA that are structurally similar to functional genes but harboring many defects that makes them incapable to produce functional protein (D'Errico et al., 2004). The most common defects include the loss of transcription starting site, presence of premature stop codons, splicing errors, frameshift mutations (insertion/deletion) and lack or abnormality of flanking regulatory regions. Pseudogenes can be classified in three main groups based on their origin: non-processed, processed and unitary pseudogenes (D'Errico et al., 2004; Hu et al., 2018; Pink et al., 2011; Poliseno, 2012).

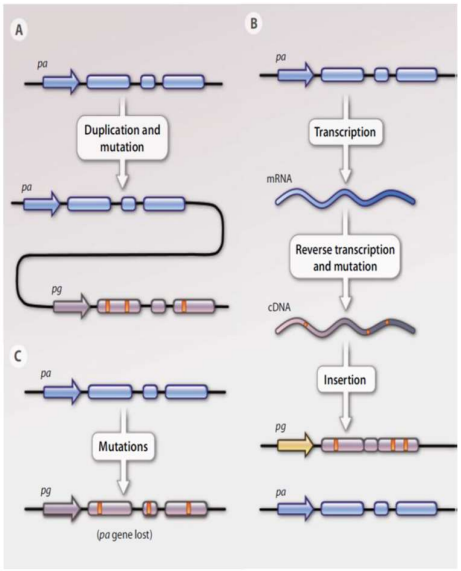

Figure 1: Types of pseudogenes. (A) Non-processed pseudogenes derive from gene duplication and are located on the same chromosome as the parental gene from which they are derived. (B) Processed pseudogenes arise by retrotransposition and are located on a different chromosome than the parental gene. (C) Unitary pseudogene derive from mutations of the parental gene, which is in turn lost. Blue boxes and lines, parental (pa) gene and mRNA; gray boxes and lines, pseudogene (pg) gene and RNA; orange dots and boxes, mutations; yellow arrow, unrelated promoter. Figure taken from Poliseno et al. 2012

Non-processed pseudogenes (Figure 1A) derive from a process of tandem duplication or uneven crossing-over (Mighell et al., 2000). Given their origin they are located in the same chromosome as the parental gene and they may retain their intron-exon characteristics. This extra copy of the gene is unnecessary for the cell and thus it can accumulate mutations without damaging the organism. Normally, the copies of duplicated genes are expected to become completely unfunctional and they continue to drift until they are either deleted or become unrecognizable as a genetic copy (D'Errico et al., 2004; Mounsey et al., 2002). However, it has been demonstrated that some non-processed pseudogenes may still be transcribed and could seldom become functional acquiring novel functions or mode of expression and thus become fixed in the genome. An interesting example of this is the "resurrection" of duplicated pseudogenes; bovine seminal ribonuclease, which seems to be a duplicated pseudogene of pancreatic RNAse, was reportedly dormant for 20 million years. However, phylogenetic conservation analysis revealed that, at some point of bovine evolution, the pseudogene was repaired and expressed with a mechanism that involved gene conversion, which involves transfer of information from the pancreatic RNAse ancestral gene to the seminal RNase pseudogene (Trabesinger-Ruef et al., 1996).

Processed pseudogenes (Figure 1B), by contrast, originate from a retrotransposition event. This process comprises three stages: (i) initially, RNA synthesis from genomic DNA template occurs; (ii) this primary transcript is processed, losing introns, and producing a mature messenger RNA (mRNA); (iii) this mRNA acts as a template for a reverse transcription process mediated by the enzyme reverse transcriptase, producing a DNA sequence which, in turn, is inserted in a random position (D'Errico et al., 2004; Tchnio and Segal-bendirdjian, 1993; Vanin, 1984). Given their origin, processed pseudogenes lack introns and possess a poly-A tail and are flanked by direct repeats at either end of the pseudogene. They can be complete or incomplete copies of the CDS produced by RNA polymerase II. Their retrotransposition seems to be mediated by LINE-1, since it can mobilize transcribed DNA not associated with LINE sequence by a process involving the diversion of the LINE enzymatic machinery by the corresponding mRNA transcripts (Ding et al., 2006; Esnault et al., 2000; Pink et al., 2011; Tchnio and Segal-bendirdjian, 1993). Thus, processed pseudogenes may also contain long interspersed nuclear elements (LINE) sequences - the most abundant transposable elements in the mammalian genome - since

they are not under selective pressure. This sequences are not normally found in coding regions, so their presence could allow the identification of a pseudogene (Cronin et al., 1998; D'Errico et al., 2004; Dubbink et al., 1998; Foord et al., 1996). Retrotranscription is far from being a high-fidelity process thus processed pseudogenes accumulate numerous mutations, which renders them very different when compared from the template RNA. In addition, they do not contain regulatory regions, which lie in non-transcribed regions, and even if their coding region remains intact, they are likely to generate inactive copies. Indeed, processed pseudogenes could be inserted in an inactive genomic region that could not activate their transcription, this pseudogenes are thus defined 'dead-on-arrival' (D'Errico et al., 2004). However, emerging evidence shows that some pseudogenes can indeed be transcribed (at least 5% in the human genome). In fact, even if they do not possess regulatory regions they could rely on different transcriptional elements of the region in which they land (Kandouz et al., 2004; Poliseno, 2012). In general, processed pseudogenes that land in intronic regions can exploit the transcriptional machinery of their host gene and, as a result, they are more frequently expressed when compared with the ones in intergenic regions (Vinckenbosch et al., 2006). Reverse transcription is a ubiquitous process, thus it is expected that mammalian genome may contain a number of these gene copies that, as seen before, are in the majority non-functional and interspersed along the genome. This process was defined as a 'Vesuvian mode' of evolution because it resembles a volcano producing lava (D'Errico et al., 2004; Leder, 1982). After their retro-insertion processed pseudogenes undergo two different evolutionary processes. First, they accumulate point mutations in a very fast rate and so they could rapidly start to diverge from their parental gene, which displays slower point mutation rate. As a result, by a process called 'compositional assimilation', processed pseudogene nucleotide composition will tend to increasingly resemble the surrounding non-functional region. The second evolutionary process is caused by the predominance of deletions over insertions, as a result processed pseudogenes reduced in size are produced. It has been estimated that they lose half of their nucleotide content in nearly 400 million years. This explains why the human genome is still full of pseudogenes. The shrinkage process is too slow to counterbalance the increase in genomic content caused by the 'Vesuvian evolution'. So, the restriction in pseudogene number and their fixation in the genome is due to other factors such as natural selection

(D'Errico et al., 2004). In fact, since pseudogenes have accumulated some mutations, they could possess different functions than the original gene that could be selected by selective pressure (Moreau-Aubry et al., 2000).

Unitary pseudogenes (Figure 1C), unlike non-processed or processed pseudogenes, do not have a parental gene that encodes a functional protein since they arise from spontaneous mutations in former protein-coding genes (Zhang et al., 2010c). They are generated by disruptive mutations occurring in functional genes and prevent them from being successfully transcribed or translated. They differ from duplicated pseudogenes in that the disabled gene had an established function rather than being a more recent copy of a functional gene (Zhang et al., 2010c). Thus, a gene present in mouse could not be present in human where it lies as an inactive ortholog (unitary pseudogene).

1.1.2 Pseudogene number and expression

The number of pseudogenes in a genome depends on the relative rate of gene duplication and pseudogene loss (Mounsey et al., 2002). As a result, the number of pseudogenes is very different in metazoan genomes (Zhang and Gerstein, 2004). However, the relationship between the amount of pseudogenes, the size of the proteome and the amount of coding versus non-coding DNA in genomes is still unclear. For example, it has been proposed that *C. elegans* displays a very high rate of gene duplication and, as a consequence, there is a large proportion of pseudogenes (at least 20% of annotated *C. elegans* genes may be, in fact, pseudogenes) (Mounsey et al., 2002). By contrast, it seems that the majority of pseudogenes found in mammalian genomes derives from mRNA. However it could be an underestimation of the number of non-processed pseudogenes due to the difficulty to distinguish non-processed pseudogenes from functional genes (D'Errico et al., 2004). In total, it has been estimated that the human genome may contain 18,000 pseudogenes of which almost two-thirds are processed (Poliseno, 2012). Despite the high number of pseudogenes only 10% of human genes have a pseudogenic counterpart and most pseudogenes derive from abundantly expressed gene families such as metabolic enzymes, structural proteins (such as tubulin and actin) and ribosomal protein genes, since they are highly transcribed and a retrotranscription event is more likely to have place for abundantly expressed mRNAs (McDonell and Drouin, 2012). In fact, 20% of the total pseudogenes comes from ribosomal proteins. As stated above, most pseudogenes lose the ability to be

transcribed due to mutations in their promoter or integration into silent regions of the genome, nevertheless computational approaches suggest that at least 5% of the human pseudogenes are indeed transcriptionally active (Harrison et al., 2005; Zheng et al., 2007). In addition, there is experimental evidence of cell type-specific transcription of pseudogenes (Balakirev and Ayala, 2003; Pink et al., 2011; Scarola et al., 2015; Zheng and Gerstein, 2007). The abundance of pseudogene transcripts is lower than that of their functional parental protein-coding genes, but exceptions exists with pseudogene transcripts levels that are comparable or even higher of their coding counterparts (Poliseno et al., 2010; Sorge et al., 1990). Even more, in some contexts only the pseudogene(s) are expressed, whereas the parental gene is not (Scarola et al., 2015; Suo et al., 2005). Interestingly, some transcribed pseudogenes RNA show tissue specificity expression and there are some examples of spatiotemporal transcription of pseudogenes different from that of their parental gene (Elliman et al., 2006; Olsen and Schechter, 1999; Pink et al., 2011; Scarola et al., 2015). Moreover, pseudogene pattern of expression can be altered under different patho-physiological conditions, such as diabetes (Chiefari et al., 2010) and cancer (Hu et al., 2018; Poliseno, 2012; Suo et al., 2005; Zou et al., 2009). Given their origin, transcriptional activity of a pseudogene depends on the promoter controlling its expression. Some pseudogenes utilize their own promoter, while others take advantage of the promoters of nearby or host genes (Harrison et al., 2005; Vinckenbosch et al., 2006). Particularly, processed pseudogenes transcription depends on the site in which they integrate. Therefore, the difference that sometimes is found between the expression of pseudogenes and their parent genes may depend merely on the fact that they are driven by a different promoter. But if this were true one would expect pseudogene RNA transcription as evolutionarily neutral (or even negatively selected in the aim of conserve cellular energy). However, analysis on processed pseudogenes shows that about 50% of them are conserved through primate evolution (Khachane and Harrison, 2009; Pink et al., 2011). Surprisingly, translated protein originating from pseudogene RNA have been identified (Balakirev and Ayala, 2003; Brosch et al., 2011; Ezkurdia et al., 2012), but, since they often harbor premature stop codons or inactivating mutations, they are shorter than proteins encoded by true genes or they are sub-functional respectively. Together, these evidences –

transcription, tissue-specificity and conservation – hint a functional role for pseudogenes in cells.

1.1.3 Pseudogene function and biological role

What is the potential benefit to mammals in retaining genes that have lost protein coding-potential? Actually, many ncRNAs have been shown to regulate a plethora of processes in cells with various mechanisms and pseudogenes derived lncRNAs could make no exception. There is evidence for an establishment of a regulatory network between pseudogenes and their parental genes but this relationship is not exclusive since there is evidence of pseudogene regulation of unrelated genes. It has been proposed a series of mechanism through which pseudogenes can act (Figure 2) and can be classified in two main categories: (i) functions related to those of the parental genes and (ii) pseudogenes functions independent to those of the ancestral gene.

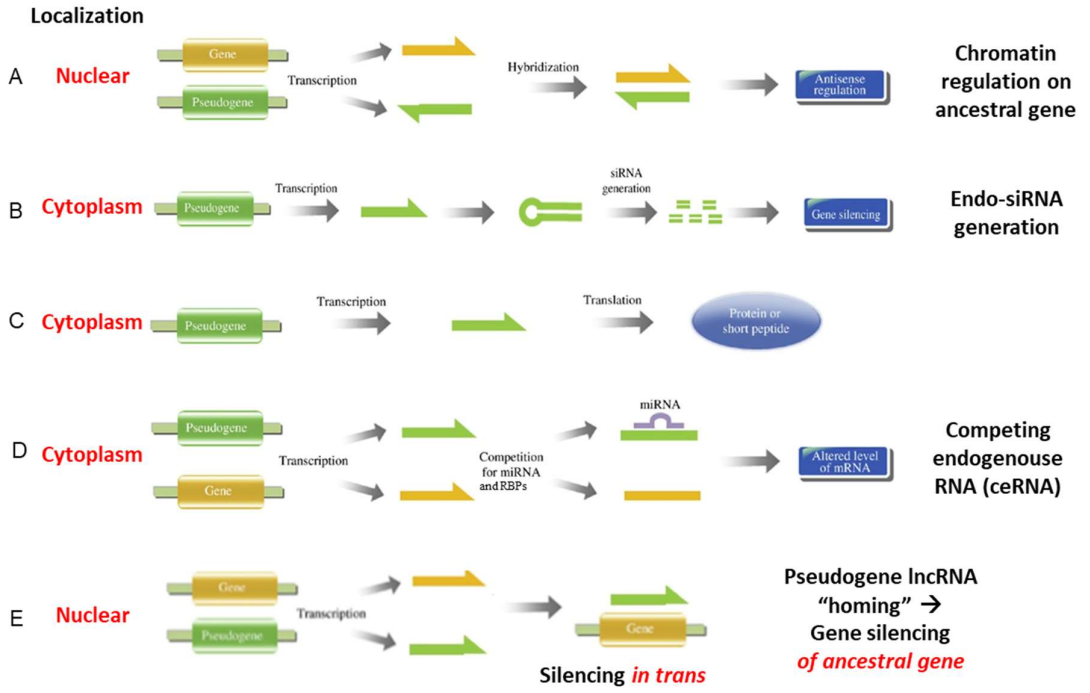

Figure 2: Mechanisms of pseudogene function. (A) Nuclear antisense pseudogene transcripts could recruit epigenetic rregulators to the ancestral gene (B) Pseudogene transcripts with inverted repeats could give rise to endo-siRNAs which regulate ancestral gene (C) Some pseudogene transcripts could be translated into proteins or short peptides with various functions (D) Pseudogene transcripts could compete fort miRNAs and RNA binding proteins affecting the abundance and stability of their ancestral gene transcripts (E) Nuclear sense transcripts could home epigenetic remodelers to thei ancestral gene promoter

1.1.3.1 Functions related to the parental gene of pseudogenes and mediated by the pseudogene DNA

Sequence similarity between pseudogenes and parental genes makes them predisposed to DNA exchange. This mechanism is called gene conversion and allows the transfer of DNA from pseudogenes to their parental genes (Figure 3A). Pseudogenes, as stated above, contain several mutations so a gene conversion event to their parental gene makes the latter non-functional. This process is at the origin of several human genetic diseases, such as autosomal dominant retinitis pigmentosa, and cancer predisposition (Bischof et al., 2006; Ganster et al., 2010). Notwithstanding, there are reported cases in which the transfer of DNA from pseudogene to parental gene does not impinge on the protein coding potential of the parent gene, in fact it could be able to produce a mutation that can modify parental gene expression in a positive or negative way (Poliseno, 2012). Another way through which pseudogenes and parental genes exchange DNA is via a homologous recombination mechanism (Figure 3B). Also in this case the recombination event could render the parental gene nonfunctional, for example because parent gene could be deprived of its promoter or initiation codon as shown for BRCA1 and its pseudogene ΨBRCA1 (Puget et al., 2002). Finally, interactions between parental gene and its pseudogene(s) could be mediated by several DNA binding proteins resulting in either a negative or positive effect on transcription of the parental gene (Figure 3C) (Chang and Cheng, 1998; Lee et al., 2002; Troyanovsky and Leube, 1994). Altogether these mechanisms provide a proof of principle that pseudogenes can act also in absence of their transcription modifying the expression of their transcribed parental gene(s) via an RNA independent mechanism.

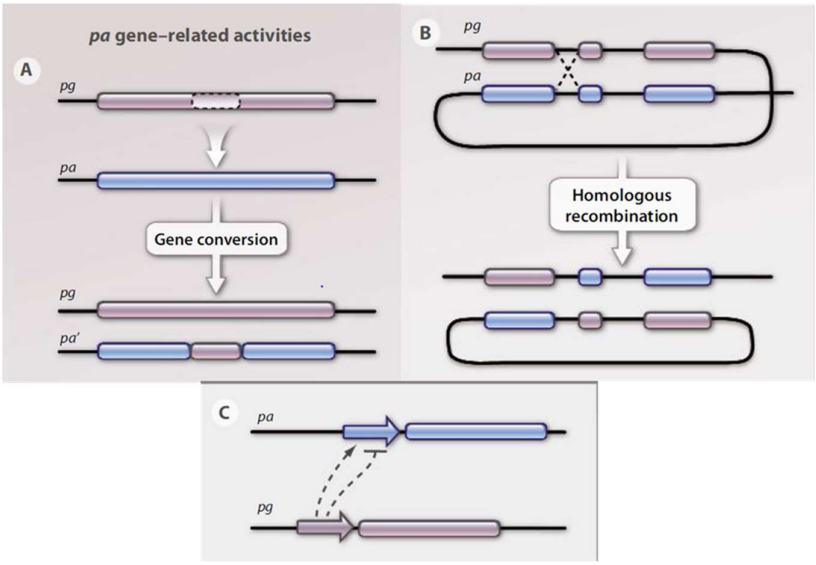

Figure 3: Mechanisms of action of pseudogenes at the DNA level. (A) Gene conversion, a monodirectional transfer of DNA from a pseudogene to its parental gene, changes the sequence of the parental gene. (B) Homologous recombination between the pseudogene and the parental gene causes a deletion of the parental gene. (C) Regulatory sequences located upstream of the pseudogene can enhance or inhibit the transcription of the parental gene. Blue boxes, parental (pa) gene; gray boxes, pseudogene (pg). Adapted from Poliseno et al. 2012

1.1.3.2 Functions related to the parental gene of pseudogenes and mediated by pseudogene RNA transcribed in sense

Pseudogene RNAs transcribed in sense, as mRNAs, can sponge microRNA (miRNAs), RNA binding proteins (RBPs) or the translational machinery (Figure 4). miRNAs are a class of small ncRNAs that act as posttranscriptional regulators of gene expression through their binding to partially complementary sites (miRNA recognition elements, MREs) mainly located in the 3'UTR of mRNAs. Their binding with the target mRNAs causes a decrease in mRNA stability and translation (Huntzinger and Izaurralde, 2011; Winter et al., 2009). One way to regulate microRNAs, aside from their expression, is the concomitant presence of multiple legitimate targets (either coding and noncoding) at the same time. These targets could act as competitive endogenous elements (ceRNAs) and in this way titrate miRNAs availability (Seitz, 2009). In this manner ceRNAs could protect each other from miRNA-mediated post-transcriptional regulation competing for the binding of the miRNA molecules. Classic vision states that miRNAs affect the abundance of their targets but with this new view also miRNA targets could affect the function of miRNAs decreasing their availability (Salmena et al., 2011). The ceRNA hypothesis provide a mechanism through

which RNA molecules can crosstalk *in trans* and, in turn, modulate each other abundance in an indirect way mediated by shared miRNAs (Marques et al., 2011). This new landscape can open a new role for many components of the transcriptome (both coding and noncoding) that contain at least a MRE. Indeed, many coding (Jeyapalan et al., 2011; Karreth et al., 2011; Lee et al., 2009, 2010, 2011; Poliseno et al., 2010; Rutnam and Yang, 2012; Sumazin et al., 2011; Tay et al., 2011) and noncoding (Cazalla et al., 2010; Cesana et al., 2011; Franco-Zorrilla et al., 2007; Iskow et al., 2012; Libri et al., 2012; Poliseno et al., 2010; Wang et al., 2010a) RNAs have demonstrated to act via a ceRNA-mediated mechanism. Among these, pseudogenes are particularly adapt to function as ceRNA for two main reasons: first, they are basically not translated, hence they can compete with miRNAs without interfering with the translational machinery (Gu et al., 2009). Second, they show high sequence homology when compared with their parental genes and they can act as multifunctional decoys that can protect their counterparts from all shared miRNAs (Poliseno, 2012). This is different from standard lncRNAs that usually share only one or few MREs (Cazalla et al., 2010; Cesana et al., 2011; Franco-Zorrilla et al., 2007; Libri et al., 2012; Wang et al., 2010a), and could explain how pseudogenes can act also when they are present in low ratio when compared with their protein coding counterparts and also it could clarify why they often show a correlation between pseudogene and parental gene expression (Poliseno et al., 2010). Examples of pseudogene lncRNAs which act as ceRNAs include PTENP1 and KRAS1P both involved in cancer (Ioffe et al., 2012; Poliseno et al., 2010). Apart from miRNAs, transcribed pseudogene RNAs could sequester also RBPs. RBPs are proteins that bind to mRNA transcripts recognizing specific sequences or secondary structures (Lunde et al., 2007) and, in this way, they can act as mRNA stabilizers or destabilizers, modulating the abundance of their coded protein. Consequently, competition for an RBP between a pseudogene and its parental gene could lead to either destabilization (if the RBP is stabilizing) or stabilization (if the RBP is destabilizing) of the parental gene. For example, Myosin light chain kinase pseudogene 1 (MYLKP1) RNA decreases MYLK parental gene RNA stability competing for still unknown stabilizing factor and thus promoting cell proliferation (Han et al., 2011). Finally pseudogenic RNAs compete for the translational machinery and thus alter the levels of the translation of parental gene. For example, Connexin 43 pseudogene (ΨCX43) which binds translational machinery more tightly than

its parental gene RNA inhibiting its translation and thus promoting tumor cell spreading (Bier et al., 2009; Carystinos et al., 2001; Kandouz et al., 2004; Poliseno, 2012).

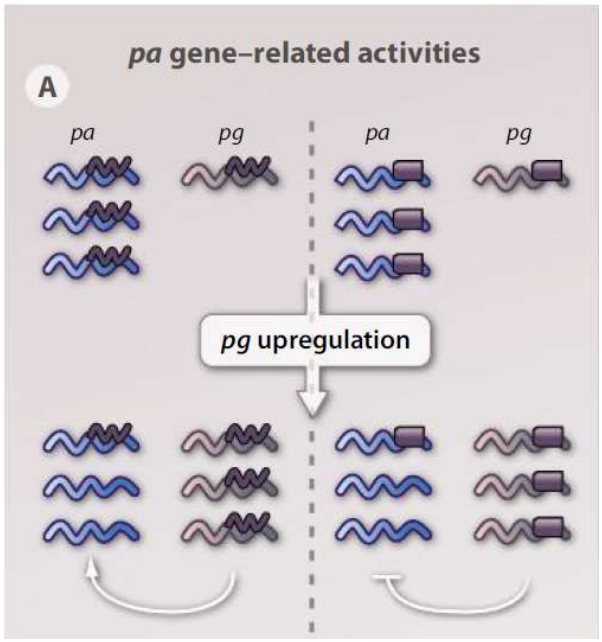

Figure 4: Mechanisms of action of pseudogenes at the RNA level. (Left) Pseudogene RNAs compete with their parental mRNAs for shared microRNAs, thus increasing the expression of the parental gene. (Right) Pseudogene RNAs compete with their parental mRNAs for shared stabilizing RBPs, thus inhibiting the expression of the parental gene. Blue boxes and lines, parental (pa) gene and mRNA; gray boxes and lines, pseudogene (pg) gene and RNA; gray rectangles, RBP; black lines, microRNAs. Adapted from Poliseno et al. 2012

1.1.3.3 Function of pseudogenes related to the parental gene and mediated by pseudogene RNA transcribed in antisense

Antisense transcription occurs in 70% of human coding and noncoding transcriptional units giving rise to NATs which modulate the expression of the corresponding sense transcripts at the epigenetical, transcriptional or posttranscriptional level (Faghihi and Wahlestedt, 2009; Morris, 2009; Werner and Swan, 2010) (Figure 5). For example, NATs form long double stranded (ds)RNA molecules with their sense RNAs that are subsequently cleaved to give rise to endogenous short interfering RNAs (endo-siRNAs) (Carlile et al., 2009; Song et al., 2011; Watanabe et al., 2008). Endo-siRNAs can, subsequently, cause the RNA interference-mediated posttranscriptional cleavage of sense transcript (Okamura and Lai, 2008; Song et al., 2011; Tam et al., 2008; Werner and Swan, 2010). Pseudogenes transcribed in antisense direction generally act in this way but they can regulate the abundance of both

their own sense transcripts *in cis* and of their parental counterparts *in trans*. Pseudogene-derived endo-siRNAs can be classified in two groups: (i) processed from dsRNAs formed after the hybridization of spliced transcripts from protein coding transcripts to antisense transcripts of cognate transcripts and (ii) processed pseudogenes from inverted-repeat pseudogenes (Tam et al., 2008; Watanabe et al., 2008).

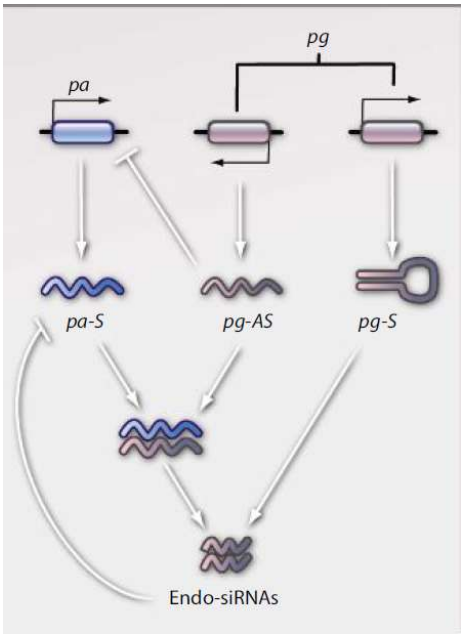

Figure 5: Antisense pseudogene transcripts recruit chromatin remodeling complexes to the parental gene and repress its transcription or are processed to form endo-siRNAs, which inhibit the expression of the parental gene at the posttranscriptional level. Endo-siRNAs can also originate from sense transcripts of pseudogenes with inverted-repeat structures. Blue boxes and lines, parental (pa) gene and mRNA; gray boxes and lines, pseudogene (pg) gene and RNA;. Adapted from Poliseno et al. 2012

1.1.3.4 Pseudogene functions related to the parental gene mediated by pseudogene-encoded proteins

As stated above, some processed pseudogenes are reported to encode protein which, although usually mutated or truncated, are partially functional (Balakirev and Ayala, 2003; Brosch et al., 2011; Ezkurdia et al., 2012). These pseudogenic proteins can be considered as intronless paralogs which exert the same function as their parental counterparts but, given their processed origin, in different tissues, cellular compartments or pathophysiological conditions (Figure 6A), CRIPTO1 is a receptor of TGF-β signaling pathway which is not expressed in cancer cells. Its pseudogene CRIPTO3 translates in a protein with a four amino acids mutation but it is expressed in cancer cells where it exerts oncogenic activities activating the extracellular signal-regulated kinase (ERK)-AKT pathway (Sun et al., 2008). In addition, pseudogene-derived proteins can carry mutations that cause the constitutive activation of the encoded product. An example of a constitutively active pseudogene product comes from H-RAS pseudogene Ha-RAS2 which encodes a constitutively active form of H-RAS due to the mutation of amino acids Gly^{12} and Glu^{63} (Takahashi et al., 2003). Finally, pseudogenic truncated proteins can still affect the function of the parental protein (Figure 6B). For example ~50 pseudokinases (i.e. ΨBRAF1) have been identified most of which act as allosteric activators or as scaffolds that bring together various components of the signaling pathway (Manning et al., 2002; Poliseno, 2012; Zeqiraj and van Aalten, 2010; Zou et al., 2009).

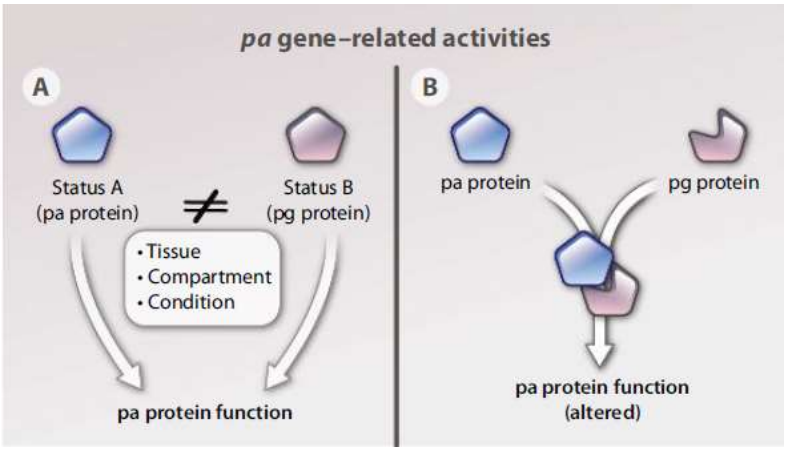

Figure 6: Mechanisms of action of pseudogenes at the protein level. (A) Pseudogenic proteins have the same activity as the parental proteins, but function in different tissues, cellular compartments, or pathophysiological conditions. (B) Partially functional pseudogenic proteins or pseudogenic proteins with altered function affect the activity of their parental proteins. blue polygon, parental (pa) protein; gray polygon, pseudogenic (pg) protein. Adapted from Poliseno et al. 2012

1.1.3.5. Parental gene-independent functions of pseudogenes mediated by the pseudogene DNA

Given their intrinsic properties (such as high mutations content), pseudogenes can show functions not related to the one of the parental gene, this phenomenon is called functional independence (Zheng et al., 2007). One mechanism through which pseudogenes acquire functions unrelated to the functions of their parental counterpart is the so-called exonization process. This consists in a change in their structure due to the acquisition of new exons (Poliseno, 2012). Processed pseudogenes that are integrated into introns exploit the transcriptional machinery of their host gene and, sometimes, they can be co-transcribed as fusion transcripts. If the insertion occurs upstream of the exons of the host gene (i.e. in an intron between the 5'UTR and the first exon) a noncoding fusion happens and, if a protein is translated from this fusion transcript, the resulting protein would be derived only from the pseudogene (Ghosh et al., 2001; Poliseno, 2012) (Figure 7D). If the intron in which a pseudogene lands is farther downstream, a coding fusion occurs. In this case the encoded protein is a chimera derived from both the pseudogene and the host gene (Rogalla et al., 2000; Vinckenbosch et al., 2006) (Figure 7D). The chimeric nature of these fusions have implications for both gene and pseudogene since they both acquire functions different from their original ones (Poliseno, 2012). An alternative exonization process happens when pseudogenes acquire exons that arise *de novo* (Figure 7E).

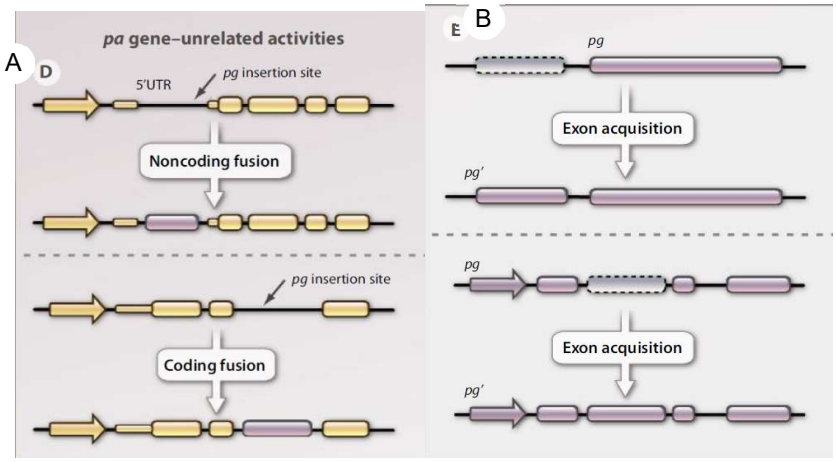

Figure 7: (A) Insertion of a processed pseudogene into either the noncoding (upper panel) or the coding (lower panel) region of an unrelated gene modifies the sequence and hence the function of both the pseudogene and the gene to which it becomes fused. (B) Acquisition of new exons by processed (upper panel) or nonprocessed (lower panel) pseudogenes modifies their sequence and hence function. Gray boxes, pseudogene (pg); yellow boxes, unrelated gene (wide boxes indicate coding sequences and narrow boxes indicate UTRs). Adapted from. Poliseno et al. 2012

1.1.3.6 Parental gene-independent functions of pseudogene mediated by the pseudogene RNA

Mutations acquired by the pseudogene RNAs allow this transcripts to perform functions unrelated to those of their parental genes and affect genes unrelated to them (Poliseno, 2012) (Figure 8C). For example, XIST has been proposed to represent an unitary pseudogene derived from the protein coding Lnx3 gene after an accumulation of mutations. In fact, among the 10 exons in XIST, 6 derive from transposable elements, but the remaining 4 show sequence similarity with Lnx3 (Duret et al., 2006; Elisaphenko et al., 2008). Pseudogenes can also act as miRNA precursors. Examples of genes that contain miRNA are well represented in the mammalian genome. miRNA can arise in intron or exons when the transcriptional units do not encode proteins or in introns when they are protein coding (Kim and Nam, 2006). Pseudogenes are an additional class of transcriptional unit source of miRNAs in which sequence drift may allow them to function as a reservoir of new miRNAs (Sasidharan and Gerstein, 2008) (Figure 8D). For example, Keratin 19 processed pseudogene (KRT19Ψ) contains miR-492 while in the parental gene the miR-492 gene is split between two exons and thus interrupted by an intron (von Frowein et al., 2011). Pseudogene RNAs can exert functions related to their ability to act as miRNA decoys or to bind RBPs. In fact, since miRNAs and RBPs have typically multiple targets, their binding to a pseudogene can sequester them not only from the parental gene but also from all their unrelated targets. In addition, because parental gene and pseudogenes do not always have identical sequences, it is conceivable that they do not bind the same set of miRNAs and RBPs (Figure 8E).

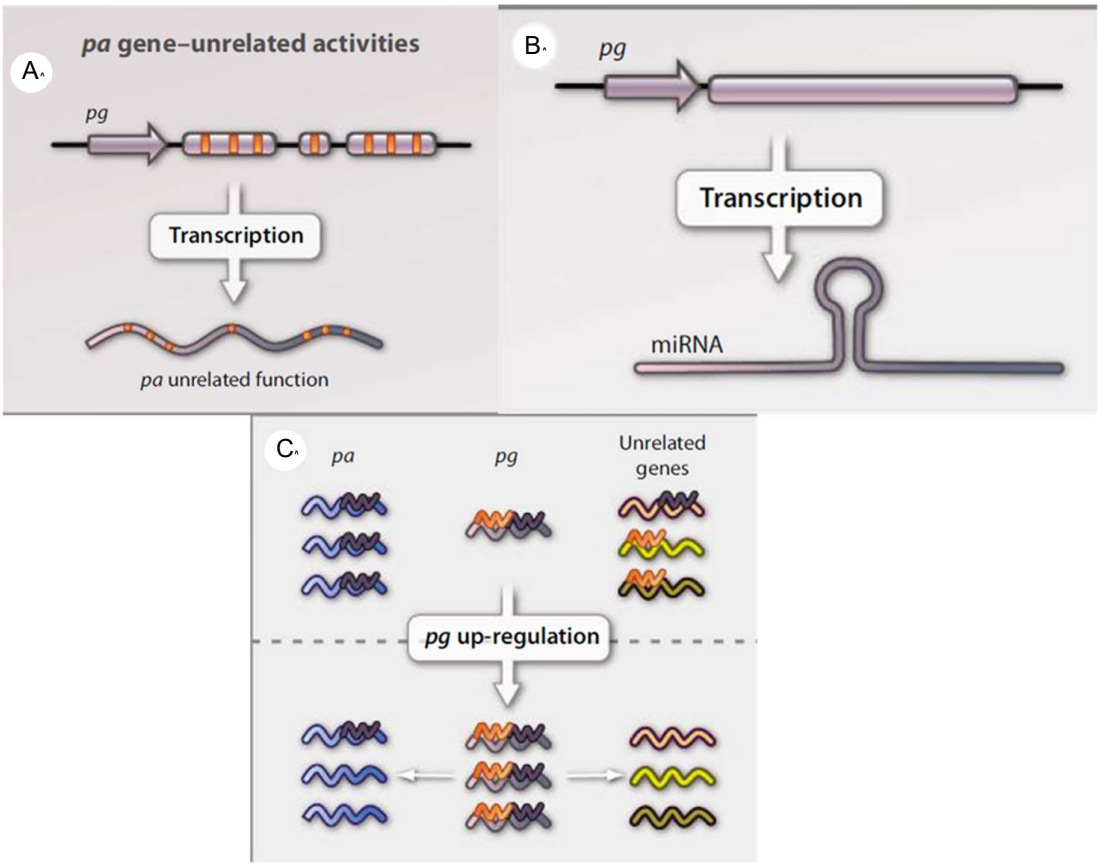

Figure 8: (A) Pseudogenes function as lncRNAs and affect genes that are unrelated to the parental gene. (B) pseudogenes encode microRNA precursors. (C) Pseudogene RNAs compete for microRNAs that target the parental gene and unrelated genes (gray microRNAs) and for microRNAs that do not target the parental gene (orange microRNAs). In both cases, the expression of the microRNA targets is increased as a result of the activity of the pseudogene as a ceRNA. gray boxes and lines, pseudogene (pg) gene and RNA; orange dots and boxes, mutations; gray and orange lines, microRNAs; gray rectangles, RBPs; brown, yellow, and pink lines, RNAs of unrelated genes. Adapted from Poliseno et al. 2012.

1.1.4 The boundary between functional pseudogenes and genes

As defined earlier, pseudogenes are known as gene relics that no longer encode functional proteins (D'Errico et al., 2004). Because it is often difficult to prove the lack of biological function for a segment of DNA, a pseudogene is operationally defined by its homology to a functional gene but does not require the presence of non-functionality at protein level (Xu and Zhang, 2016). This non-functionality is often due to a disruption of the canonical ORF rendering the pseudogene unable to encode a functional protein. RNA-mediated gene duplication, which gives rise to processed pseudogenes, only copies the transcribed and processed region of a gene, thus, the duplicated gene lacks the original promoter and introns. Consequently, processed pseudogenes are frequently not transcribed (D'Errico et al., 2004). However, in rare cases RNA-mediated duplicates may, in some way, retain the

canonical ORF (with promoter and/or introns) and provide the possibility of transcription and, more rarely, translation. In this case, they should be annotated as a retrogenes instead of as processed pseudogenes (Pei et al., 2012). Using this operational criteria, a number of pseudogenes have been annotated in sequenced genomes (Karro et al., 2007; Sisu et al., 2014) but, as stated earlier, the operational definition of pseudogene does not require proof of non-functionality at protein level. Thus, it is not surprising that many operationally-defined pseudogenes have been proposed to be transcribed and to have a function on the protein level (Balakirev and Ayala, 2003; Pink et al., 2011; Poliseno, 2012). Indeed, a human proteomic study identified peptides encoded by 107 pseudogenes (Kim MS et al., 2014). However, these peptides may represent a function for pseudogenes at protein level but they could also be resulting from spurious translation with no protein function. Thus, the presence of a translated protein it is not itself a proof of functionality that would render a misannotated pseudogene as a functional gene. In particular, to proof true a biological function of a pseudogene it is required to demonstrate that the activity of the pseudogene-encoded protein is under natural selection (Doolittle, 2013; Doolittle et al., 2014; Graur et al., 2013). In 2016, Xhu and Zhang demonstrated that at least 15% of the translated pseudogenes possess evolutionary selected functions at the protein level while the majority of them possess no selected function at the protein level (Xu and Zhang, 2016). Thus they conclude that, translation per se is not a guarantee of functionality. However, the presence of selective pressure on some translated pseudogenes opens the possibility that these pseudogenes could be misannotated. In fact, by definition they are not pseudogenes if they are under purifying selection. However, it should be stated that there was no evidence for their mRNA or protein expression at the time of annotation rendering the initial annotation not erroneous. In conclusion, if an annotated pseudogene is transcribed, translated and under purifying selection, it should be reannotated as gene.

1.2. Oct4 (Octamer-binding transcription factor 4)

Oct transcription factors are crucial developmental regulators of embryonic pluripotent state to terminal differentiation. Oct proteins are members of the POU family of transcription factors and are able to recognize the so-called “octamer motif”, an eight base pair consensus sequence [ATGC(A/T)AAT] and its variants, located in the regulatory

regions of both ubiquitous and tissue-specific target genes (Bodner et al., 1988; Ingraham et al., 1988; Kemler et al., 1989).

The 5′ ATGC motif associates with the 80 amino acid POU-specific domain (POU_S), while the 3′ with the 60 amino acid POU homeodomain (POU_H). Interestingly, the two domains are independently folded and joined by a flexible linker (Figure 9).

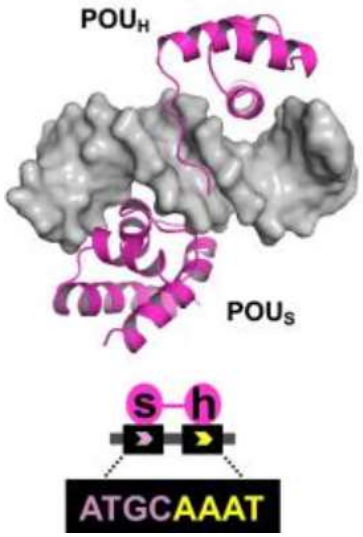

Figure 9: POU family of transcription factors are characterized by the presence of an homeodomain (POU_H, of about 60 amino acid) joined by a flexible linker to a second, independently folded DNA-binding domain termed the POU specific domain (POU_S) (Adapted from Tantin D., 2013).

The best-known Oct protein is the pluripotency regulator Oct4 (Okamoto et al., 1990; Rosner et al., 1990; Schöler et al., 1990), which is encoded by the Pou5f1 gene. Oct4 is expressed both in early mammalian embryo and also in the germline. In the former case, it is found in the pluripotent cells of the blastocyst inner cell mass (ICM) and epiblast. Oct4 knockdown in embryos results in a failure in pluripotency establishment: instead of forming an ICM, the whole embryo differentiates into trophoectoderm and fails to implant (Nichols et al., 1998). By contrast, a less than twofold increase in expression causes differentiation into primitive endoderm and mesoderm. Thus a critical amount of Oct4 is required to sustain stem-cell self-renewal, and up- or downregulation induce divergent developmental programs (Niwa et al., 2000). Altogether these data demonstrate that Oct4 is critically involved in the self-renewal of undifferentiated embryonic stem cells. As such, it is frequently used as a marker for undifferentiated cells and Oct4 expression must be closely regulated since Oct4 protein expression is crucial to maintain lineage-specific embryonic stem cell (ESC) differentiation and developmental fate. Importantly, Oct4 is

expressed in ESC cells derived from ICM and ectopic expression of the key ESC self-renewal transcription factors Oct4, Sox2, Klf4 and c-Myc induces pluripotency in differentiated cells, resulting in induced pluripotent stem (iPS) cells (Kim et al., 2009; Takahashi et al., 2006; Zhu et al., 2010). Out of this basic set of transcription factors, Oct4 is the only indispensable factor for iPS cell generation. This highlights a central role of Oct4 for ESC pluripotency and self-renewal (Sterneckert et al., 2012). Oct4 targets can be categorized in three broad classes: (i) positive targets that are actively expressed; (ii) positive targets that are transcriptional silent; and (iii) negative (repressed) targets (Tantin, 2013). (i) Positive targets that are actively expressed in ESCs include active pluripotency genes such as Nanog and its own gene (Pou5f1). Nanog, Pou5f1 and others make up "the core pluripotency network" which expression is maintained not only by Oct4 but also by other proteins encoded in this network such as Nanog and Sox2 (Boyer et al., 2005; Rodda et al., 2005). Positive targets that are actively expressed genes comprise also constitutively expressed proteins such as histones and metabolic enzymes (Chen et al., 2008a). (ii) In positive targets that are transcriptional silent, e.g. Hoxa5, Hoxc6, Pax6, Otx2, Gata2 and Pou4f1, Oct4 maintain an epigenetically "poised" state in ESCs but they become activated in developing tissues after the appropriate developmental cue or become stably repressed if cells proceed down other developmental trajectories. In ESCs, the "poised" state consists in a chromatin state free of DNA methylation and containing both activating and repressive histone marks (Bernstein et al., 2006; Meissner et al., 2008). Finally, negative Oct4 targets include, for example Cdx2, which is actively repressed by Oct4 and thus silent in ESCs (Yeap et al., 2009; Yuan et al., 2009).

Given the key role of Oct4 in mESC self-renewal, discovering pathways that control Oct4 expression has primary importance. Induction of ESC differentiation leads to rapid transcriptional and epigenetic silencing of Oct4 mediated by different players. For example, Cdx2 was shown to regulate Oct4 via a Bsg1-mediated epigenetic mechanism that seems to involve histone deacetylases HDAC1/2 (Wang et al., 2010b); Sato et al. demonstrated that GCNF recruits DNA methyltransferase (Dnmt3a/Dnmt3b) to the Oct-3/4 promoter and facilitates its methylation (Sato et al., 2006) and Gu et al. demonstrated that GCNF is required for Oct4 repression (Gu et al., 2005). Finally, Bright/Arid3A binds directly the Oct4

promoter to contribute to its repression (Popowski et al., 2014). Other pathways that regulate Oct4 expression involve posttranslational modifications, such as serine 111 phosphorylation and O-GlcNAc (Brumbaugh et al., 2012; Jang et al., 2012; Saxe et al., 2009; Spelat et al., 2012) or miRNAs, such as miR-145 and miR-335 (Schoeftner et al., 2013; Tay et al., 2008; Xu et al., 2009). Upon induction of differentiation of mESCs or embryonal carcinoma cells, the Oct4 promoter is subjected to several histone modifications (Kellner and Kikyo, 2010). Markers of transcriptionally active genes, such as acetylation of lysine 9 and 14 and methylation of lysine 4 on histone H3, begin to be removed after retinoic acid-mediated induction of differentiation (Feldman et al., 2006). Conversely, methylation of lysine 9 in histone H3, a marker of inactive genes, increases in an opposite manner. In this context, the best characterized histone modifying enzyme involved in Oct4 suppression is the histone methyltransferase (HMTase) G9a, which imposes di- and trimethylation of histone H3 (Feldman et al., 2006; Tachibana et al., 2002). Interestingly, G9a induces a primary dimethylation on lysine 9, but it seems that the trimethylation could be an indirect effect through recruitment of another HMTase (Kellner and Kikyo, 2010), for example Suv39h1 which was shown to mediate Oct4 promoter H3K9me3 imposition in a pseudogene-mediated manner (Scarola et al., 2015). These lysine methylations are then able to recruit the heterochromatin protein HP1α, resulting in heterochromatin formation and Oct4 gene silencing. In addition, it seems that G9a may be able to induce DNA methylation mediated by Dnmt3a and Dnmt3b in a histone modification-independent way (Epsztejn-Litman et al., 2008). Recent studies indicate that Oct4 pseudogenes are gaining importance regulating the expression of their ancestral gene through imposition of heterochromatin on Oct4 promoter as explained further in this work (Hawkins and Morris, 2010; Scarola et al., 2015).

1.2.1 Oct4 and cancer

The first evidence of the involvement of OCT4 in the context of cellular transformation came in 2003 when it was demonstrated that ectopic expression of OCT4 increased the malignant potential of ESCs when injected in nude mice. OCT4 is expressed in various tumor types such as bladder, lung, pancreas, gastric, esophageal and ovarian cancer (Huang et al., 2012; Jóźwicki et al., 2014; Kong et al., 2014; Li et al., 2014; Lin et al., 2014; Rijlaarsdam et al., 2011;

Zhang et al., 2010b; Zhao et al., 2012). In addition, tumors with high OCT4 expression are associated with metastasis and shorter patient survival rates, when compared with tumors which display low OCT4 expression (Chen et al., 2012; He et al., 2012; Karoubi et al., 2009; Li et al., 2015; Liu et al., 2014; Yin et al., 2015). For example, in oral squamous cell carcinoma OCT4 overexpression correlates with increased tumorigenicity and its silencing leads to tumor volume reduction and abolishment of the stemness characteristics. Furthermore, metastatic tissues and recurrent tumor samples showed higher levels of OCT4 when compared to primary tumor regions (Tsai et al., 2014). In addition, ectopic expression of OCT4 in primary breast cancer cell led to the selection of tumor-initiating cells that developed to high-grade, poorly differentiated breast carcinomas when injected in nude mice (Beltran et al., 2011). As stated above, OCT4 plays a critical role in self-renewal, supporting findings that link OCT4 expression with the maintenance and enhancement of the so-called cancer-stem cell (CSC) like characteristics (Chen et al., 2008b; Kim and Nam, 2011). Resistance to therapy occurs in a small proportion of cells which were identified as CSCs (Dean et al., 2005). In line with the role of OCT4 in promoting CSC-like features, OCT4 expression is correlated with increased resistance to various chemotherapeutic agents such as cisplatin and doxorubicin (Wang et al., 2010c). In addition, OCT4 expression was found high in various chemoresistant cancer cell lines (Linn et al., 2010) and promotes dysplastic growth in epithelial tissues, inhibiting cellular differentiation of adult progenitor-cells (Hochedlinger et al., 2005). Altogether, these evidences underline the importance of OCT4 in tumor initiation and progression towards chemoresistant disease.

1.2.2 OCT4 and ovarian cancer

Ovarian cancer is the 8th most common malignancy in women worldwide and the third most common female-specific cancer after breast and cervix cancer, with an incidence rate of 10-15 per 100,000. It is also considered the 5th most lethal cancer in women worldwide (World Cancer Research Funds, 2018). Mortality rates are falling in Western countries, but this is more likely due to a change in lifestyle (use of oral contraceptives and declined use of postmenopausal hormone replacement) than to better treatment options (Malvezzi et al., 2016). The reasons for this high death rates are the absence of specific sign and symptoms, the lack of specific screening techniques and the extremely metastatic nature of the disease.

This means that most affected women are diagnosed already with advanced or metastatic stage of the disease. In addition, the majority of the patients develops resistance to chemotherapy that may be linked to the precence of CSCs. OCT4 was first linked to ovarian cancer in 2004 when Cheng et al. described OCT4 as a novel biomarker for ovarian dysgerminoma, an ovarian tumor derived from undifferentiated germ cells (Cheng et al., 2004). Afterwards, OCT4 was found to be expressed in immature teratoma of the ovary and in serous and mucinous epithelial ovarian cancer (Abiko et al., 2010; Zhang et al., 2010a). Interestingly, OCT4 expression was shown to be increased from normal surface ovarian surface epithelium/Fallopian tube epithelium, to benign/borderline tumors to high-grade serous carcinomas, further suggesting a relationship between OCT4 expression, tumor initiation and progression of serous ovarian cancer (Samardzija et al., 2015; Zhang et al., 2010a). The role of CSCs in ovarian cancer has been very little analyzed in the past, however more recently research on ovarian cancer CSCs as the origin of therapy resistance has been intensified. CSCs from ovarian cancer were firstly isolated via a serial dilution from ascites of a patient with advanced epithelial ovarian cancer (EOC). In this way, it was possible to isolate a single tumorigenic clone among a mixed population of cells from the ascites (the abnormal and pathological accumulation of serous fluid in the cavity of the abdomen). This clone was able to continuously give rise to new tumors when transplanted in mice via a stem and progenitor cell transformation(Bapat et al., 2005). In ovarian cancer, CSCs have found to be responsible for the generation of hundreds of tumor multicellular aggregates (spheroids) in the peritoneal cavity of advanced-stage patients after they detach from the primary tumor (Zeimet et al., 2012). CSCs are presumed to be responsible for chemoresistance and relapse in tumors. In fact, chemoresistant ovarian tumor cells are found to be enriched in CSC-like cells and activate stem cell pathways, including OCT4 expression, indicating that CSCs may contribute to the progression of the disease (Abubaker et al., 2013; Hu et al., 2010; Latifi et al., 2012; Samardzija et al., 2015). Considering the role of OCT4 as pluripotent regulator in developing embryo, the core of cancer spheroids may resemble the inner cell mass (ICM) structure that contains chemoresistant CSC-like cells, capable to evade chemotherapy and reform a secondary tumor. Similar to this, core spheroids cells expressing OCT4 in the ascites microenvironment will survive chemotherapy treatment and will serve as a niche for regenerating cancer cells. Thus, the

surviving spheroids are supposed to be the main cause of recurrence and treatment failure (Samardzija et al., 2012). Importantly, our research group recently pointed out an important role of OCT4 is able to drive ovarian cancer cell proliferation through NIPP1 and CCNF and the inhibition of PP1 (Comisso et al., 2017). In addition we also demonstrated that OCT4/NIPP1/CCNF axis promotes the expression of the central chromosomal passenger complex components to increase mitotic stability and improved chemoresistance (Comisso et al., 2017). Together, activation of these pathways leads to dramatically reduced overall survival of OCT4 high expressing, high grade ovarian cancer patients (Comisso et al., 2017).

1.2.3 Oct4 pseudogenes

The human OCT4 gene has at least seven processed pseudogenes in the human genome (Poursani et al., 2016). All of them have been shown to be transcribed in various cancer cell lines and tissues (Suo et al., 2005). Of these, OCT4-pg1, OCT4-pg3 and OCT4-pg4 have very similar exon structure when compared to the ancestral OCT4A splicing isoform (Figure 10).

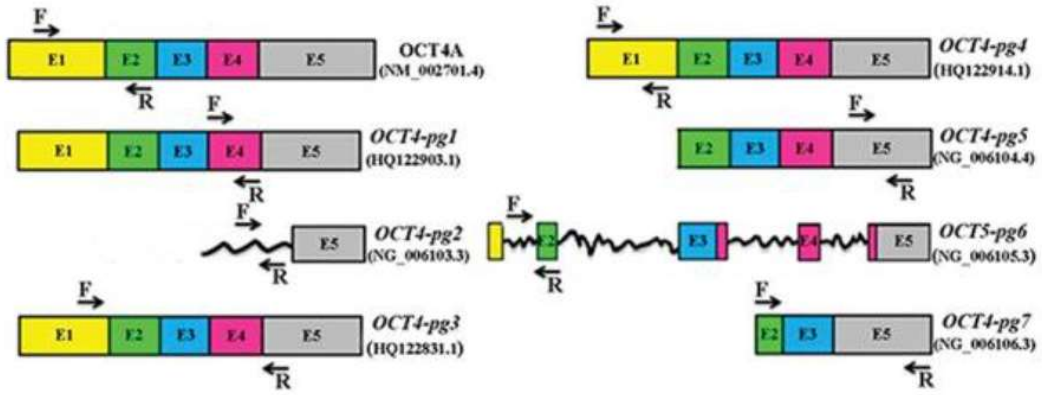

Figure 10: Schematic representation of OCT4 pseudogenes. OCT4-pg1, OCT4-pg3 and OCT4-pg4 have highly similar nucleotide sequences to that of the OCT4A transcript. OCT4-pg5 transcript lacks exon1, and OCT4-pg7 lacks exon1, exon4, and part of exon2. OCT4-pg2 has a part of exon5, and OCT4-pg6 has all five exons, incompletely. Rough lines in OCT4-pg2 and OCT4-pg6 are remained sequences which are derived from OCT4 introns. Adapted from Poursani et al. 2016

Several studies anticipate a relevance for pseudogene-derived ncRNAs in the regulation of the ancestral OCT4 gene in ESCs but also cancer (Jez et al., 2014; Kaltz et al., 2008; Kastler et al., 2010; Liedtke et al., 2007; Lin et al., 2007; Pain et al., 2005; Panagopoulos et al., 2008; Redshaw and Strain, 2010; Suo et al., 2005; Wezel et al., 2013; Zhao et al., 2011). Their expression is not always correlated to the one observed for OCT4 ancestral gene and have

been shown to be linked to stem cell capability but also to cancer susceptibility and aggressiveness. However, the exact mechanism through which they act remains mostly unknown. Nevertheless some mechanisms have been proposed: the human OCT4 pseudogene OCT4-pg4 was demonstrated to sponge OCT4-specific miRNAs in hepatocellular carcinoma cells via a ceRNA-mediated mechanism (Wang et al., 2013a); human OCT4-pg5 antisense transcripts, expressed in cancer but not in normal tissue (Suo et al., 2005), reduce OCT4 promoter activity via a mechanism which involves epigenetic remodeling complexes such as Ezh2 and G9a and this may be a mechanism by which the expression of OCT4 could be repressed in some human cancers (Hawkins and Morris, 2010). Interestingly, OCT4-pg1, OCT4-pg3 and OCT4-pg4 might potentially produce proteins. For instance, OCT4-pg1 transcript can produce a protein similar to OCT4A, containing the NTD, CTD and POU domain. Due to point mutations, OCT4-pg3 encodes a truncated protein with a complete NTD and a partial POUs domain. Hypothetical OCT4-pg4 protein misses a large part of CTD, but has intact NTD and POU domain. However, even if they are able to produce a protein it is unstable and so their activity must rely only on their lncRNAs (Poursani et al., 2016).

By contrast, only one murine Oct4 pseudogene (Oct4-pg1 also known as Oct4P1), which impacts on mesenchymal stem cell differentiation via a still unknown mechanism was known. In our laboratory, using genome wide bioinformatic and gene expression analysis, we demonstrated the transcription of four additional candidate processed pseudogenes found in the murine genome (Oct4P2, Oct4P3, Oct4P4 and Oct4P5) with high homology to Oct4 transcript 1 (Pou5f1) cDNA in both open reading frame and untranslated regions (UTRs) (Figure 11) (Scarola et al., 2015).

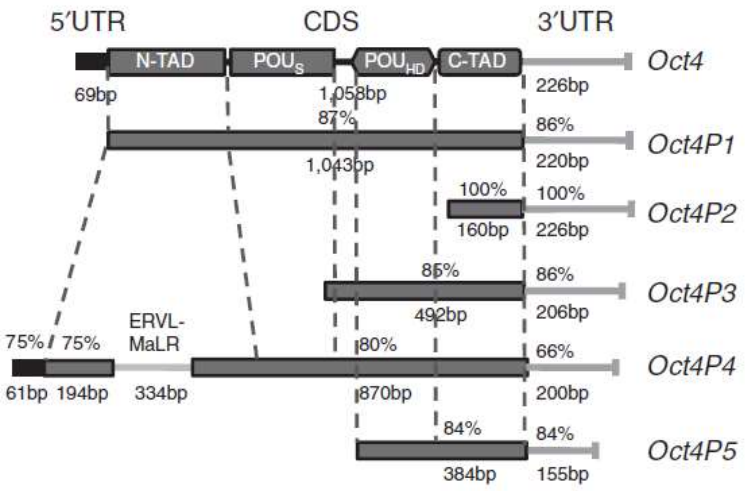

Figure 11: Schematic representation of Oct4 and Oct4 pseudogenes transcripts. Pseudogenes length and percentage of sequence homology to Oct4 are indicated. Regions with high homology to the Oct4 5′UTR or 3′UTR are indicated by black boxes or grey lines, respectively. ERVL-MaLRs, spliced region of Oct4P4 with homology to the LTR of ERVL-MaLRs retrotransposons. OCT4 protein domains are indicated: N-terminal trans-activating domain (N-TAD), POU-specific DNA-binding domain (POUs), DNA-binding homeodomain (POUHD), C-terminal trans-activating domain (C-TAD). Adapted from Scarola et al. 2015

mESC differentiation led to downregulation of Oct4P1 and activation of Oct4P2 and Oct4P4 expression while it does not impinge on Oct4P3 and Oct4P5 expression (Scarola et al., 2015). Also the subcellular localization is different among Oct4 pseudogenes with Oct4P1 mainly localizing to the cytoplasm, Oct4P2 and Oct4P4 lncRNAs to the nucleus. In contrast Oct4P3 and Oct4P5 lncRNAs do not have preferential localization (Scarola et al., 2015). Interestingly, Oct4P4 contains a 334 bp insertion which shows high homology to the ERLV-MaLRs retrotransposon family and, remarkably this insert is subjected to RNA splicing (Scarola et al., 2015). Thus, the unique properties of Oct4P4 such as extended homology with ancestral Oct4 gene, nuclear localization, splicing, upregulation during mESC differentiation and absence of any ORF prompted the extended investigation of mOct4P4 lncRNA. The nuclear localization excluded its function as ceRNA (which requires a cytoplasmatic localization for an efficient binding with miRNAs) and the absence of antisense transcripts hinted a role for a sense oriented based lncRNA function. Overexpression of Oct4P4 in self-renewing mESCs impaired self-renewal via a mechanism which involves an interference with transcriptional regulatory circuits that controls pluripotency via Oct4 ancestral gene. Remarkably, Oct4P4 lncRNA was found able to bind SUV39H1 and Oct4 ancestral gene promoter (Scarola et al., 2015). Binding was paralleled by an increased SUV39H1 HMTase content, and consequently elevated H3K9me3 (a marker of repressive heterochromatin) of the Oct4 promoter. Thus, the proposed model was that

sense-oriented Oct4P4 lncRNA is upregulated during mESC differentiation, forms a complex with SUV39H1 and then translocate to the promoter of the ancestral Oct4 gene, thus leading to heterochromatinization, reduced Oct4 expression and reduced expression of self-renewal markers and mESC differentiation (Figure 12) (Scarola et al., 2015).

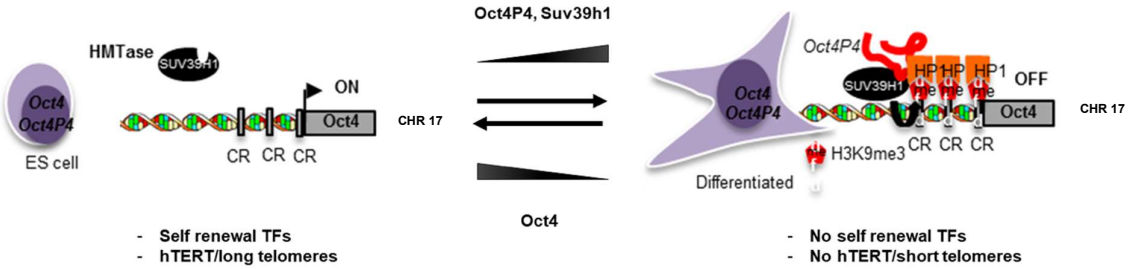

Figure 12: The nuclear, long non-coding Oct4P4 is upregulated during mESC differentiation and recruits the H3K9-specific HMTase SUV39H1 to impose H3K9me3 and HP1a at the promoter of ancestral Oct4 gene in trans, leading to Oct4 gene silencing. Adapted from Scarola et al. 2015

1.3. FUS RNA binding protein

The thesis work revealed an involvement of the RNA-binding protein FUS/TLS (Fused in sarcoma/Translocated in sarcoma) in Oct4P4 mediated regulation of Oct4 expression. FUS is a multifunctional protein component of the heterogeneous nuclear ribonuclein (hnRNP) complex (Figure 13). FUS/TLS is a member of the FET family of RNA-binding proteins, which have been implicated in cellular processes that include regulation of gene expression, maintenance of genomic integrity and mRNA/microRNA processing. This family is characterized by an N-terminal domain with a glutamine-glycine-serine-tyrosine (QGSY-rich region), a highly conserved RNA recognition motif (RRM) domain, and multiple RGG repeats and a C-terminal zinc finger motif (Crozat et al., 1993; Deng et al., 2014; Iko et al., 2004; Morohoshi et al., 1998; Prasad et al., 1994). *In vitro* studies demonstrated the ability of FUS/TLS to bind RNA, ssDNA, dsDNA (even if with lower affinity) and proteins (Baechtold et al., 1999; Crozat et al., 1993; Perrotti et al., 1998; Prasad et al., 1994; Ratti and Buratti, 2016; Wang et al., 2008; Zinszner et al., 1997). The exact FUS/TLS recognition motif on RNAs has not been established yet. Nevertheless, an *in vitro* selection (SELEX) revealed that a common GGUG motif has been identified in about half of the RNA sequences bound by FUS/TLS (Lerga et al., 2001). In addition, it has also been proposed that FUS/TLS could be able to interact with multiple RNA-structural motifs or recognizes secondary structures of target RNAs (Fujii, 2005). Finally, FUS/TLS has been shown to bind telomeric RNA $(UUAGGG)_4$ and single-stranded human telomeric DNA *in vitro* (Takahama et al., 2008).

Remarkably, FUS/TLS can bind its own pre-mRNA and autoregulate its expression (Zhou et al., 2013) through a mechanism which involves non-sense mediated decay. In addition, the autoregulation of FUS/TLS protein can also be mediated by specific miRNAs which are induced by FUS/TLS itself in a feed-forward regulatory mechanism (Dini Modigliani et al., 2014). FUS/TLS is a mainly nuclear RNA binding protein with a variety of functions at transcriptional and post-transcriptional level both in nuclear compartment as well as in the cytoplasm.

1.3.1 FUS as Transcription regulator

The QGSY-rich region at the N-terminal is responsible for FUS/TLS dimerization and, in turn, for chromatin binding and regulation of transcription initiation (Yang et al., 2014). FUS/TLS can associate with the general transcription machinery and may be able to regulate transcription initiation interacting with RNA polymerase II and mediating its phosphorylation during transcription (Bertolotti et al., 1996; Petermann et al., 1998; Schwartz et al., 2012; Yang et al., 2000). It has also been shown to associate with various other protein factors involved in the initiation of transcription (Law et al., 2006) such as several nuclear receptors (Powers et al., 1998) and gene-specific transcription factors such as Spi-1/PU.1 (Hallier et al., 1998) or Nf-κB (Uranishi et al., 2001). During its transcriptional regulatory activity FUS/TLS also regulates the alternative polyadenylation signals selected by numerous transcripts in neuronal cells in a position dependent manner (Masuda et al., 2015). In fact, FUS/TLS binding downstream an alternative polyadenylation signal was able to recruit polyadenylation machinery and to induce polyadenylation of the alternative short transcripts. By contrast, an upstream binding determined the downregulation of these RNA species (Masuda et al., 2015).

1.3.2 FUS as regulator of pre-mRNA splicing

Apart from transcriptional regulation, FUS/TLS has also been linked to splicing. Actually, it has been demonstrated to bind SMN protein, U1 small nuclear ribonucleoprotein (snRNP) and Sm-snRNP complex which are all essential members of the spliceosomal machinery (Gerbino et al., 2013; Yamazaki et al., 2012). FUS/TLS is important for the formation and maintenance of nuclear gems (a subnuclear structure involved in snRNP biogenesis) and mutations in FUS/TLS can lead to a general impairment of splicing since it

is able to bind nascent pre-mRNAs acting as a molecular mediator between RNA-pol II and U1-snRNP for splicing (Sun et al., 2015; Tsuiji et al., 2013; Yamazaki et al., 2012; Yu and Reed, 2015; Yu et al., 2015).

1.3.3 FUS modulates mRNA stability, transport and translation

FUS/TLS has also a role in mRNA stability, it is able to bind the 3'UTR sequence of selected target mRNAs (Colombrita et al., 2012; Hoell et al., 2011; Lagier-Tourenne et al., 2012; Rogelj et al., 2012) and in this way is able to control additional aspects of an mRNA life, including mRNA stability/decay. FUS/TLS was found to form a complex with factors controlling the mRNA 3'-end processing and polyadenylation (PAN2, PABC1 and CPSF6 proteins) (Udagawa et al., 2015). Evidence reports that FUS/TLS transports RNA granules into dendrites to regulate local translation at synapse (Belly et al., 2005; Fujii et al., 2005). In addition, FUS/TLS relocalizes at dendritic spines upon glutamate receptor activation and transports actin and actin-stabilizing factor mRNAs to regulate spine remodeling (Fujii, 2005).

1.3.4 FUS involvement in noncoding RNA processing

FUS/TLS is also involved in ncRNA maturation. It was found to localize at the Drosha complex (Gregory et al., 2004) and by binding nascent pri-miRNAs is able to recruit Drosha at chromatin sites of active transcription to promote pri-miRNA processing (Morlando et al., 2012). Also lncRNA are bound by FUS/TLS, in fact 30% of all literature annotated lncRNAs show a consistent binding to FUS/TLS (Ratti and Buratti, 2016).

1.3.5 FUS and DNA damage repair

Finally, FUS/TLS has also been linked to DNA damage repair. Indeed, the recruitment of FUS/TLS to double-strand breaks is very rapid (Wang et al., 2013b) and seems to orchestrate the DNA repair response mediated by its interaction with histone deacetylase 1 (HDAC1) and poly (ADP-ribose) polymerase (PARP). When DNA damage is induced, FUS/TLS is phosphorylated in its N-terminal Serine residues by ATM and DNA-PK kinases (Gardiner et al., 2008) and is able to promote D-loop formation and homologous recombination during DNA double strand break (Baechtold et al., 1999). In line with this, FUS/TLS loss-of-function resulted in increased DNA damage in neurons and an altered FUS/TLS nuclear

localization impairs the PARP-dependent DNA damage response (Naumann et al., 2018) leading to neurodegeneration. As a further response to DNA damage, FUS/TLS was shown to mediate the transcriptional inhibition of Cyclin D1 gene: after binding to several ncRNAs transcribed by Cyclin D1, FUS/TLS undergoes a protein conformation change that increases its binding affinity with two histone acetyltransferases (CBP and p300) repressing their transcriptional activity (Wang et al., 2008). Given the plethora of functions related to FUS/TLS and its ability to bind lncRNAs it could play a role also in still unknown pseudogene-derived lncRNA functions.

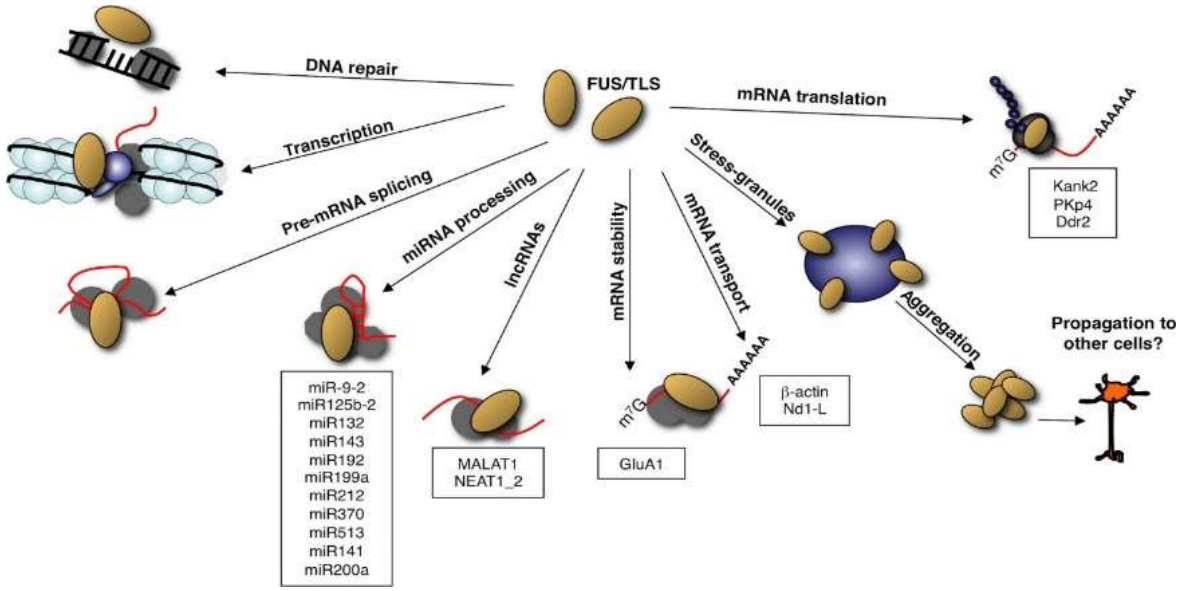

Figure 13: Schematic diagram of known FUS/TLS/TLS-regulated cellular functions in the nuclear and cytoplasmic compartments. The boxes below each function report only the names of the RNA targets that have been extensively validated at the experimental level. Possible causes of aggregation following stress granule formation and the consequent seeded spreading of the disease are also indicated. Figure taken from. Ratti and Buratti, 2016

2. Aim of the thesis

Only two percent of the whole genome encodes for protein-coding genes and the remaining 98 percent was, in the past, regarded as "junk DNA" (Wright and Bruford, 2011). However in recent years, great interest in understanding the non-coding part of the vertebrate genome has risen. It is believed that the vast majority of the genome is transcribed as non-protein-coding RNA (ncRNA) giving rise to a plethora of diverse ncRNAs with different functions ranging from gene regulation to protein translation going through mRNA stability regulation. One of the most studied class of ncRNAs is the long noncoding RNAs (lncRNAs), which are mostly studied for their role in epigenetic landscape remodeling. Examples of this are found for HOTAIR (Rinn and Chang, 2012) and XIST (Lee and Bartolomei, 2013; Schoeftner et al., 2006; Wutz et al., 2002) lncRNAs which act as molecular scaffolds to impose heterochromatin on HOXC/D loci and to silence X chromosome, respectively. Pseudogenes are defined as non-functional copies of protein coding genes that were recently demonstrated to produce lncRNA populations in a regulated manner, regulating important functions in normal cell physiology and in pathology. Differently from canonical lncRNAs, pseudogene-derived lncRNAs often control ancestral gene expression acting as competitive endogenous RNAs (ceRNAs), sponging miRNAs, given their high sequence similarity to the respective ancestral gene (Salmena et al., 2011). Recently, we proposed a new role for a murine Oct4 pseudogene, which is able to control ancestral gene expression via an epigenetic regulatory mechanism *in trans*. In particular, we demonstrated that the murine Oct4 pseudogene 4 (mOct4P4) encodes a nuclear restricted sense lncRNAs which is upregulated upon mESC differentiation induction. In this context, mOct4P4 is able to recruit a SUV39H1 containing silencing complex to the promoter of the ancestral Oct4 gene, inducing local H3K9me3 imposition and Oct4 gene silencing, promoting cell differentiation (Scarola et al., 2015). However, we lack mechanistic insights on mOct4P4 function. Thus, the aim of this study was to **obtain understandings into the molecular mechanism of mOct4P4 pseudogene derived lncRNA dependent gene silencing**. We aimed to characterize its role in a loss-of-function context in mESCs and identify functional relevant regions in the mOct4P4 lncRNA required for SUV39H1 recruitment and targeting of the ancestral Oct4 gene promoter. In addition we wanted to

identify novel interactors of the mOct4P4 lncRNA/SUV39H1 complex and clarify their role within it. Finally, we wanted to demonstrate functional conservation of the human mOct4P4 homolog (hOCT4P3) in controlling OCT4 gene expression in human ovarian cancer cells providing new insights on pseudogene-mediated regulation of OCT4 also in this context. Indeed, OCT4 expression have been shown to be linked to several human cancers, such as ovarian cancer, where its high expression is linked to poor prognosis and increased chemoresistance (Comisso et al., 2017).

3. Results

We recently demonstrated that the murine Oct4 pseudogene (mOct4P4) encodes a nuclear restricted lncRNA that is upregulated upon the induction of mESC differentiation. mOct4P4 is able to recruit a SUV39H1 containing silencing complex to the promoter of the ancestral Oct4 gene, mediating local imposition of H3K9me3 and silencing of the ancestral Oct4 gene upon mESCs differentiation (Scarola et al., 2015).

3.1 Murine Oct4 Pseudogene 4 is essential for mESC differentiation

In order to functionally characterize mOct4P4 lncRNA function, we generated a mOct4P4 loss-of-function mESC line using the CRISPR/dCas9 system. This genome editing tool consists in two components which are the Cas9 endonuclease and a target-identifying single guide RNA (sgRNA). The sgRNA base pairs with the target DNA and directs the Cas9 nuclease to this site. To obtain a transcriptional repression of the target gene, it is possible to use a catalytically dead Cas9 mutant (dCas9), which is defective in DNA cleavage. dCas9 is able to act as transcriptional repressor by sterically hindering the transcriptional activity of RNA polymerase (Figure 14A) (Qi et al., 2013). We used a vector expressing an HA-tagged version of catalytically inactive dCas9, fused to the Kruppel associated box (KRAB) repressor domain to enhance its silencing activity (La Russa and Qi, 2015). We also generated a plasmid expressing a sgRNA specifically targeting mOct4P4 promoter (sgOct4P4). HA-dCas9-KRAB plasmid was infected in mESC generating the "dCas9 empty" (dCas9 Ø) mESC cell line (Figure 14B, top panel). Subsequently, we infected the sgOct4P4 vector in dCas9 empty cells giving rise to the "dCas9 sgOct4P4" cell line (Figure 14B, bottom panel).

Endogenous mOct4P4 is almost undetectable in undifferentiated mESCs and it is efficiently upregulated during mESC embryoid body differentiation (Scarola et al., 2015). To confirm the importance of mOct4P4 in regulating mESC differentiation, we took advantage of the embryoid body (EB) ES cell differentiation protocol, which mimics early embryonic development (Rungarunlert et al., 2013). Firstly, we performed EB differentiation experiments to test whether expression of mOct4P4 is impaired in dCas9 sgOct4P4 cells. As expected, mOct4P4 expression increases during 10 days of EB differentiation in control cells

(dCas9 empty); by contrast, mOct4P4 upregulation is impaired in dCas9 sgOct4P4 EBs. Indeed, starting from the third day of differentiation single guide expressing cells showed 20% mOct4P4 RNA levels compared to control EBs (Figure 14C). We then tested whether the silencing of mOct4P4 maintains high Oct4 expression upon differentiation. Interestingly, starting with the sixth day of differentiation we observed a statistical significant increase (8/9 fold) in Oct4 mRNA in mOct4P4 knockdown cells compared to control cells (Figure 14D). This indicates that a reduction in mOctP4 levels impairs Oct4 repression upon differentiation.

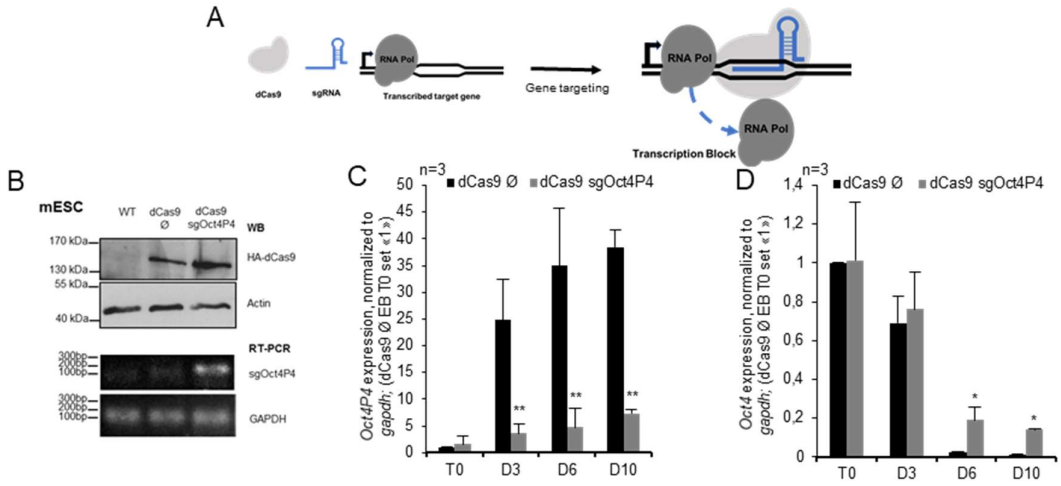

*Figure 14: (A) Schematic representation of the CRISPR/dCas9 targeting mOct4P4 promoter. (B) Top panel: Verification of dCas9-HA-KRAB (Addgene plasmid #50919) expression in embryonic stem cells by western blotting using an HA antibody. Actin was used as loading control. Bottom Panel: sgOct4P4 lncRNA expression as determined by semi-quantitative PCR. PCR products are visualized on a 1.2% agarose gel. (C) mOct4P4 lncRNA expression in mESC at Day 0, 3, 6 and 10 of embryoid body differentiation, as determined by quantitative RT-PCR. mOct4P4 expression was normalized to gapdh. sgOct4P4 expression downregulates mOct4P4 expression (D) Oct4 mRNA expression in mESC at Day 0, 3, 6 and 10 of embryoid body differentiation, as determined by quantitative RT-PCR. Oct4 expression was normalized to gapdh. Error bars represent s.d; *p<0.05: **p<0.01, n: number of independent experiments carried out on the same cell line*

To investigate whether impaired mOct4P4 expression in differentiating mESC is paralleled with persistent expression of self-renewal markers, we measured RNA expression of Sox2, Nanog and Gdf3. Interestingly, at Day 6 of EB differentiation we found a statistical significant increase in Sox2, Nanog and Gdf3 RNA levels in mOct4P4 loss-of-function EBs when compared to the control (Figure 15A). Furthermore, we investigated the expression of early differentiation markers Fgf5 and Nestin. In line with the expression of self-renewal markers, we found a decreased RNA expression for both Fgf5 and Nestin in differentiating dCas9 sgOct4P4 cells (Fig. 14B). Altogether, this data shows that mOct4P4 silencing can

block effective mESC differentiation, maintaining stem cell features even after the induction of mESC differentiation. Embryoid bodies are commonly produced to mimic cardiomyogenesis since they can form contractile structures, which represent an hallmark of effective cardiomyocyte differentiation (Rungarunlert et al., 2013). As shown in Figure 15C, mOct4P4 silencing in dCas9 sgOctP4 cells causes a reduction in the number of contractile EBs. The nature of the beating is another hallmark of effective differentiation. Interestingly, in line with the decreased number of contractile embryoid bodies, also the morphology and features of the beating contractile structures are modified upon mOct4P4 silencing. In fact, in control cells the beating involves the whole embryoid body, while in dCas9 sgOct4P4 cells the beating is limited to small foci (Figure 15D). Together, these data show that mOct4P4 is important to regulate the shift between self-renewal to differentiation and its precise regulation of expression is required for a correct differentiation.

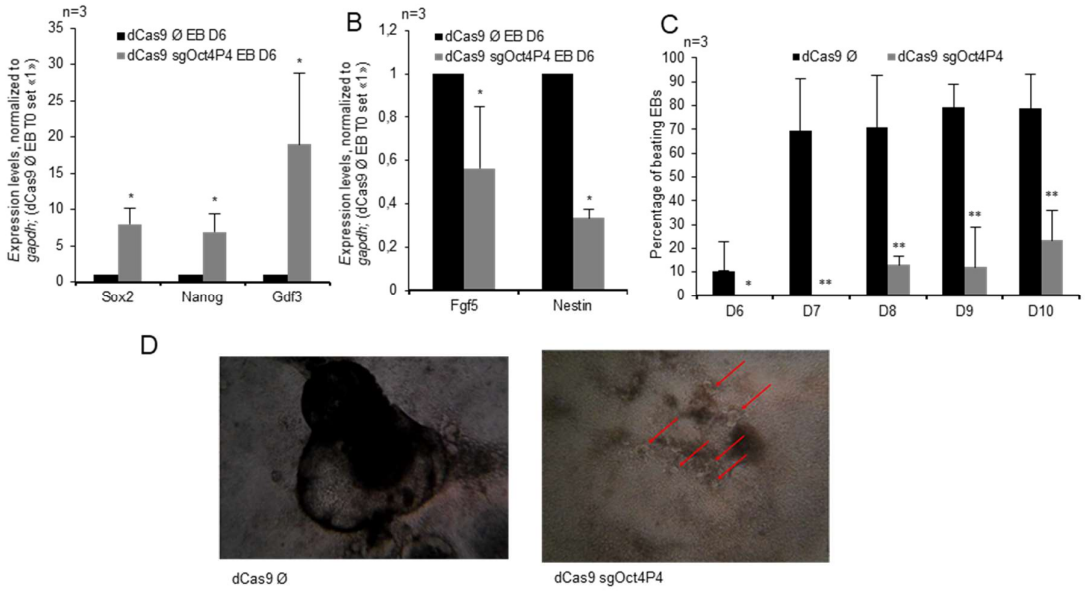

*Figure 15: (A) Self-renewal markers mRNA are upregulated in dCas9/sgOct4P4 mESC at Day 6 of embryoid body differentiation, as determined by quantitative RT-PCR. Expression levels were normalized to gapdh. (B) Early differentiation markers mRNA are downregulated in dCas9/sgOct4P4 mESC at Day 6 of embryoid body differentiation, as determined by quantitative RT-PCR. Expression levels were normalized to gapdh. (C) Percentage of beating embryoid bodies at day 6, 7, 8, 9 and 10 of differentiation are lower in dCas9/sgOct4P4 condition. (D) Representative images of adhesion grown embryoid bodies. Arrows in dCas9 sgOct4P4 indicate foci of beating. Error bars represent s.d; *p<0.05: **p<0.01; . n: number of independent experiments carried out on the same cell line*

Given the vast repertoire of vertebrate pseudogenes, we speculate that mOct4P4 mechanism of action could be conserved among species. For example, human genome

contains at least 7 OCT4 processed pseudogenes (Poursani et al., 2016), thus we hypothesize that a conservation of mOct4P4 mechanism of action could be found also for a human OCT4 pseudogene.

3.2 hOCT4P3 is the functional homolog of mOct4P4

Data obtained in our laboratory (Scarola et al., 2015) show that the human OCT4 pseudogene OCT4P3 (hOCT4P3) shares high structural and sequence homology with the murine Oct4P4 lncRNA (Figure 16A) suggesting a conserved role in the regulation of ancestral Oct4 gene. Information on subcellular localization of lncRNAs can provide important information about a possible function as nuclear restricted epigenetic regulator (as mOct4P4) or as cytoplasmatic ceRNA (Ponting et al., 2009; Tay et al., 2014). Similar to mOct4P4, hOCT4P3 is enriched in the nucleus of several ovarian cancer cell lines, such as OVCAR-3, SKOV3, TOV-112D and CAOV3 (Figure 16B). We previously demonstrated that ovarian cancer cell lines expressed the highest levels of OCT4 when compared to other cancer cell lines (Comisso et al., 2017). Co-expression of OCT4 and hOCT4P3 may enable functional interaction between these players that may impact on cancer cell features. The nuclear localization of hOCT4P3 lncRNA excludes its activity as ceRNA, even though its 3'UTR harbors conserved miRNA responsive elements (Poursani et al., 2016). We were then interested in testing whether hOCT4P3 could regulate OCT4 in human cancer cells. To achieve this aim, we ectopically expressed hOCT4P3 in OVCAR-3 cells. OVCAR-3 cells are extensively used as model of high grade serous ovarian cancer and were recently shown to express elevated levels of OCT4 (Comisso et al., 2017). Importantly, ectopic expression of hOCT4P3 was able to induce a significant reduction of OCT4 ancestral gene expression which was paralleled by a reduced expression of self-renewal markers SOX2, NANOG and KLF4 (Figure 16C). This data underlines a conserved role for hOCT4P3 in controlling OCT4 ancestral gene expression in human cancer cells.

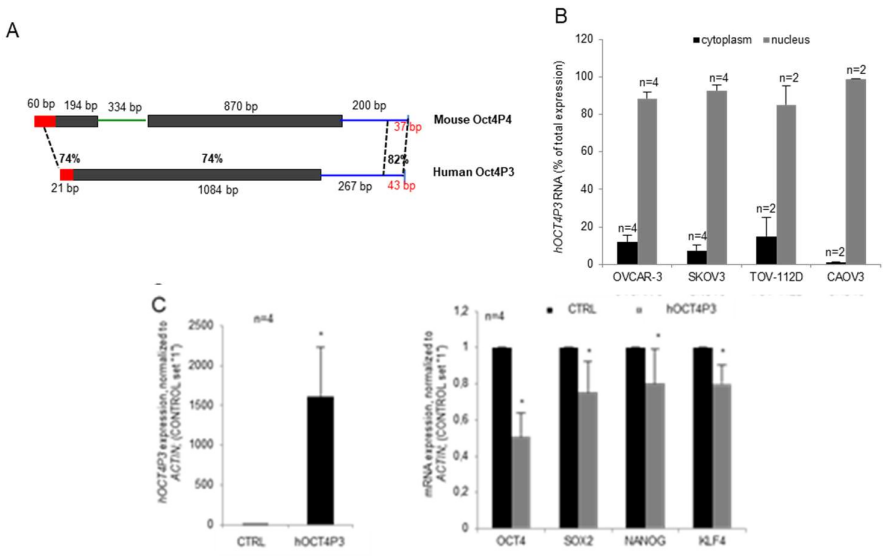

*Figure 16: (A) Alignment of mouse Oct4 pseudogene mOct4P4 with human OCT4 pseudogene hOCT4P3. Pseudogene length and percentage of sequence homology are indicated. Sequence identity is presented in percentage to mouse Oct4P4. Thin lines represent 3'UTR at the 3' ends of the transcripts and a centrally located, spliced fragment exclusively present in mOct4P4. (B) Subcellular localization of hOCT4P3 pseudogene in human Ovarian Cancer cell lines OVCAR-3, SKOV3, TOV-112D and CAOV3. Expression values are shown as percentage of total RNA expression, as determined by quantitative real time (RT-PCR). (C) hOCT4P3, OCT4, SOX2, NANOG and KLF4 mRNA levels are lower in OVCAR-3 hOCT4P3 overexpressing cells as determined by qRT-PCR. Error bars represent s.d. *p<0.05. n: number of independent experiments carried out on the same cell line*

To provide evidence for a cross-talk between hOCT4P3 lncRNA and the promoter of the OCT4 parental gene, we established a hOCT4P3-MS2 RNA stemloop tethering system in OVCAR-3 cells: the hOCT4P3 cDNA, fused to 24 MS2 (phage coat protein) binding sites (OCT4P3-24xMS2 stemloop) was stably co-expressed with flag-tagged MS2 protein (MS2-flag) which binds with high affinity to the MS2 RNA stemloops. This model system allows the study of trafficking, localization and interaction proteins with the hOCT4P3 lncRNA using an anti-flag antibody in chromatin and RNA Immunoprecipitation experiments (Figure 17A). The system was tested and validated for further experiments: in fact, hOCT4P3-24xMS2 stemloop overexpression (Figure 17B) caused a decrease of OCT4 mRNA levels paralleled by significantly reduced OCT4 protein levels (Figure 17C). Importantly, we were able to amplify hOCT4P3 RNA by RT-PCR in an anti-flag RIP

experiment using OVCAR-3 cells stably co-expressing MS2-flag and hOCT4P3 24xstemloop constructs (Figure 17D).

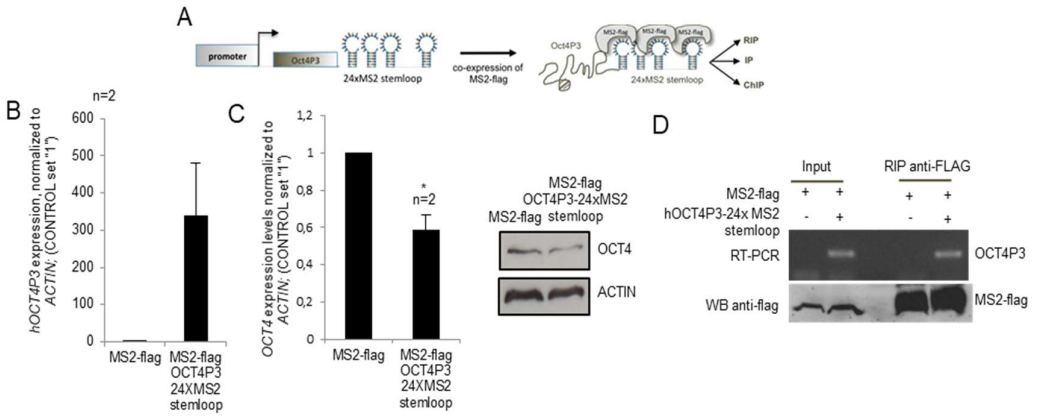

*Figure 17: (A) Schematic representation of the hOCT4P3-24xMS2 stemloop construct used for the generation of the MS2-flag/OCT4P3-24xMS2 cell model system. (B) Left panel: ectopic hOCT4P3 stable expression in hOCT4P3-24xMS2 stemloop OVCAR-3 cells, as determined by qRT-PCR. (C) hOCT4P3-24xMS2 stemloop RNA overexpression reduces OCT4 expression, as determined by qRT-PCR (left) and western blot (right). (D) Anti-flag RNA immunoprecipitation (RIP) using OVCAR-3 cells overexpressing MS2-flag/OCT4P3-24xMS2 stemloop RNA or MS2-flag control cells. qRT–PCR followed by agarose gel electrophoresis verified the presence of OCT4P3-24xMS2 RNA in anti-flag RIP (top). Immunoprecipitation of MS2-flag validated by western blot (bottom) Error bars represent s.d., *p<0.05. n: number of independent experiments carried out on the same cell line*

To demonstrate that hOCT4P3 localizes to the OCT4 ancestral gene promoter, we performed anti-flag ChIP experiments taking advantage of the hOCT4P3-MS2 tagging system. We demonstrated the presence of MS2-flag protein at the ancestral OCT4 promoter only when co-expressed with hOCT4P43-24xMS2 stemloop RNA (Figure 18A). This data indicate that, in analogy to mOct4P4 lncRNA in mESCs (Scarola et al., 2015), hOCT4P3 is able to bind OCT4 ancestral gene promoter in OVCAR-3 cells. To further confirm the conservation of hOCT4P3 lncRNA function, we moved to test whether hOCT4P3 lncRNA physically interacts with the SUV39H1 HMTase in OVCAR-3 cells. Anti-flag RIP experiments demonstrated that SUV39H1 co-immunoprecipitates with MS2-flag only in hOCT4P3-24xMS2 stemloop RNA expressing OVCAR-3 cells, demonstrating SUV39H1/hOCT4P3 lncRNA binding (Figure 18B). Since SUV39H1 is a H3K9-specific HMTase, we wished to test whether hOCT4P3, when interacting with SUV39H1 is able to impose H3K9me3 at the ancestral OCT4 gene promoter. We performed anti-H3K9me3 ChIP experiments on hOCT4P3-24xMS2 stemloop OVCAR-3 cells showing that hOCT4P3

overexpression resulted in a fivefold increase of H3K9me3 at ancestral OCT4 promoter (Figure 18C). Altogether, these data indicate that hOCT4P3 in complex with SUV39H1 is able to impose H3K9me3 repressive heterochromatin marks at the OCT4 promoter leading to OCT4 repression.

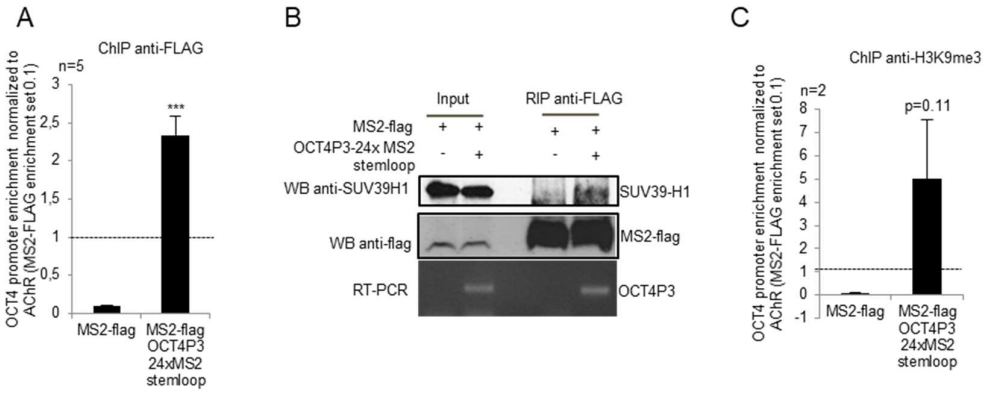

*Figure 18: (A) Chromatin immunoprecipitation (ChIP) analysis of OCT4 promoter regions in MS2-flag/OCT4P3-24-MS2 stemloop and MS2-flag OVCAR-3 cells using anti-flag antibodies. ChIP data were quantified versus input and unrelated HA-specific antibodies. ChIP data confirmed hOCT4P3 presence at OCT4 promoter. (B) Anti-flag RIP using MS2-flag/OCT4P3-MS2 stemloop OVCAR-3 cell model system. Western blot for MS2-FLAG and SUV39H1. qRT–PCR followed by agarose gel electrophoresis verified the presence of OCT4P3-24xMS2 RNA and SUV39H1 protein in anti-flag RIP. (C) ChIP analysis of OCT4 promoter regions in hOCT4P3 OVCAR-3 cells using H3K9me3 antibodies demonstrated an increased heterochomatinization of OCT4 promoter in hOCT4P3 expressing OVCAR-3 cells. ChIP data were quantified versus input and the unrelated AChR promoter. Error bars represent s.d.; *p<0.05; **p<0.01; ***p<0.001; n: number of independent experiments carried out on the same cell line*

We next wanted to confirm the role of hOCT4P3 in the epigenetic control of ancestral OCT4 gene also in a loss-of-function condition. To this aim, we generated a hOCT4P3 loss-of-function cell line taking advance of the CRISPR/dCas9 system. We expressed dCas9-HA-KRAB in OVCAR-3 cells generating "dCas9 empty" OVCAR-3 cell line in which we co-expressed a sgRNA targeting hOCT4P3 promoter ("dCas9 sgOCT4P3" cell line) (Figure 19A). As expected, dCas9/sgOCT4P3 co-expression resulted in decreased hOCT4P3 lncRNA expression as determined by qRT-PCR (Figure 19B). We next tested whether repression of hOCT4P3 expression releases silencing of ancestral OCT4. As expected, "dCas9 sgOCT4P3" OVCAR cells show a significantly increased OCT4 RNA content compared to "dCas9 empty" cells (Figure 19C), which is paralleled with an increase in OCT4 protein expression (Figure 19D). Finally, to study the chromatin status of OCT4 promoter, we performed anti-H3K9me3 ChIP experiments demonstrating a fivefold reduction of H3K9me3 on OCT4 ancestral gene promoter in dCas9 sgOCT4P3 OVCAR-3

cells (Figure 19E). These loss-of-function experiments confirm a role for the hOCT4P3 lncRNA in controlling OCT4 expression at the epigenetic level.

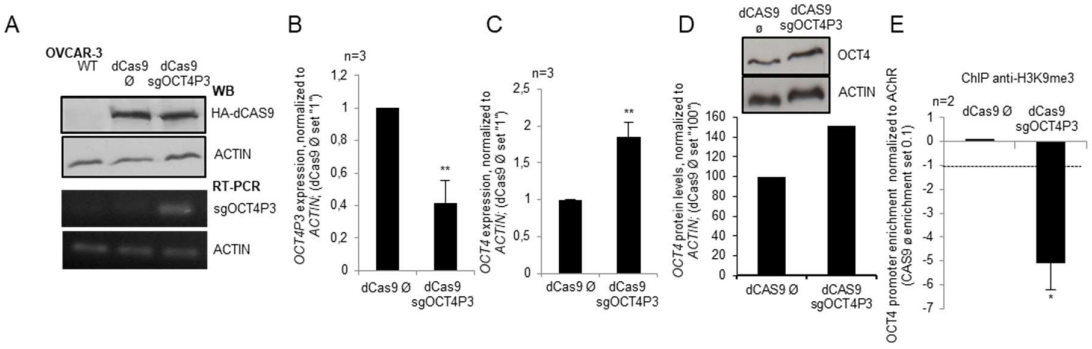

*Figure 19: (A) Certification of dCAS9-HA-KRAB (Addgene plasmid #50919) expression in OVCAR-3 cells by western blotting, using an HA antibody (top panel) and sgOCT4P3 expression evaluation in OVCAR-3 cells by qRT-PCR, visualized on a 1.2% agarose gel (bottom panel). (B) hOCT4P3-lncRNA expression is reduced in OVCAR-3 cells expressing dCas9 and sgOCT4P3, as determined by qRT-PCR. (C) Specific knockdown of hOCT4P3 increases OCT4 expression in OVCAR-3 cells as determined by qRT-PCR. (D) Specific knockdown of hOCT4P3 increases OCT4 expression in OVCAR-3 cells as determined by western blotting (Top panel) representative image of OCT4 western blotting in experimental cell lines. (Bottom panel) OCT4 band intensity quantification (E) ChIP analysis of OCT4 promoter regions in hOCT4P3 knockdown OVCAR-3 cells using H3K9me3 antibodies verified a reduced H3K9me3 presence on OCT4 promoter in hOCT4P3 knockdown cells. ChIP data were quantified versus input and the unrelated AChR promoter. Error bars represent s.d.; *p<0.05, **p<0.01; n: number of independent experiments carried out on the same cell line*

Altogether, these our data show that Oct4 pseudogene lncRNA-mediated silencing of ancestral Oct4 gene is a conserved mechanism in vertebrates.

3.3 mOct4P4 deletion studies to identify functionally relevant lncRNA regions

In order to get insights into functional relevant mOct4P4 lncRNA domains involved in SUV39H1 recruitment and Oct4 promoter binding we performed a mOct4P4 deletion analysis using mESCs as model system. Previous findings demonstrated that the sequences in mOct4P4 lncRNA that show high homology to the Oct4 5′ and 3′UTR (60bp at 5′ UTR and the 82bp at the 3′end of mOct4P4 lncRNA) were important for mOct4P4 nuclear localization and thus cooperate to repress the expression of the Oct4 gene (Scarola et al., 2015). We generated an additional panel of mOct4P4 deletion constructs progressively deleting 200 bp regions starting from the 3′UTR of the full length lncRNA, but maintaining the previously identified 5′ and 3′ high homology regions. In this way we were able to

create the constructs mOct4P4 - Δ200, Δ400, Δ600, Δ800, Δ994 and 5'+3' respectively (Figure 20A). Vectors were infected and stably expressed in mESCs reaching ectopic expression levels that were comparable to the full-length construct (Figure 20B). We first tested whether these constructs maintain a nuclear localization. qRT-PCR using total RNA from nuclear and cytoplasmic fractions of undifferentiated mESCs revealed a relative enrichment in the nucleus for all the deletion constructs encoded lncRNAs (Figure 20C), excluding a possible impairment of Oct4 silencing caused by a defective lncRNA localization. We next wanted to investigate the effect of deletion construct expression on Oct4 ancestral gene expression. Interestingly, only full length, Δ200 and Δ400 were able to silence Oct4 ancestral gene expression, as determined by qRT-PCR (Figure 20D) and western blotting (Figure 20E). In contrast, all other constructs (Δ600, Δ800, Δ994 and 5'+3') did not show any effect on Oct4 RNA and protein expression (Figure 20D and E).

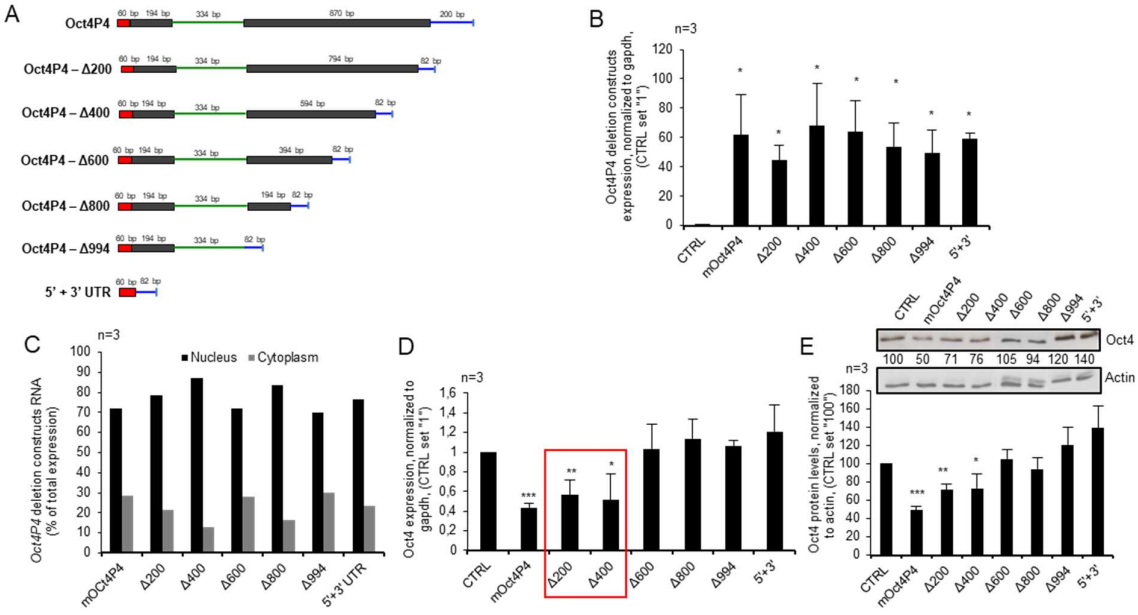

*Figure 20: (A) Schematic representation of the mOct4P4 deletion constructs. Thin lines represent 3'UTR at the 3' ends of the transcripts and a centrally located, spliced fragment (B) mOct4P4 deletion constructs expression in mESCs as determined by qRT-PCR. (C) Subcellular localization of mOct4P4 deletion constructs in mESCs. Expression values are shown as percentage of total RNA expression, as determined by quantitative real time (RT-PCR). (D) Oct4 expression in deletion construct expressing mESCs, as determined by quantitative RT-PCR. (E) Oct4 expression in experimental mESCs as determined by western blotting. (Top panel) representative image of Oct4 western blotting in deletion construct expressing mESCs. Actin was used as loading control. Numbers represent Oct4/Actin ratio (control was set to 100) (Bottom panel) Oct4 band quantification of three independent experiments. Error bars represent s.d. *p<0.05, **p<0.01, ***p<0.001; n: number of independent experiments carried out on the same cell line*

We next wished to understand if the effects observed in Δ200 and Δ400 mESCs on Oct4 gene expression were mediated by H3K9me3 imposition by the SUV39H1/mOct4P4 lncRNA complex. Exploiting the MS2 tethering system, we carried out anti-flag ChIP experiments on mOct4P4-24xMS2 stemloop deletion constructs to test their interaction with the promoter of the ancestral Oct4 gene. Interestingly, we found MS2-flag at the Oct4 ancestral gene promoter only upon expression of the full length mOct4P4-24xMS2 stemloop, the Δ200-24xMS2 stemloop and the Δ400-24xMS2 stemloop lncRNAs (Figure 20A). We next wished to test which mOct4P4 deletion constructs maintain the ability to bind SUV39H1. For this purpose, we performed anti-SUV39H1 RIP experiments. Remarkably, we were able to specifically amplify lncRNA derived from mOct4P4, Δ200 and Δ400 deletion constructs by qRT-PCR in anti-SUV39H1 RIP eluates (Figure 21B). We did not detect any interaction of SUV39H1 with any other lncRNA produced by Δ600, Δ800, Δ994 and 5'+3' deletion constructs and in unrelated anti-HA RIP eluates (Figure 21B).

This indicates that mOct4P4 physically interacts with SUV39H1 using a region spanning positions 984-1183 of the mOct4P4 lncRNA.

Finally, we wanted to confirm that mOct4P4 deletion constructs Δ200 and Δ400 are able to impose a H3K9me3 repressive histone mark at the gene promoter of the ancestral Oct4. Thus, we performed an anti-H3K9me3 ChIP analysis using a primer pair that specifically amplify Oct4 promoter region. ChIP data confirmed our previous data, demonstrating that Oct4 promoter showed an enrichment in H3K9me3 content only upon expression of full length, Δ200 and Δ400 mOct4P4 lncRNA (Figure 21C).

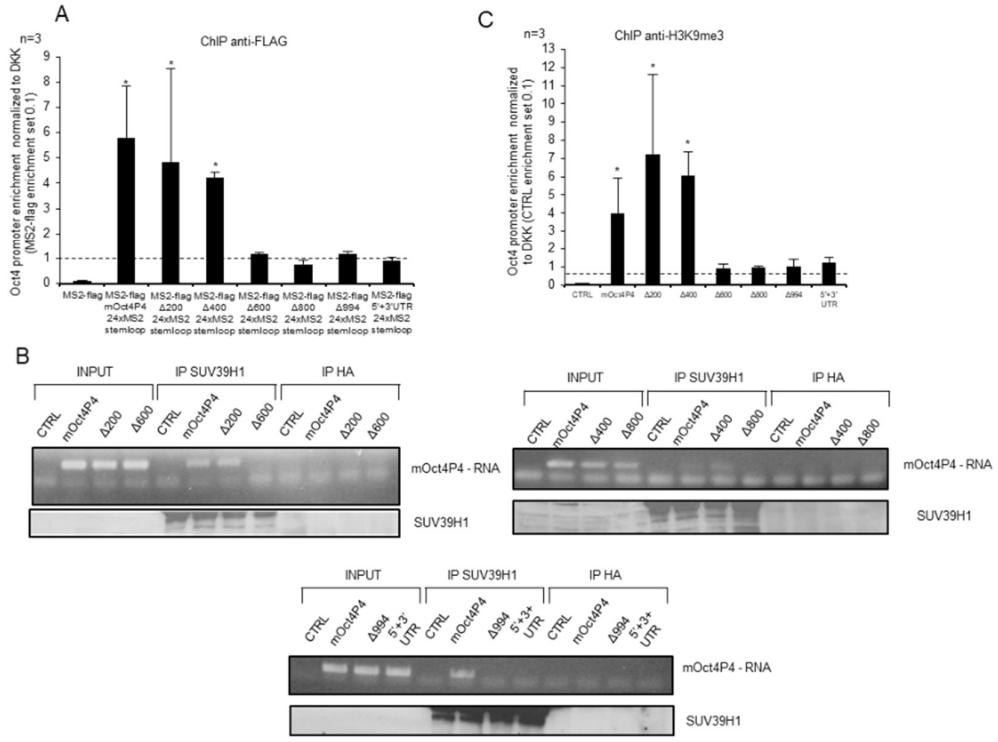

*Figure 21: (A) ChIP analysis of Oct4 promoter region in mOct4P4 deletion constructs expressing mESCs using anti-flag antibody. ChIP data were quantified versus input and unrelated DKK1 promoter. mOct4P4, Δ200 and Δ400 localize to the Oct4 promoter. (B) Anti-SUV39H1 RIP using mESC model system. (Top panel) qRT–PCR followed by agarose gel electrophoresis. (Bottom panel) Western blot for SUV39H1 confirms the immunoprecipitation. Anti-HA RIP was used as negative control. mOct4P4, Δ200 and Δ400 are able to bind SUv39H1 (C) ChIP analysis of Oct4 promoter region in mOct4P4 deletion constructs expressing mESCs using anti-H3K9me3 antibody. ChIP data were quantified versus input and unrelated DKK1 promoter. Increased H3K9me3 is found only in mOct4P4, Δ200 and Δ400 mESCs. Error bars represent s.d. *p<0.05. n: number of independent experiments carried out on the same cell line*

These data identify a functional mOct4P4 region that is able to recruit the mOct4P4 lncRNA/SUV39H1 complex to the Oct4 promoter and, in turn, to impose H3K9me3 to silence Oct4 expression. The minimum functional region appears to comprise a 200 nucleotide sequence (base pairs 984-1183 in mOct4P4) upstream the deletion originating Δ400 construct and hence deleted starting from the unfunctional Δ600 deletion construct.

3.4 Functional characterization of crucial mOct4P4 lncRNA region

In order to functionally characterize the mOct4P4 region spanning positions 984-1183, we generated two additional constructs. The first (hereafter referred as 200bp-mOct4P4) comprises mOct4P4 base pairs 984-1183 fused with the nuclear localization 5′ and 3′ mOct4P4 regions. The second (hereafter referred as -200bp-mOct4P4) is a mOct4P4

construct lacking base pairs 984-1183 (Figure 22A). We hypothesized that 200bp-mOct4P4 could resemble the features of full length mOct4P4 in terms of SUV39H1 complex formation and Oct4 gene promoter silencing, while -200bp-mOct4P4 should still localize to the nucleus, but neither bind Oct4 promoter nor SUV39H1 HMTase. Firstly, both constructs were infected and stably expressed in mESCs (Figure 22B). We next tested the subcellular localization of these constructs. As expected, preferential nuclear localization of both 200bp-mOct4P4 and -200bp-mOct4P4 constructs was confirmed by cytoplasmatic/nuclear RNA sub fractionation followed by qRT-PCR (Figure 22C). Exploiting the MS2 tethering system, we performed anti-flag ChIP experiments on mOct4P4-24xMS2 stemloop, 200bp-mOct4P4-24xMS2 stemloop and 200bp-mOct4P4-24xMS2 stemloop constructs to verify the ability of these lncRNAs to bind the Oct4 promoter. Remarkably, MS2-flag showed an enrichment to Oct4 promoter only when coexpressed with mOct4P4-24xMS2 stemloop or 200bp-mOct4P4-24xMS2 stemloop, but not with -200bp-mOct4P4-24xMS2 stemloop construct (Figure 22D). This demonstrates, again, that the Oct4 promoter is bound by mOct4P4 constructs only in presence of mOct4P4 base pairs 984-1183 (mOct4P4 minimal region) and this minimal region is sufficient to direct mOct4P4/Oct4 promoter binding. We analyzed SUV39H1 binding to this minimal region by an anti-SUV39H1 RIP. Full length mOct4P4 and 200bp-mOct4P4 lncRNA were able to be co-immunoprecipitated with SUV39H1 in anti-SUV39H1-RIP (Figure 22E), however, we were not able to amplify -200bp-Oct4P4 lncRNA in anti-SUV39H1 RIP eluates and any lncRNA in unrelated anti-HA RIP (Figure 22E). This data shows that mOct4P4 minimal region is necessary and sufficient for SUV39H1 binding. Subsequently, we investigated whether 200bp-mOct4P4 lncRNA/SUV39H1 and 200bp-mOct4P4/Oct4 promoter binding are followed by an increased H3K9me3 content on Oct4 promoter. ChIP experiments performed with anti-H3K9me3 antibodies showed a comparable enrichment of repressive heterochromatin marker on Oct4 promoter only in presence of mOct4P4 and 200bp-mOct4P4 but not in mESCs ectopically expressing -200bp-mOct4P4 (Figure 22F). This confirms the previous findings about the necessity and sufficiency of the minimal functional mOct4P4 region for mOct4P4 action. Finally, we measured whether increased H3K9me3 at the Oct4 promoter was paralleled with a decreased expression of Oct4 both at protein and RNA level. As expected, mOct4P4 and 200bp-mOct4P4 expression led to Oct4 mRNA downregulation

(Figure 22G) that is paralleled by a reduction in Oct4 protein level (Figure 22H). In line with the unchanged levels of H3K9me3 in -200bp-mOct4P4 mESCs we did not observe any change in Oct4 protein and RNA levels in these cells (Figure 22G and H).

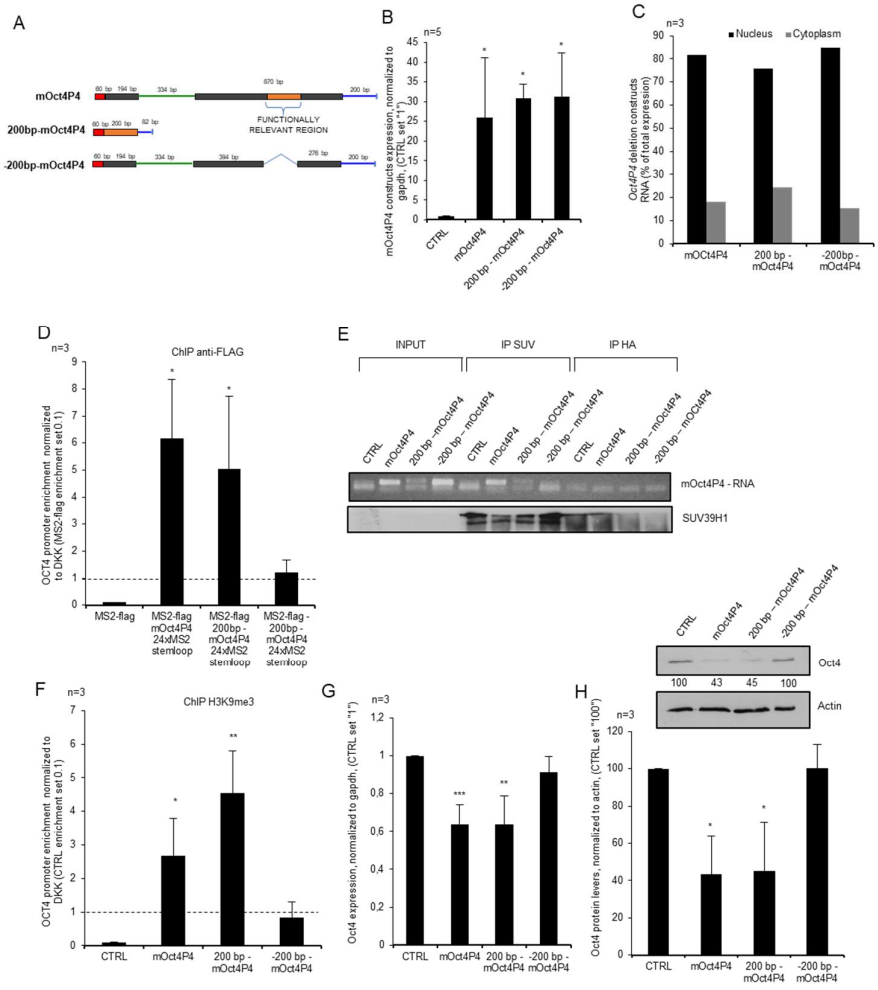

*Figure 22: (A) Schematic representation of the mOct4P4 deletion constructs. Thin lines represent 3'UTR at the 3' ends of the transcripts and a centrally located, spliced fragment. Orange box indicate identified minimal 200bp region. Red boxes indicate 5'UTR at the 5' ends of the transcripts. (B) Expression levels of 200bp-mOct4P4 and -200bp-mOct4P4 RNA in experimental mESCs. (C) Subcellular localization of the indicated constructs. Expression values are shown as percentage of total RNA expression, as determined by quantitative real time (RT-PCR). (D) ChIP analysis of Oct4 promoter region in indicated cell lines using anti-flag antibody. ChIP data were quantified versus input and unrelated DKK promoter. Oct4 promoter binding is found only for mOct4P4/24xMS2 stemloop and 200bp-mOct4P4/24xMS2 stemloop constructs. (E) Anti-SUV39H1 RIP using mESCs 200bp-mOct4P4 and -200bp-mOct4P4 stable mESCs cell lines. qRT-PCR followed by agarose gel electrophoresis verified the presence of mOct4P4 and 200bp-mOct4P4 RNAs in anti-SUV39H1 RIP (Top panel) Western blotting for SUV39H1 confirms the immunoprecipitation (bottom panel). Anti-HA RIP was used as negative control (F) H3K9me3 ChIP analysis of Oct4 promoter region in indicated cell lines. ChIP data were quantified versus input and unrelated DKK promoter. Oct4 promoter H3K9me3 enrichment is found only for mOct4P4 and 200bp-mOct4P4 constructs. (G) Oct4 expression is downregulated only in mOct4P4 and 200bp-mOct4P4 mESCs as determined by qRT-PCR (H) (Top panel) Oct4 expression in the indicated experimental cells as determined by Western blotting. Actin was used as loading control. Numbers represent Oct4/Actin ratio (control was set to 100). (Bottom panel) Western blot quantification of three independent experiments using actin as loading control. Error bars represent s.d. *p<0.05, **p<0.01, ***p<0.001; n: number of independent experiments carried out on the same cell line*

Together, these findings further identify the base pairs 984-1183 of mOct4P4 lncRNA as the minimum mOct4P4 functional region, sufficient for directing mOct4P4 to the Oct4 gene, SUV39H1 recruitment, heterochromatin formation and silencing Oct4 ancestral gene expression.

3.5 The role of SUV39H1 role mOct4P4 for minimal functional region silencing activity

SUV39H1 plays an important role in mOct4P4 lncRNA function in terms of Oct4 promoter binding and H3K9me3 deposition (Scarola et al., 2015). Since we showed that SUV39H1 specifically binds mOct4P4 minimal region (Figure 22E), we hypothesized that this important role is conserved also for the minimal functional region when expressed alone. To test this hypothesis, we transfected mOct4P4 and 200bp-mOct4P4 mESCs with a SUV39H1-specific siRNA oligo. An impairment of SUV39H1 function by RNA interference (Figure 23A) rescues Oct4 protein (Figure 23B) and mRNA levels (Figure 23C) in both mOct4P4 and 200bp-mOct4P4 cells. This rescue is paralleled to a reduced H3K9me3 content on Oct4 promoter as determined by an anti-H3K9me3 ChIP experiment performed on these cell lines (Figure 23D) indicating that SUV39H1 is necessary also for mOct4P4 minimal region mediated Oct4 silencing. Furthermore, an anti-flag ChIP experiment was carried out in order to determine whether SUV39H1 knockout impairs 200bp-mOct4P4-24xMS2 stemloop localization on Oct4 promoter. Interestingly, absence of SUV39H1 led to a decreased MS2-flag presence on Oct4 ancestral gene promoter in both mOct4P4-24xMS2 stemloop and 200bp-mOct4P4-24xMS2 stemloop expressing mESCs, indicating an impaired Oct4 promoter binding for mOct4P4 and its minimal region upon SUV39H1 silencing (Figure 23E). Altogether, these data indicate that SUV39H1 is essential for the function of mOct4P4 and its minimal functional region in terms of mOct4P4 promoter binding and H3K9me3 deposition.

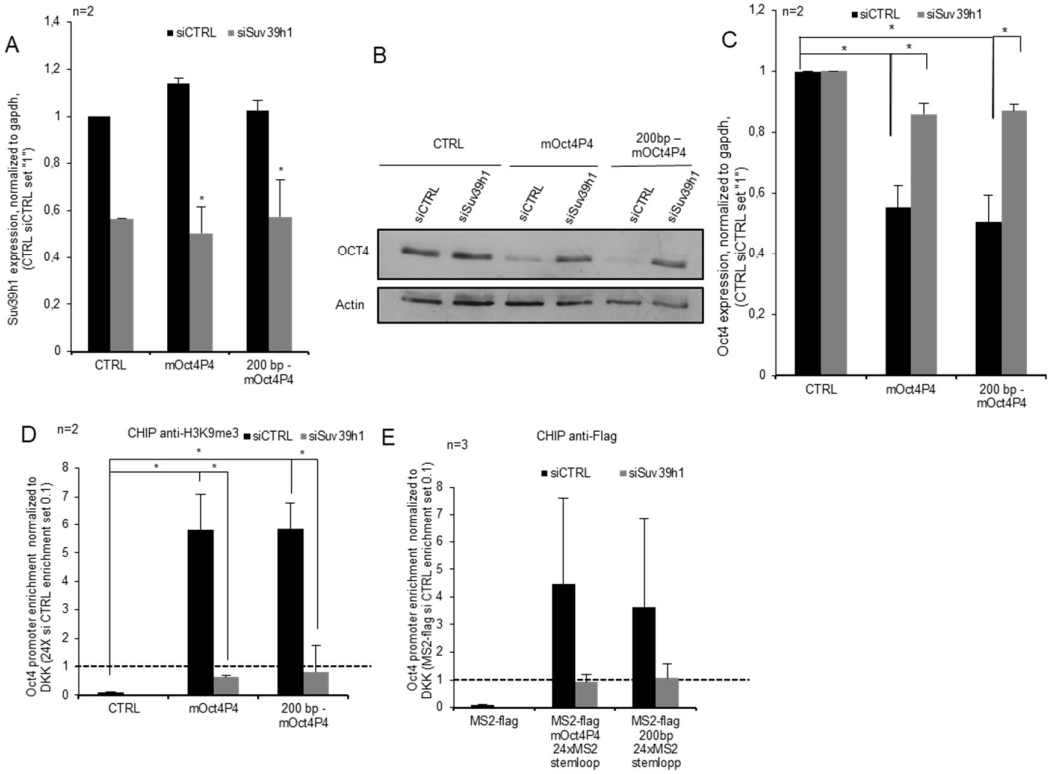

*Figure 23: (A) SUV39H1 expression in mOct4P4 and 200bp-mOct4P4 mESCs transiently transfected with SUV39H1 siRNA oligo as determined by qRT-PCR. (B) SUV39H1 silencing rescues Oct4 expression in experimental cells as determined by western blotting. Actin was used as loading control (C) SUV39H1 silencing rescues Oct4 RNA expression in experimental cells as determined by qRT-PCR. (D) Anti-H3K9me3 ChIP analysis of Oct4 promoter region under SUV39H1 knock-down conditions. SUV39H1 silencing rescues H3K9me3 levels on Oct4 promoter. ChIP data were quantified against input and unrelated DKK promoter. (E) ChIP analysis of Oct4 promoter region in mOct4P4 overexpressing cell lines transfected with the indicated siRNA using anti-flag antibody. mOct4P4 and 200bp-mOct4P4 presence on Oct4 promoter is reduced upon SUV39H1 silencing. ChIP data were quantified versus input and unrelated DKK promoter. Error bars indicate s.d.;*p<0.05; **p<0.01; ***p<0.001 n: number of independent experiments carried out on the same cell line*

3.6 The long non coding RNA mOct4P4 meets FUS (Fused in Sarcoma)

In order to better define the mOct4P4 lncRNA/SUV39H1 epigenetic remodeling complex, mass spectrometry analysis was carried out on anti-flag RIP eluates obtained from mESC carrying the mOct4P4-24xMS2stemloop/MS2-flag tethering system and control cells expressing only MS2-flag and the endogenous, un-tagged Oct4P4 lncRNA. We identified bands which are present only in the mOct4P4-24xMS2 stemloop anti-flag eluates. Proteomic analysis on these bands revealed several mOct4P4 interaction partners such as Peptidyl-prolyl cis-trans isomerase B (PPIB), High Mobility Group Box 1/2 (HMGB1/2), fused in

Sarcoma/Translated in Sarcoma (FUS/TLS) RNA binding protein, Eukaryotic translation Initiation Factor 4B (eIF4B) and nucleolin (NCL) (Figure 24A). Interestingly, we did not find Suv39h1 in mass spec results. This maybe do to that fact that Suv39h1 locates in protein dense region of the polyacrylamide and might have been missed during visual inspection of differentially eluted proteins. We decided to focus on the RNA binding protein FUS/TLS, since this RNA binding protein has multifaceted roles in cell physiology which were described in the Introduction section. We validated mOct4P4/FUS interaction in both anti-flag and anti-FUS RIP experiments in MS2-flag/mOct4P4-24xMS2 stemloop mESCs using MS2-flag mESCs as control (Figure 24B and C). Anti-flag RIP eluates contained FUS only in ES cells expressing mOct4P4 lncRNA (Figure 24B). Accordingly, anti-FUS RIP eluates contained MS2-flag when cells expressed the mOct4P4-24xMS2 stemloop construct (Figure 24C). This data indicates that FUS co-immunoprecipitates with flag only when mOct4P4-24xMS2 stemloop is expressed, confirming that FUS interacts with the mOct4P4 lncRNA.

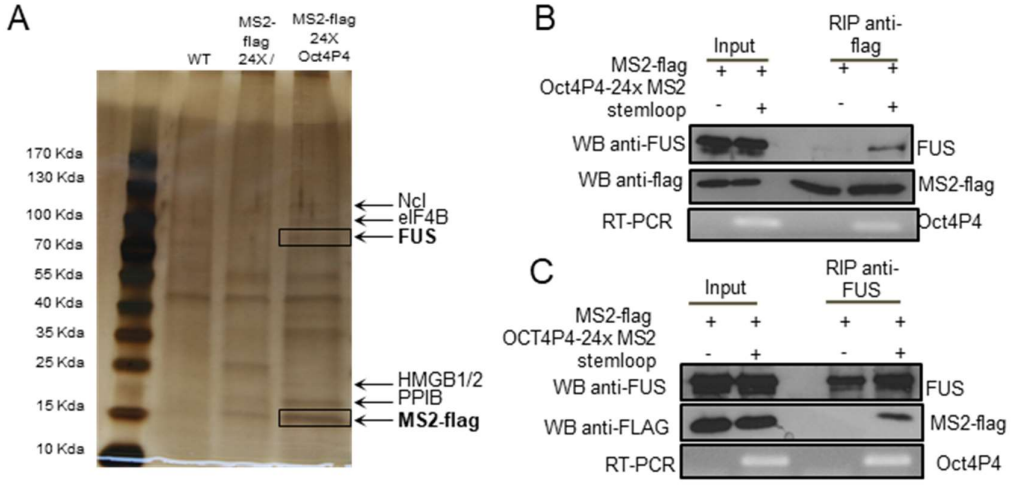

Figure 24: Representative image of gel used for mass spectrometry analysis of mOct4P4 interactors. Wild type, MS2-flag control and MS2-flag/Oct4P4-24xMS2 stemloop mESCs were used to perform anti-flag RNA immunoprecipitation (RIP). The immunoprecipitates were separated by SDS-PAGE and silver staining was used to detect protein on gel. Gel slices, as indicated, were cut out for mass spectrometry analysis. (B) Anti-flag RIP using mESCs overexpressing MS2-flag/mOct4P4-24xMS2 stemloop RNA or MS2-flag control cells. Western blotting for MS2-flag and FUS are shown (top panel). qRT-PCR followed by agarose gel electrophoresis verified the presence of mOct4P4-24xMS2 stem loop RNA in anti-flag RIP experiments (bottom panel). (C) Anti-FUS RIP using MS2-flag/mOct4P4-24xMS stemloop RNA RNA or MS2-flag control cells. MS2-flag co-immunoprecipitates with FUS only in presence of mOct4P4-24xMS2 stem loop RNA, as determined by western blot (top). qRT-PCR followed by agarose gel electrophoresis verified the presence of mOct4P4-24xMS2 stem loop RNA in anti-flag RIP experiments (bottom panel).

To address the direct impact of the mOct4P4 binding partner FUS on Oct4 expression, we carried out FUS loss-of-function experiments in mESC stably overexpressing mOct4P4. Short interfering (siRNA)-mediated knockdown of FUS rescues Oct4 mRNA levels in

mOct4P4 overexpressing cells (Figure 25A, left panel), an effect that is paralleled by a rescue of Oct4 protein levels (Figure 25A, right panel). Accordingly, deposition of H3K9me3 on Oct4 ancestral gene promoter in mOctP4 overexpressing cells is impaired in absence of FUS (Figure 25B). These data indicate that FUS RNA binding protein is required for mOct4P4-mediated deposition of repressive histone marks on ancestral Oct4 promoter.

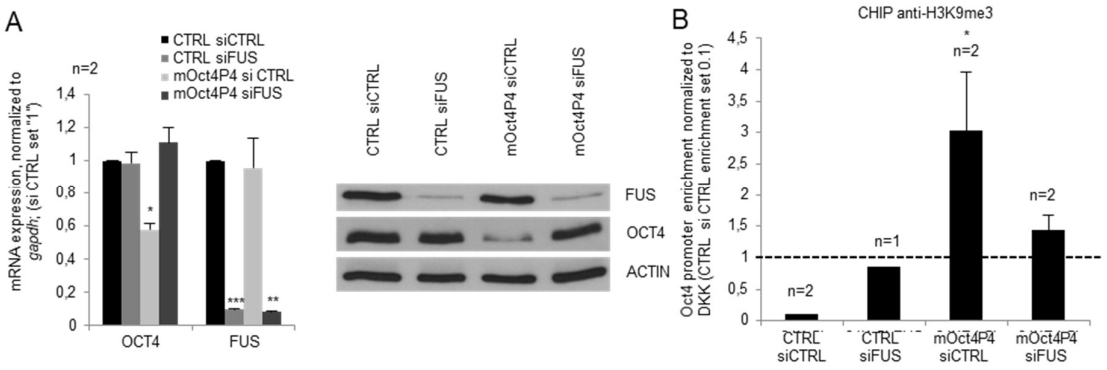

*Figure 25: (A) FUS knockdown in mOct4P4 overexpressing mESCs rescues Oct4 expression, as determined by qRT-PCR (left) and western blotting (right). Actin was used as loading control. (B) ChIP analysis of Oct4 promoter region using an H3K9me3 antibody in mOct4P4 overexpressing mESCs transfected with the indicated siRNAs. FUS silencing reduces Oct4 promoter H3K9me3 in mOct4P4 expressing mESCs. ChIP data were quantified versus input and the unrelated DKK promoter. Error bars indicate s.d.;*p<0.05; **p<0.01; ***p<0.001; n: number of independent experiments*

We next wished to get insights about the mechanism through which FUS may regulate mOct4P4 function. To address this aim, we performed anti-flag ChIP experiment using mOct4P4-24xMS2 stemloop/MS2-flag expressing mESCs transfected with a siRNA specific for FUS. In FUS knockdown condition, we found a reduction in mOct4P4-24xMS2 stemloop abundance at the Oct4 ancestral gene promoter, when compared to mOct4P4-24xMS2 cells transfected with a control siRNA (Figure 26A). Interestingly, an anti-SUV39H1 ChIP in mOct4P4 overexpressing FUS knockdown mESCs reduced SUV39H1 enrichment at the Oct4 promoter (Figure 26B). Together, these findings suggest that FUS silencing lead to an impaired mOct4P4/SUV39H1 complex recruitment to the Oct4 ancestral gene promoter.

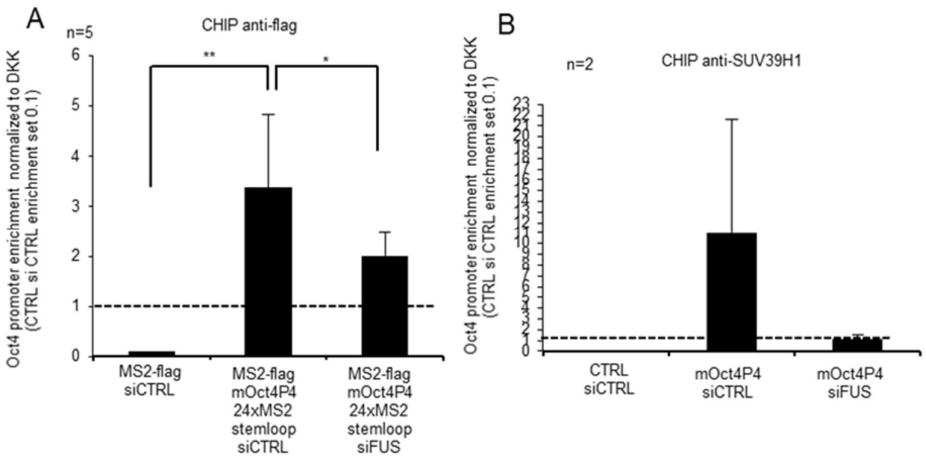

*Figure 26: (A) ChIP analysis of Oct4 promoter region in mOct4P4 overexpressing cell lines transfected with the indicateed siRNAs using anti-flag antibody. FUS silencing reduces mOct4P4 presence at Oct4 promoter. ChIP data were quantified versus input and unrelated DKK promoter. (B) Anti-SUV39H1 ChIP experiment revealed a decreased SUV39H1 binding to Oct4 promoter upon FUS knockdown. Error bars indicate s.d.;*p<0.05; **p<0.01; n: number of independent experiments carried out on the same cell line.*

3.7 FUS role in mOct4P4 minimal functional region

We next aimed to test whether FUS is functionally linked to the mOct4P4 minimal functional region. Surprisingly, FUS silencing (Figure 27A, top; Figure 27B, left panel) in 200bp-mOct4P4 overexpressing undifferentiated mESCs failed in rescuing Oct4 protein (Figure 27A, bottom) and RNA expression (Figure 27B, right panel). This suggests that the mOct4P4 minimal functional region does not require FUS for its silencing activity on the Oct4 ancestral gene promoter.

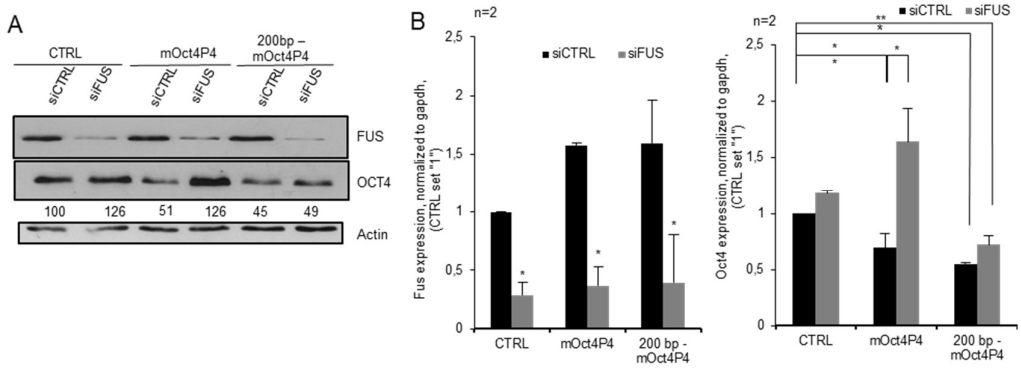

*Figure 27: FUS and Oct4 expression in mOct4P4 and 200bp-mOct4P4-24xMS2 stemloop mESCs under FUS knockdown condition. siRNA mediated FUS knockdown rescues Oct4 expression only in mOct4P4-24xMS2 stemloop mESCs, as determined by western blotting (A) and qRT-PCR (B). Actin was used as loading control in western blot and gapdh dor qRT-PCR. Numbers represent Oct4/Actin ratio (control was set to 100). Error bars represent s.d.; *p<0.05; **p<0.01; n: number of independent experiments*

To further confirm these results, we next wanted to assess the chromatin state of Oct4 promoter upon FUS knockdown in mOct4P4 and 200bp-mOct4P4 overexpressing mESCs. Accordingly to the previous findings, anti-H3K9me3 ChIP experiments on 200bp-mOct4P4 cells revealed that the H3K9me3 content at the Oct4 promoter remained unaltered after FUS knockdown. However, FUS depletion in full-length mOctP4 overexpressing cells is paralleled by a decreased H3K9me3 at the Oct4 promoter (Figure 28A). To highlight the role of FUS in mOct4P4-mediated epigenetic silencing of Oct4, we investigated mOct4P4 and 200bp-mOct4P4 binding on Oct4 promoter under FUS knockdown condition. Anti-flag ChIP experiments on Oct4 promoter exploiting the MS2-flag tethering system showed that 200bp-mOct4P4-24xMS2 stemloop lncRNA is still able to bind Oct4 promoter even after FUS knockdown (Figure 28B). Altogether, these findings indicate that FUS is critically involved in Oct4 promoter silencing by mOct4P4, On the contrary mOct4P4 minimal region function is independent of FUS.

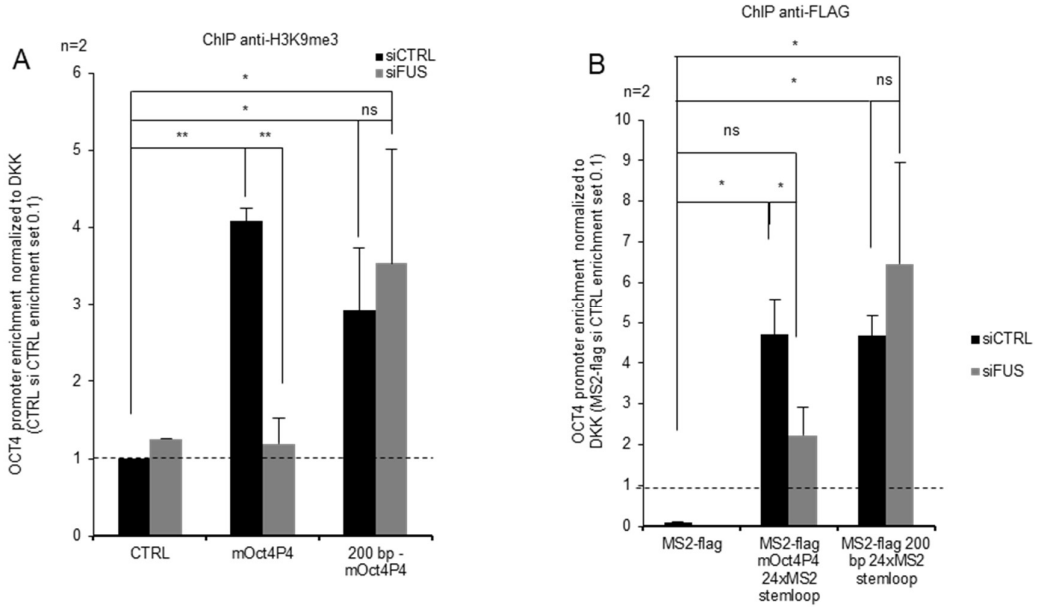

*Figure 28: 200bp minimal region is independent of FUS presence (A) H3K9me3 ChIP analysis of Oct4 promoter region in indicated cell lines. FUS knockdown rescues H3K9me3 abundance on Oct4 promoter only in mOct4P4 overexpressing cells. ChIP data were quantified versus input and unrelated DKK promoter. (B) ChIP analysis of Oct4 promoter region in mOct4P4 overexpressing cell lines transfected with the indicateed siRNAs using anti-flag antibody. FUS silencing redeuces mOct4P4 lncRNA prensence at Oct4 promoter. ChIP data were quantified versus input and unrelated DKK promoter. Error bars indicate s.d.;*p<0.05; **p<0.01; ns: not significant; n: number of independent experiments*

To better clarify the role of FUS in mOct4P4 function we wanted to map its binding site on the mOct4P4 lncRNA by performing an anti-FUS RIP using full length mOct4P4, 200bp-mOct4P4 and -200bp-mOct4P4 overexpressing mESCs. Interestingly, we were able to

efficiently amplify mOct4P4, 200bp-mOct4P and -200bp-mOct4P4 RNA by RT-PCR in anti-FUS RIP eluates, indicating that FUS binds to all these RNAs. Of notice, we did not amplify any RNA coming from these constructs in anti-tubulin control RIP, confirming the specificity of our experimental conditions (Figure 29). These data suggest that FUS binds mOct4P4 lncRNA at multiple regions, both inside and outside minimal functional region. This hypothesis was supported by the fact that mOct4P4 harbors 17 GGUG putative RNA binding motifs for FUS along its sequence, of which two reside in the minimal functional region. Future experiments will be aimed to insert mutations into the GGUG sequences in the minimal region to test their importance in mOct4P4 function.

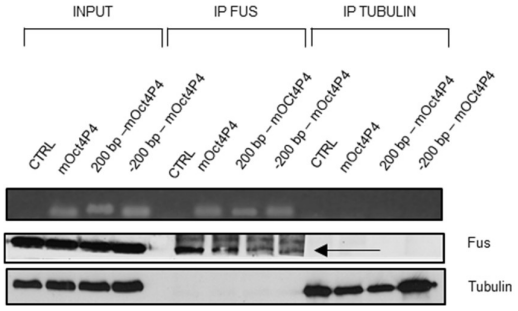

Figure 29: Anti-FUS RIP in experimental cells. qRT-PCR and agarose gel electrophoresis confirmed the FUS-mOct4P4 binding in anti-FUS RIP (top panels). Western blotting confirms the immunoprecipitation of FUS (bottom panel). Anti-tubulin RIP was used as negative control

This suggests that mOct4P4 may require FUS binding to trigger SUV39H1 recruitment and silencing at the Oct4 promoter. We therefore speculated that FUS may induce a change in mOct4P4 conformation, exposing mOct4P4 minimal region to recruit SUV39H1 and rendering it functionally active.

3.8 FUS role conservation in hOCT4P3 function

To find additional evidence for the importance of FUS/mOct4P4 interaction, we addressed its functional conservation in human cells. Taking advantage of MS2-flag system established in OVCAR-3 cell line, we wished to confirm hOCT4P3 binding to FUS. To address this aim, we performed anti-flag RIP experiments in MS2-flag/hOCT4P3-24xMS2 stemloop expressing OVCAR-3 cells. We detected FUS in anti-flag RIP eluates only in presence of hOCT4P3-24xMS2 stemloop RNA (Figure 30A) confirming a conservation of hOCT4P3/FUS interaction in humans. Subsequently, we wanted to determine the conservation of the role of FUS for hOCT4P3 silencing function. Interestingly, transiently

siRNA-mediated knockdown of FUS in OVCAR-3 cells stably overexpressing hOCT4P3, showed a rescue in OCT4 levels in both mRNA and protein (Figure 30B). Altogether, these data support a conservation of the mOct4P4/SUV39H1/FUS complex for OCT4 ancestral gene silencing also in human cancer context.

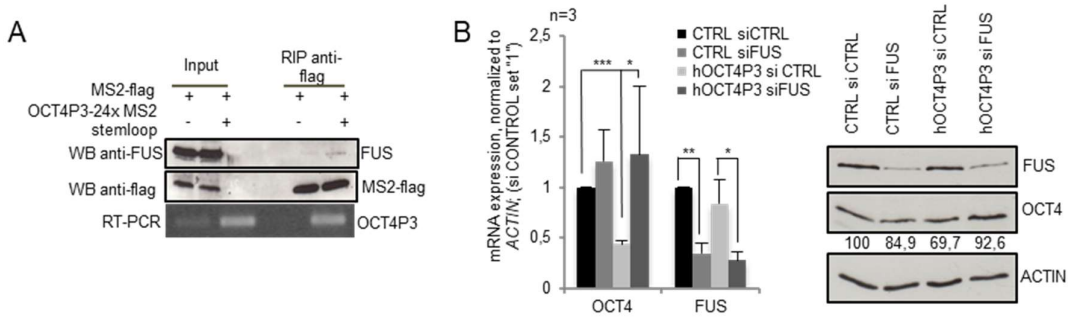

Figure 30: (A) Anti-flag RIP using MS2-flag/hOCT4P3-MS2 stemloop OVCAR-3 cell model system. Western blot for MS2-flag and FUS are shown. qRT-PCR followed by agarose gel electrophoresis verified the presence of OCT4P3-24xMS2 RNA in anti-flag RIP. FUS has been immunoprecipitated only in presence of hOCT4P3 indicating its binding (B) FUS knockdown in hOCT4P3 overexpressing OVCAR-3 cells rescues OCT4 expression, as determined by qRT-PCR (left panel) and western blot (right panel). n: number of independent experiments

3.9 The role of the FUS/SUV39H1/mOct4P4 interaction

In order to further elucidate the interaction between FUS, SUV39H1 and mOct4P4 driving Oct4 ancestral gene silencing we performed an additional series of experiments. We performed anti-FUS ChIP experiments in undifferentiated control mESCs, which express undetectable levels of mOct4P4, and in mOct4P4 overexpressing mESCs. We found that FUS is bound to Oct4 promoter only in presence of mOct4P4 (Figure 31A), suggesting that FUS requires mOct4P4 for being transported to the Oct4 promoter. In addition, to test the requirement of SUV39H1 for FUS recruitment of OCT4 promoter, we performed ChIP experiments in mOct4P4 overexpressing SUV39H1 knockdown mESCs. We previously demonstrated that mOct4P4 does not bind Oct4 promoter in this context (Scarola et al., 2015). Anti-FUS ChIP confirmed that FUS does not bind Oct4 ancestral gene promoter in absence of SUV39H1 (Figure 31B) indicating that SUV39H1 and the mOct4P4 lncRNA presence at the Oct4 promoter are required for FUS recruitment. We demonstrated that FUS binds mOct4P4 and that mOct4P4 binds SUV39H1, thus we were interested in testing whether FUS and SUV39H1 directly bind to each other. Interestingly, we were not able to co-immunoprecipitate FUS in an anti-SUV39H1 IP experiment, neither in presence nor in absence of mOct4P4 (Figure 31C). Accordingly, we were not able to find SUV39H1 in

eluates from anti-FUS immunoprecipitates in both mOct4P4 overexpressing and control mESCs (Figure 31D). This data demonstrate that FUS does not directly bind SUV39H1.

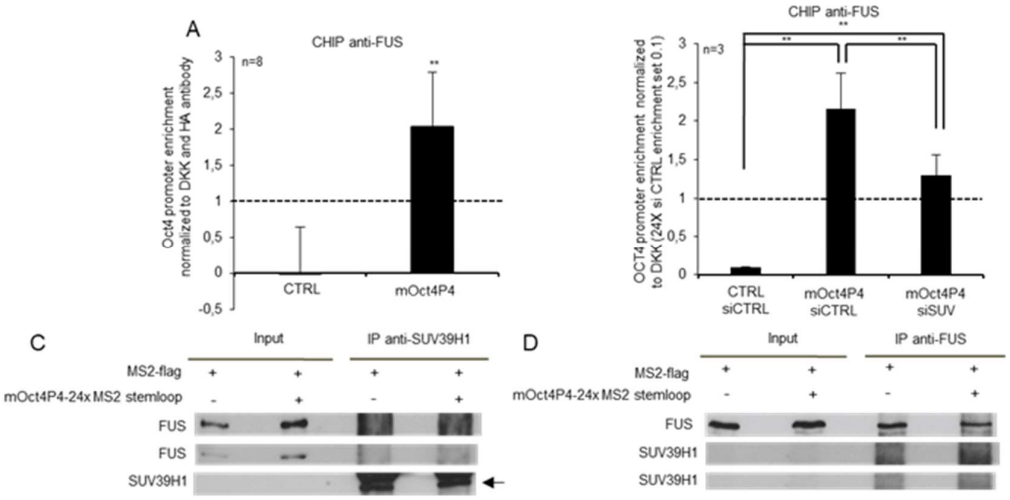

*Figure 31: (A) ChIP analysis of Oct4 promoter region using an anti-FUS antibody in mOct4P4 and control cells. FUS binds Oct4 promoter only in presence of mOct4P4. ChIP data were quantified versus input and the unrelated DKK promoter, and normalized to unrelated HA-specific antibody. (B) Anti-FUS ChIP analisys of Oct4 promoter region under SUV39H1 knock-down conditions. SUV39H1 knockdown reduces FUS abundance at the Oct4 promoer in mOct4P4 expressing mESCs. ChIP data were quantified versus input and the unrelated DKK promoter. (C) Anti SUV39H1 RIP using mOct4P4 expressing cells. Western blotting confirmed SUV39H1 immunoprecipitation but not a FUS/SUV39H1 interaction neither in presece nor in absence of mOct4P4 (D) Anti FUS RIP using mOct4P4 expressing cells. Western blotting confirmed FUS immunoprecipitation but not a FUS/SUV39H1 interaction neither in presece nor in absence of mOct4P4. Error bars represent s.d.; *p<0.05; **p<0.01. n: number of independent experiments carried out in the same cell line*

With this set of experiments we identified the novel mOct4P4 interacting partner FUS that is required for mOct4P4 action. We speculate that FUS binding on mOct4P4 gives rise to a change in its conformation causing the exposure of mOct4P4 minimal functional region. This renders mOct4P4 functionally active and able to bind SUV39H1 and Oct4 promoter and, in turn, silence Oct4 gene expression.

3.10 Final model

In this study we showed novel insights into the molecular mechanism of mOCt4P4 pseudogene lncRNA mediated silencing of the ancestral Oct4 gene. Using a loss-of-function model, we demonstrated that the nuclear restricted mOct4P4 lncRNA function is essential for an effective mESC differentiation since it is able to regulate Oct4 ancestral gene. Performing a mOct4P4 deletion analysis, we identified a mOct4P4 minimal region important for its function. Mass spectrometry analysis identified FUS RNA binding protein as a novel mOct4P4 interactor which was shown to be essential for mOct4P4 function along with SUV39H1 HMTase. Our data suggest that the 5′ and 3′ UTR regions of mOct4P4 localize mOct4P4 lncRNA to the nucleus. There, FUS binds mOct4P4 lncRNA (Figure 32A) and this interaction could, via a still unknown mechanism, change the conformation of mOct4P4 lncRNA (Figure 32B). This may expose mOct4P4 minimal functional region (base pairs 984-1183 in mOct4P4 lncRNA), for binding by SUV39H1 (Figure 32C). In fact, FUS depletion caused a loss of full-length mOct4P4 function while it did not alter mOct4P4 minimal region mechanism of action. In a subsequent step, a mOct4P4 lncRNA/SUV39H1 complex may then translocate to the Oct4 promoter to mediate H3K9me3 deposition leading to Oct4 ancestral gene silencing (Figure 32D). Our data also show that this mechanism is recapitulated by the human homolog of mOct4P4, hOCT4P3, demonstrating the importance of Oct4 pseudogenes in precisely controlling Oct4 expression.

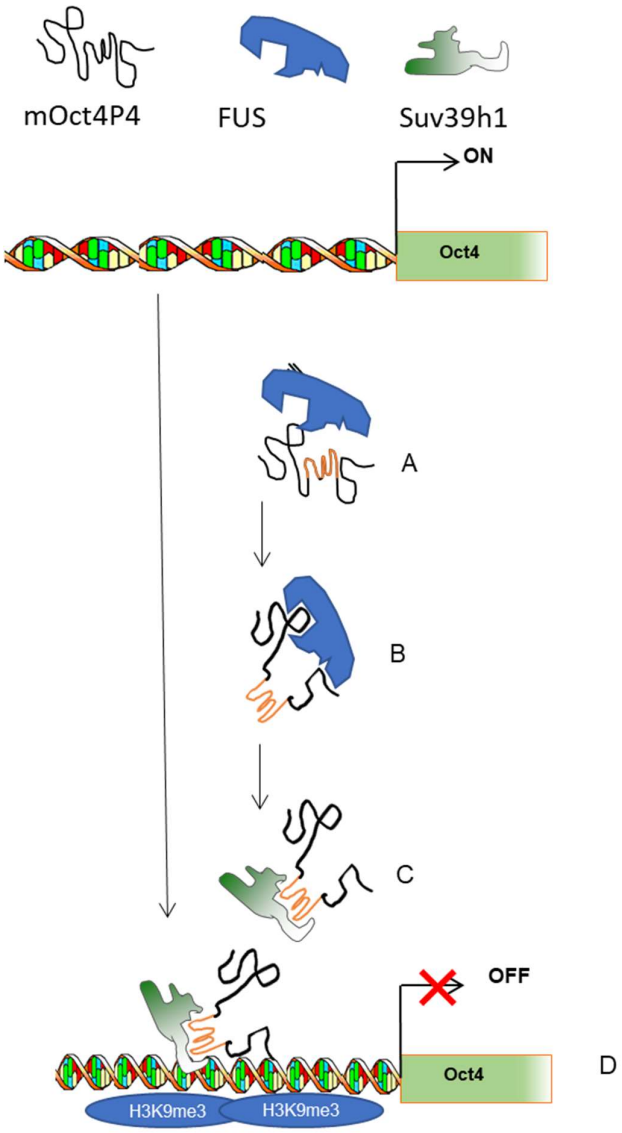

Figure 32: The nuclear restricted mOct4P4 lncRNA is bound by FUS (A); this binding causes a change in the mOct4P4 conformation which exposes mOct4P4 200bp minimal functional region (in orange) (B); this region is then bound to Suv39h1 (C). The complex then translocates to the ancestral Oct4 promoter where it leads to the imposition of H3K9me3 and gene silencing.

4. Conclusions and discussion

Canonical lncRNAs and pseudogenes-derived lncRNAs have emerged as important regulators in cell physio-pathology. The most studied mechanism of canonical lncRNA gene regulation is their role as epigenetic *cis* and *trans*-acting modulators for the expression of protein-coding genes by recruiting chromatin-remodeling complex to a specific chromatin locus (Han and Chang, 2015; Tsai et al., 2010). For example, the lncRNA Xist provides an interesting example of chromatin remodeling lncRNA. The Xist RNA B-repeat is bound by hnRNPK which, in turn, recruit Polycomb proteins through PCGF3/5-PRC1 complex (Almeida et al., 2017; Pintacuda et al., 2017). PRC2 is subsequently activated and recruited in a Jarid2 dependent manner causing the H3K27 trimethylation and gene inactivation along the X chromosome (da Rocha et al., 2014). Pseudogene derived lncRNA, by contrast, are mostly studied for acting as competitive endogenous RNAs or altering the stability of ancestral mRNA (Johnsson et al., 2013; Poliseno et al., 2010; Tam et al., 2008; Wang et al., 2010a; Watanabe et al., 2008). Still, there is evidence for alternative mechanism which involve anti-sense pseudogene transcripts that impact on the promoter activity of ancestral genes (Hawkins and Morris, 2010; Johnsson et al., 2013, 2014; Korneev et al., 1999; Pain et al., 2005). The self-renewal Oct4 transcription factor offers a paradigmatic example of pseudogene-derived lncRNA mediated gene regulation with different mechanisms. We demonstrated that the X-linked sense oriented and nuclear restricted mOct4P4 lncRNA is upregulated upon differentiation; it forms a complex with the SUV39H1 HMTase, translocates to the promoter of the ancestral Oct4 gene on chromosome 17, leading to gene silencing of the ancestral Oct4 gene *in trans* (Scarola et al., 2015). This provides a proof-of-principle for a novel pseudogene-derived sense lncRNA in modulating of ancestral genes on the epigenetic level.

Based on data previously obtained in our laboratory, in this study we wanted to gain further insights into mOct4P4-mediated Oct4 ancestral gene silencing. Interfering with mOct4P4 expression impaired long-term embryoid body differentiation. This is reflected by persisting expression of self-renewal markers and reduced upregulation of early differentiation factors and other hallmarks of effective differentiation. This indicates that mOct4P4 lncRNA expression is essential for effective mESC differentiation regulating Oct4

expression. In fact, this was supported by the fact that interfering with the mOct4P4/SUV39H1 complex in differentiated cells led to re-activation of self-renewal transcription factors expression representing a partial re-acquisition of mESC stem cell features in mOct4P4 lncRNA knock-down experiments in terminally differentiated cells (Scarola et al., 2015). This hints a possible role for mOct4P4 silencing in cell reprograming. Yamanaka and Takahashi introduced the concept of induced pluripotent cell (iPSC) generation through the expression of four self-renewal transcription factors (Oct3/4, Sox2, Klf4 and c-Myc) in terminally differentiated cells (Takahashi et al., 2006). However, the features of iPSCs were not identical to ES cells and further works tried to improve iPSCs generation methodologies without fully resolving the intrinsic inefficiency of these protocols (Okita et al., 2007, 2008; Wernig et al., 2007). Interestingly, some groups pointed out that c-Myc is dispensable for iPSC generation (Nakagawa et al., 2008, 2010; Wernig et al., 2008) indicating that some factors could be replaced or eliminated. It will be interesting to test whether loss of mOct4P4 function can increase the efficiency of induced pluripotent stem cell generation protocols in addition to the classical Yamanaka factors.

As mentioned above, lncRNAs can be incorporated into chromatin-modifying complexes and function as active components or as a scaffold to assemble the complex for chromatin modification. In the fruit fly, for example, the lncRNA roX2, after being structurally remodeled by an RNA helicase MLE1 (Ilik et al., 2013; Maenner et al., 2013) becomes incorporated into the male-specific lethal (MSL) protein complex, which then recruits a histone acetyltransferase (MOF) to induce histone acetylation and activate gene transcription in the male X chromosome (Akhtar and Becker, 2000; Sass et al., 2003). Interestingly, roX2 lncRNA contains a minimal functional unit, structured as two mutually exclusive stem-loops that exist in a peculiar structural arrangement essential for its function (Ilik et al., 2017). In order to identify whether mOct4P4 harbors a minimum functional region, we performed a deletion study analysis that revealed a region comprising position 984-1183 of the mOct4P4 lncRNA essential for SUV39H1 recruitment and imposition of H3K9me3 at the Oct4 promoter. Importantly, expression of this portion of the mOct4P4 lncRNA is sufficient to repress Oct4. In addition, we identified a novel mOct4P4/SUV39H1 complex interactor, FUS/TLS. This was interesting because FUS was shown to be involved

in several mRNA regulation pathways but little is known about its role in pseudogene-derived lncRNA function (Ratti and Buratti, 2016) and embryonic stem cells (Ratti and Buratti, 2016). In *X. laevis*, fus/tls was linked to mRNA processing of several developmental regulators during gastrulation (Dichmann and Harland, 2012). Inbred Fus/Tls−/− mice have been reported to be small at birth and die within a few hours with major defects in B-lymphocyte development (Hicks et al., 2000; Lagier-Tourenne et al., 2010), whereas similar mice in an outbred background survive until adulthood but develop male sterility (Kuroda, 2000; Lagier-Tourenne et al., 2010). Interestingly, FUS depletion impaired mOct4P4-mediated Oct4 silencing but not the function of its minimal functional region alone which is still able to bind Oct4 promoter and impose H3K9me3 repressive histone marks. By contrast SUV39H1 seems to play an essential role for the function of both full length and minimal region, since loss of SUV39H1 impaired the binding of the mOct4P4 lncRNA to the Oct4 promoter and mOct4P4 lncRNA-mediated Oct4 silencing. We hypothesize that the mOct4P4 lncRNA binds FUS resulting in a change of mOct4P4 lncRNA conformation. In this way mOct4P4 may expose the minimal functional region to be then recognized by the SUV39H1 HMTase. This could be explained by the fact that mOct4P4 minimal region does not require FUS-mediated change of conformation since it lies in a "always-exposed" state always able to be recognized by SUV39H1. Subsequently, the mOct4P4 lncRNA/SUV39H1/FUS complex may translocate to the Oct4 promoter where it imposes H3K9me3 repressive chromatin modification and, in turn, silence Oct4 expression. Further experiments will be required to validate the conformational change of mOct4P4 induced by FUS. For example, we propose to either delete or mutate GGUG FUS binding sites in mOct4P4 lncRNA and test their ability to repress Oct4 expression. In addition, Chen et al. demonstrated that the human telomerase RNA hTR changes its conformation upon binding with TCAB1 using a technique called icSHAPE (Chen et al., 2018; Flynn et al., 2016). The use of this technique will allow to demonstrate the exposure of mOct4P4 minimal region upon FUS binding. Finally, we speculate that deleting the 200 nucleotide functional sequence from the endogenous mOct4P4 gene locus may result in a loss of Oct4 silencing in mES cell differentiation.

Our data provide important insights into the epigenetic mechanism that the mOct4P4 lncRNA uses to control ancestral Oct4 expression *in trans* to regulate self-renewal and differentiation. In addition, it would be exciting to find out the developmental phenotype of the knockout of mOct4P4 *in vivo*. Since we demonstrated that mOct4P4 loss in mESCs impaired EB differentiation we speculate that its loss *in vivo* could be embryonically lethal or lead to developmental defects. This last hypothesis could be supported by the fact that a BLAST analysis revealed conservation of mOct4P4 in various placental mammals (for example deer mouse, walrus and bonobo).

To our knowledge, mOct4P4 lncRNA represent the first example of a sense-oriented pseudogene-derived lncRNA able to direct chromatin-modifying enzymes to the promoter of its ancestral gene *in trans*. However, the exact mechanism through which mOct4P4 lncRNA targets the ancestral Oct4 gene promoter *in trans* remains to be unraveled. A first working hypothesis is based on the fact that the minimum functional region of mOct4P4 harbors two stretches of 11 nucleotides that are complementary with ancestral Oct4 promoter (considering a region of 2000bp upstream of the TSS). This may enable that the mOct4P4/SUV39H1 complex could be targeted to the Oct4 promoter by sequence complementarity, eventually by the formation of RNA:DNA hybrids. Deletion of these stretches will be fundamental to test this hypothesis. We also speculate that the mOct4P4 minimal region could intercalate into the DNA of Oct4 promoter promoted by the structural conformation of mOct4P4 (and its minimum functional region). Another hypothesis is based on the observation of a transient pairing of Oct4 alleles at the onset of the early embryonic stem cell differentiation that precedes Oct4 epigenetic silencing (Hogan et al., 2015). It would be interesting to find out if this allele pairing may also involve the mOct4P4 gene locus, enabling the neosynthesized Oct4P4 lncRNA to recruit SUV39H1 to the Oct4 promoter. We also need to consider that mOct4P4 may be recruited to the Oct4 promoter by SUV39H1, since we observed that loss of SUV39H1 leads to reduced mOct4P4 lncRNA at the Oct4 promoter. Finally, mOct4P4/SUV39H1 recruitment on Oct4 promoter could be mediated by an unknown additional interactor which could be both a protein or a RNA.

Interestingly, we found that the mOct4P4 lncRNA gene silencing activity could evolutionarily conserved. In fact, we showed that human OCT4 pseudogene-derived

hOCT4P3 lncRNA shares high homology with mOct4P4 in terms of sequence, subcellular localization and mechanism of function. Several pseudogenes have been to linked cancer (Bier et al., 2009; Han et al., 2011; Hayashi et al., 2015; Hu et al., 2018; Ioffe et al., 2012; Poliseno, 2012; Puget et al., 2002), some of which have been recently demonstrated to act via controlling epigenetic gene regulation (Lister et al., 2017; Wei et al., 2017). Interestingly, OCT4 pseudogenes could play different roles at various stages of cancer progression (Hayashi et al., 2015; Kastler et al., 2010; Poliseno, 2012; Poursani et al., 2016; Suo et al., 2005; Wang et al., 2013a). Preliminary data obtained in our laboratory indicate that hOCT4P3 is able to reduce cancer cell proliferation by controlling OCT4 expression. Nevertheless, it could be interesting to assess whether the effects of mOct4P4/hOCT4P3 are ascribable exclusively to Oct4 or whether the Oct4 pseudogene lncRNA/SUV39H1/FUS complex is able to regulate other candidate genes *in trans*. In addition, we speculate that mOct4P4/hOCT4P3 lncRNA could represent a prototype for a category of pseudogenes that are able to impose epigenetic modifications on their ancestral gene. Given the vast repertoire of vertebrate pseudogenes, most of which have no or still unknown function, we anticipate the existence of a series of pseudogene-derived lncRNAs that may use an analogous mechanisms to control the expression of ancestral genes on the epigenetic level. Identifying sense nuclear restricted pseudogene lncRNA whose expression is inversely correlated with their ancestral gene may open the window on a new functional class of pseudogene-derived lncRNAs. Notably, with this work we anticipate an unprecedented role for the RNA binding protein FUS in changing a lncRNA conformation and in pseudogene-derived lncRNA biology which could be conserved through other canonical and pseudogene-derived lncRNA species.

Finally, lncRNAs have been studied for their ability to regulate many cellular processes via various mechanisms, however feasibility for the biotechnological application of lncRNAs is still in its infancy. Nonetheless, few examples for lncRNAs adapted for a biotechnological application do exist. SINEUPs are modular antisense lncRNAs able to increase synthesis of target proteins in cells. They act through two functional domains - a SINEB2 inverted sequence and an antisense overlapping region to target gene sense mRNA – and they are able to upregulate translation in a gene-specific manner (Indrieri et al., 2016; Patrucco et al.,

2015; Sasso et al., 2018; Schein et al., 2016; Takahashi et al., 2018; Zucchelli et al., 2015b, 2015a, 2016). Similarly, our new insights into mOct4P4 lncRNA mechanism of action may be used to generate a lncRNA/SUV39H1 based tool to target repressive epigenetic modifications at gene promoters of interest. To address this aim, future experiments will be designed in order to better define structural mOct4P4 lncRNA motifs, present in the minimal functional region. In particular, it will be crucial to understand the exact mechanism through which mOct4P4 lncRNA/SUV39H1 complex recognizes Oct4 promoter and translate this insights to other genes of interest.

5. Materials and methods

5.1 Cell culture and generation of Embryoid Bodies (EBs)

mouse Embryonic Stem cells (mESCs) were cultured in a feeder-independent condition on 0.2% gelatin-coated plates in mESC self-renewal medium (Dulbecco's modified Eagle's medium) supplemented with 15% knockout serum replacement (Lonza), 1% non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 0.1 mM β-mercaptoethanol, 1% penicillin/streptomycin (Lonza) and 1,000 U/ml mouse leukemia inhibitory factor. OVCAR-3 cells were cultured in RPMI-1640 medium (BioWhittaker, Lonza) supplemented with 20% (v/v) Fetal Bovine Serum (FBS), insulin (10 μg/ml; I9278,Sigma) and 1% (v/v) penicillin/streptomycin (Lonza). Cell lines were maintained as monolayers at 37(°C in a humidified 5% CO2 atmosphere.

mESCs differentiation was obtained with a Dulbecco's Eagle's medium (DMEM) supplemented with 15% ES cell certified serum (Invitrogen), 1% non-essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), 1% L-glutamine (Invitrogen), 0.1 mM β-mercaptoethanol and 1% penicillin/streptomycin (Invitrogen). Embryoid bodies (EBs) were generated using 300 cells in hanging drops culture for 3 days. Subsequently, EBs were transferred to a low-attachment 24-well plates (Euroclone) and grown in suspension for the indicated days. For beating embryoid bodies after hanging drop culture cells were plated in a Ø10cm dish after 3 days and maintained for the indicated days.

5.2 Generation of plasmid

To generate the pLX-sgOCT4P3 and pLx-sgOct4P4 we specifically amplified pLX-sgRNA plasmid (Addgene Plasmid #50662) using two PCR amplifications:

pLX sgRNA cloning	Forward/reverse cloning primers
PCR 1	F: AACTCGAGTGTACAAAAAAGCAGGCCTTTAAAG
	R: rc(GN_{19})GGTGTTTCGTCCTTTCC
PCR 2	F: GN_{19}GTTTTAGAGCTAGAAAATAGCAA
	R: AAAGCTAGCTAATGCCAACTTTGTACAAAGAAAGCTG

Where GN_{19}=Target sequence and rc(GN_{19})= reverse complement of the target sequences which were the following: sgOct4P4 GN_{19}=GAAGTTGGGCACCCAAGTTGG; sgOCT4P3 GN_{19}=GCCAGGCTGGTCTTGAACTTCTGG

PCR1 and PCR2 amplicons were fused using a third PCR using Forward primer from PCR1 and Reverse primer from PCR2. PCR products were then cloned into pLX vector using NheI and XhoI restriction sites

To generate the hOCT4P3-24xMS2 stem loop construct, 24 repeats of the MS2 stem loop RNA motif (obtained from pSL-MS2-24X) were cloned downstream of the OCT4P3 cDNA into pLPC. hOCT4P3 cDNA was PCR amplified using the following oligonucleotides: forward: GGGAATTC**AAGCTT**CAGCAAAGAACTAGAAGATAT; reverse: GGGAATTC **AGATCT**TTGTCTATCTACTGTGTTGA and cloned into a pLPC-24xMS2 stemloop vector using HindIII and BglII restriction sites.

To generate miRVec-hOCT4P3 vector hOCT4P3 was PCR amplified using the following nucleotides: forward: GGGAATTC **GGATCC** CAGCAAAGAACTAGAAGATAT; reverse: GGGAATTC**GAATTC**TTGTCTATCTACTGTGTTGA and cloned into miRVec using BamHI and EcoRI restriction sites.

pLPC-Oct4P4-24x stem loop deletion constructs were obtained deleting progressively 200bp starting from 3'UTR of the gene (hereafter indicated as Δ200, Δ400, Δ600, Δ800, Δ994, respectively). pLPC-Oct4P4-24x deletion constructs were generated by amplifying a specific portion of pLPC-Oct4P4-24x construct by PCR (forward and reverse primers are reported in the table below) and fused with the 3'UTR region of the gene using XhoI restriction enzyme. The obtained constructs were next cloned into pLPC-24xMS2 stemloop vector via HindIII and BglII restriction sites.

pLPC-Oct4P4-24x deletion constructs	**Restriction sites and Forward/reverse cloning primers**
Δ200	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R XhoI: GATTC **CTCGAG** AATGTGTACAGTGTGGTG
Δ400	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R XhoI: GATTC **CTCGAG** CCCAAGCTGTTTGATGAT
Δ600	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R XhoI: GATTC **CTCGAG** AGTTTACACAAGCTCTTG
Δ800	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R XhoI: GATTC **CTCGAG** GGGACTTGGTTCCAGCTT

Δ994	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R XhoI: GATTC **CTCGAG** TTATGGTAGGTGATGGCT
3'UTR	F XhoI: GATTC **CTCGAG** TAGAGAGGAAGATGAAGT
	R BglII: GATTC **AGATCT** TGTGTCCCAGGCTTTTTA

For the generation of pLPC-200bp-Oct4P4-24x construct were amplified the 5'UTR, a region corresponding to 200bp (minimal functional region) and 3'UTR of pLPC-Oct4P4-24x by PCR. Subsequently, the amplification products were next fused using the restriction sites indicated in the table below. The obtained construct was cloned into pLPC-24xMS2 stemloop vector via HindIII and BglII restriction sites.

pLPC-200bp-Oct4P4-24x construct	**Restriction sites and Forward/reverse cloning primers**
5'UTR	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R KpnI: GATTC **GGTACC** GGGGAAGTTGGGCACCCC
200bp	F KpnI: GATTC **GGTACC** GCGGCCCCTGCTGGAGAA
	R XhoI: GATTC **CTCGAG** CCCAAGCTGTTTGATGAT
3'UTR	F XhoI: GATTC **CTCGAG** TAGAGAGGAAGATGAAGT
	R BglII: GATTC **AGATCT** TGTGTCCCAGGCTTTTTA

pLPC-(-200bp)-Oct4P4-24x construct was generated amplifying the mOct4P4 region upstream and downstream of the 200 bp of pLPC-Oct4P4-24x by PCR. The final fused product was cloned in pLPC-24x vector via HindIII and BglII restriction sites (see table below).

PLPC-(-200bp)-Oct4P4-24x construct	**Restriction sites and Forward/reverse cloning primers**
Upstream	F HindIII: GATTC **AAGCTT** GTCCCTAGGTGACCAACT
	R XhoI: GATTC **CTCGAG** AGTTTACACAAGCTCTTG
Downstream	R BglII: GATTC **AGATCT** TGTGTCCCAGGCTTTTTA
	R BglII: GATTC **AGATCT** TGTGTCCCAGGCTTTTTA

5.3 Viral transduction and generation of stable cell lines

ES and OVCAR-3 viral infection was obtained after retroviral vector packaging in 293GP or lentiviral vector packaging in 293T cells. The transduced retroviral vectors used are:

pLPC-24x, pLPC-Oct4P4-24x, pLPC-Oct4P4-deletion contructs pLPC-200bp-Oct4P4-24x, pLPC-(-200bp)-Oct4P4-24x, pPLC-hOCT4P3-24xMS2 and pMSCV-HA-MS2-Flag. The transduced lentiviral vectors used are: pLX-sgOCT4P3, pLX-sgOct4P4 and pHAGE-EF1α dCas9-HA-KRAB (Addgene plasmid #50919). Stable cell lines infected with pLPC and pHAGE vectors were maintained in culture with 3 μg/mL of Puromycin and pMSCV and pLX vectors with 4 μg/mL of Blasticidine.

5.4 Transient transfection

Transient transfections of plasmids were performed using TransIT®-LT1 transfection reagent (#MIR-2300, Mirus). Transient transfection of siRNA was performed using Lipofectamine RNAiMAX reagent (Invitrogen) according to manufacturer's suggestions. The sense FUS siRNA sequence is the following: GCAACAAAGCUACGGACAA (Eurofins Genomics); the sense SUV39H1 siRNA sequence is the following: CCAAUUACCUGGUGCAGAA (Thermo Scientific Dharmacon). Control siRNA was used as a negative control: Non-Targeting siRNA#1, TAGCGACTAAACACATCAA (Thermo Scientific Dharmacon).

5.5 Mass spectrometry analysis

To perform mass spectrometry analysis of Oct4P4 interactors wild type, MS2-flag control and MS2-flag/Oct4P4-MS2 stem loop mESCs were used to perform anti-flag RNA immunoprecipitation (RIP). In brief, cells were cross-linked in culture medium with 1 % formaldehyde for 10 minutes. After 125 mM glycine (PBS) was added to block formaldehyde action and cells were washed twice with cold PBS. Crosslinked cells were scraped in lysis buffer (50 mM Tris-HCl, pH 7.4, 1% Triton X-100, 1mM EDTA, 150 mM NaCl) supplemented with protease inhibitors and RNaseOUT (Invitrogen). The cell lysates were sonicated, centrifugated and the supernatant were pre-cleared for 1h at 4°C with protein A/G PLUS-Agarose (sc-2003; Santa Cruz Biotechnology). The pre-cleared lysates were incubated overnight at 4°C with Anti-FLAG M2 Magnetic Beads (M8823, Sigma-Aldrich) following manufacturer's suggestion. RNA-protein complexes were collected with magnetic beads and washed 10 times in lysis buffer. To elute the bound FLAG fusion complexes from the anti-FLAG M2 antibody, magnetic beads were incubated for 1h at 4°C with 50 μl TBS solution containing 15μg FLAG Peptide (F3290, Sigma-Aldrich). The

immunocomplexes proteins were resuspended in SDS-PAGE sample buffer and boiled for 15 minutes at 95°C to reverse cross-linking and denature proteins.

The immunoprecipitates were separated by SDS-PAGE and a mass spectrometry-compatible silver staining (SilverQuest™ Kit, Invitrogen) or Coomassie Brilliant blue R-250 (42660, Sigma) was used to detect proteins on gel. Selected lanes were excised from the gels, de-stained and washed twice in 50% ACN with 50 mM ammonium bicarbonate, dehydrated in 100% ACN. In-gel reduction was performed by incubating lanes in a 10 mM DTT, 100 mM ammonium bicarbonate solution for 30 min at 56°C and alkylation in 55 mM iodoacetamide, 100 mM ammonium bicarbonate for 20 min at room temperature. Protein samples were dehydrated again in 100% ACN and digested by rehydrating gel pieces in 50 mM ammonium bicarbonate containing 4 ng/uL of trypsin (#V5111, Promega) overnight at 37°C. Tryptic peptides were desalted and concentrated using ZipTip mC18 pipet tips (Millipore) and co-eluted onto the MALDI target in 1 μL of α-cyano-4-hydroxycinnamic acid matrix (5 mg/mL in 50% ACN,0.1% TFA).

Mass spectra were acquired over a mass range of 800–4000 m/z (Nd:YAG laser at 355 nm, 40 Shots/Sub-Spectrum for 2,000 Total Shots/Spectrum) by reflectron positive mode on an Applied Biosystems 4,800 Proteomics Analyzer mass spectrometer (Applied Biosystems) and calibrated using a standard mixture (Mass Standards Kit, AB SCIEX). MS/MS spectra were acquired in positive mode (Nd:YAG laser at 355 nm, 40 Shots/Sub-Spectrum for 4,000 Total Shots/Spectrum) and MS/MS calibration was achieved by using the default calibration method.

Protein identifications were performed with the ProteinPilotTM software (version 2.0.1; Applied Biosystems) using the ParagonTM algorithm as the search engine. Each MS/MS spectrum was searched against an Uniprot/SwissProt database of mouse. The search parameters allowed for cysteine modification by iodoacetamide and biological modifications programmed in the algorithm (i.e., phosphorylations, semitryptic fragments, etc.). The detected protein threshold (ProtScore) in the software was set to 1.3 to achieve 95% confidence interval.

5.6 RNA immunoprecipitation

Experimental cells were scraped in RIPA buffer (50mM Tris-Cl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium deoxycholate, 0.05% SDS, 1mM EDTA, 150mM NaCl) and supplemented with protease inhibitors (Complete, Roche) and RNaseOUT (Invitrogen). After incubation at 4°C for 20 min, the cell lysate was centrifuged. The supernatant was pre-cleared for 1 h at 4°C with protein A/G PLUS-Agarose (protein A/G PLUS-Agarose – sc-2003; Santa Cruz Biotechnology supplemented with yeast transfer RNA. 0.1 mg/mL). The pre-cleared supernatant was incubated overnight at 4 °C with rabbit polyclonal anti-TLS/FUS (ab23439, abcam), mouse monoclonal anti-KMT1A/SUV39H1 (2.5 mg/ml, ab12405, Abcam) and mouse monoclonal anti-FLAG M2, clone M2 (2.5 mg ml-1; F1804; Sigma). RNA–protein complexes were recovered with protein A/G PLUS-Agarose beads and were washed six times in RIPA buffer. An aliquot of beads containing immunoprecipitated samples were saved for western blotting analysis. Remaining beads were used to obtain immunoprecipitated RNA that was analyzed by qRT-PCR. A mouse monoclonal anti-HA, clone HA-7 (2.5 mg/ml, H9658) was used as negative control for immunoprecipitation.

5.7 Western blotting and antibodies

An ice-cold lysis buffer containing 10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM PMSF, 20 mM β–glycerophosphate (Sigma-Aldrich), 1 mM Na3VO4, 1 mM Na2F, 5 mM EDTA, protease inhibitor cocktail (Sigma) and 1% Triton X-100 (Sigma-Aldrich) was used to prepare OVCAR-3 and mESCs lysates. For the analysis of immunocomplexes we resuspended input and immunoprecipitated proteins in SDS–PAGE sample buffer, boiled for 15 min and analyzed by western blotting. The following primary antibodies were used: rabbit polyclonal anti-Oct4 (ab19857, abcam), rabbit polyclonal anti-actin (A2066, Sigma), mouse monoclonal anti-Flag (clone M2, F1804), mouse monoclonal anti-KMT1A/SUV39H1 (ab12405) and rabbit polyclonal anti-TLS/FUS (ab23439, abcam). Secondary antibodies coupled to horseradish peroxidase were obtained from Sigma (anti-rabbit IgG peroxidase conjugate A-6154; anti-mouse IgG peroxidase conjugate A-4416). ImageJ software (http://imagej.nih.gov/ij/) was used to quantify protein bands.

5.8 RNA extraction and analysis

Total and immunoprecipitated RNA were extracted with the Qiazol lysis reagent (Qiagen), subjected to DNase treatment (Qiagen) and subjected to reverse transcription (Quantitect reverse transcription kit; Qiagen). The obtained cDNA used for quantitative real-time PCR (SYBR Green Universal PCR Master Mix, Applied Biosystems) on a StepOnePlus real-time PCR machine (Applied Biosystems). mRNA levels were normalized against Gapdh or ACTIN.

Nuclear and cytoplasmic RNA fractions were obtained by the following protocol: cells were collected, resuspended in lysis buffer (10mM NaCl, 20mM MgCl, 10mM Tris-Cl, pH 7.8, 5mM DTT, 0.5% NP-40) and kept in ice for 5 minutes. Nuclei were pelleted by centrifugation at 8000 rpm for 5 minutes at 4°C, pellets were washed and resuspended in lysis buffer. The cytoplasmic fraction was collected to a new tube and clarified by centrifugation. Nuclear and cytoplasmic fractions were subjected to protease treatment for 20 minutes at 37°C by adding an equal volume of proteinase K solution (300 mM NaCl, 0.2 M Tris-Cl, pH7.5, 25mM EDTA, 2% SDS and 0.1 mg/ml proteinase K), and RNA was purified using the Qiazol lysis reagent (Qiagen) and processed as indicated above.

qRT-PCR primer sequences are reported below:

Mouse

- 200bp-mOct4P4 F: TGGAGGAAGCGGACAACAAT
- 200bp-mOct4P4 R: TTCCTCTCTACTCGAGCCCAA
- Fgf5 F: GTTTCCAGTGGAGCCCTT
- Fgf5 R: GAGACACAGCAAATATTTCCAAAA
- Fus F: CAGTCCTCGCGGCATCGCTT;
- Fus R: GGCCCCGTAGCTTTGAGTTGCT.
- Gapdh F: TTCACCACCATGGAGAAGGC;
- Gapdh R: CCCTTTTGGCTCCAC;
- Gdf3 F: TCAGCTTCTCCCAGACCAGGGTTT
- Gdf3 R: CACACGCCCCGGTCCTGAAC
- Nanog F: TTCTTGCTTACAAGGGTCTGC

- Nanog R: AGAGGAAGGGCGAGGAGA
- Oct4 F: CAGGGACACCTTTCCCAGGG
- Oct4 R: TTTAAGAACAAAATGATGAG
- Oct4P4 F: TGGCACCTGGCTTTAGACTTT;
- Oct4P4 R: CCAGGCCAACTTAGGGCATT;
- Sox2 F: TGCTGCCTCTTTAAGACTAGGG
- Sox2 R: TCGGGCTCCAAACTTCTCT
- sgOct4P4 F: GTTGGGCACCCCAAGTT
- sgOct4P4 R: TCAAGTTGATAACGGACTAGCCT
- SUV39H1 F: CGGATCACCGTGGAGAAT
- SUV39H1 R: CACTCACAGCCAACAGCTACCT

Human:

- ACTIN F: CCAACCGCGAGAAGATGA
- ACTIN R: CCAGAGGCGTACAGGGATAG
- FUS F: ATGGCCAGAGCCAGAACACAGG
- FUS R: CGAGGTGCTGCTGGGAGCTG
- KLF4 F: CAAGTCCCGCCGCTCCATTACCAA
- KLF4 R: CCACAGCCGTCCCAGTCACAGTGG
- NANOG F: ATGCCTCACACGGAGACTGT
- NANOG R: AGGGCTGTCCTGAATAAGCA
- OCT4A F: GGAGCCCTGCACCGTCA
- OCT4A R: ATGGTCGTTTGGCTGAAT
- OCT4P3 F: CTTCGGATTTCGCCTTCTCA
- OCT4P3 R: GGGCACTAGCCCCACTCCAGT
- sgOCT4P3 F: GCTGGTCTTGAACTTCTGGGT
- sgOCT4P3 R: TCAAGTTGATAACGGACTAGCCT
- SOX2 F: CCCACCTACAGCATGTCCTACTC
- SOX2 R: TGGAGTGGGAGGAAGAGGTAAC

5.9 Chromatin immunoprecipitation

Cells were cross-linked in culture medium with 1% formaldehyde for 15 min, neutralized using 125mM glycine in PBS and washed in PBS. Nuclei were obtained by lysing scraped cells in hypotonic buffer (5mM Pipes pH 6.8, 85mM KCl, 0.5% NP-40 and protease inhibitors), followed by centrifugation. Nuclei were resuspended in RIPA 100mM buffer (20mM Tris-HCl pH 7.5, 100mM NaCl, 1mM EDTA, 0.5% NP-40, 0.5% Na-Deoxycholate, 0.1% SDS supplemented with protease inhibitors). Chromatin was sonicated to 500—800 bp average fragment size and pre-cleared for 1 h at 4°C with protein A/G PLUS-Agarose beads (Santa Cruz, sc-2003). Agarose was removed by centrifugation and an aliquot of supernatant was taken as input. Chromatin was immunoprecipitated overnight at 4 °C with the following antibodies: mouse monoclonal anti-Flag M2, clone M2 (2.5 mg/ml, F1804) and anti-H3K9me3 (Upstate), mouse monoclonal anti-KMT1A/SUV39H1 (ab12405) and rabbit polyclonal anti-TLS/FUS (ab23439, abcam). DNA–protein complexes were recovered with protein A/G PLUS Agarose beads and washed with RIPA 100mM buffer, RIPA 250mM buffer (20mM Tris-HCl pH 7.5, 250mM NaCl, 1mM EDTA, 0.5% NP-40, 0.5% Na-Deoxycholate and 0.1% SDS), LiCl solution (10mM Tris-HCl pH 8.0, 1mM EDTA, 250mM LiCl, 0.5% NP-40 and 0.5% Na-Deoxycholate) and 1xTris-EDTA (TE). RNase treatment was performed in TE for 30 min at 37°C. Crosslinking was reversed by overnight incubation at 68°C after adding an equal volume of proteinase K solution (200mM NaCl, 1% SDS and 0.3 mg ml^{-1} proteinase K). Samples were resuspended in ddH2O after phenol/chloroform extraction and ethanol precipitation. Co-immunoprecipitated DNA was analyzed by qRT-PCR. ChIP data of Oct4 promoter regions were normalized against input and the Oct4 unrelated Dkk or AChR gene. Primer sequences are the following:

Mouse

- Oct4 (primer A) F: TGCACCCCCTCCTCCTAATCC;
- Oct4 (primer A) R: CCCTAAACAAGTACTCAACCC;
- Dkk F: GGGAACCAGGGAAAGAGGA;
- Dkk R: GGGAAATA GGCACCCGATAA.

Human

- OCT4 F: GAGGATGGCAAGCTGAGAAA
- OCT4 R: CTCAATCCCCAGGACAGAAC
- AChR F: CAACCAAAGCCCATGTCCTC
- AChR R: AGGCACGCTACAGGGCTTC

5.10 Co-Immunoprecipitation

For co-immunoprecipitation experiments cells were lysed in RIPA buffer (50 mM Tris-Cl, pH 7.5, 1% Nonidet P-40 (NP-40), 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA,150 mM NaCl) with protease inhibitors. Samples were pre-cleared with protein A/G PLUS-Agarose (sc-2003; Santa Cruz Biotechnology) for 1 h and cleared by centrifugation. The lysates were incubated overnight at 4 °C with the following antibodies: rabbit polyclonal anti-TLS/FUS (ab23439, abcam) and mouse monoclonal anti-KMT1A/SUV39H1 (2.5 mg/ml, ab12405, Abcam). After 3 h of incubation with protein A/G PLUS-Agarose, immunoprecipitates were washed three times in RIPA buffer, resuspended in sample buffer, and analyzed by western blotting.

5.11 Statistical analysis

A two-tailed t-test was performed to calculate P values and statistical significance was set at $p<0.05$. Each finding was confirmed by three independent biological replicates, unless specified. Error bars represent standard deviation.

Bibliography

Abiko, K., Mandai, M., Hamanishi, J., Matsumura, N., Baba, T., Horiuchi, A., Mikami, Y., Yoshioka, S., Wakasa, T., Shiozawa, T., et al. (2010). Oct4 expression in immature teratoma of the ovary: Relevance to histologic grade and degree of differentiation. Am. J. Surg. Pathol.

Abubaker, K., Latifi, A., Luwor, R., Nazaretian, S., Zhu, H., Quinn, M.A., Thompson, E.W., Findlay, J.K., and Ahmed, N. (2013). Short-term single treatment of chemotherapy results in the enrichment of ovarian cancer stem cell-like cells leading to an increased tumor burden. Mol. Cancer.

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol. Cell.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science (80-.).

Baechtold, H., Kuroda, M., Sok, J., Ron, D., Lopez, B.S., and Akhmedov, A.T. (1999). Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. J. Biol. Chem.

Balakirev, E.S., and Ayala, F.J. (2003). Pseudogenes: Are They "Junk" or Functional DNA? Annu. Rev. Genet.

Bapat, S.A., Mali, A.M., Koppikar, C.B., and Kurrey, N.K. (2005). Stem and progenitor-like cells contribute to the aggressive behavior of human epithelial ovarian cancer. Cancer Res.

Belly, A., Moreau-Gachelin, F., Sadoul, R., and Goldberg, Y. (2005). Delocalization of the multifunctional RNA splicing factor TLS/FUS in hippocampal neurones: Exclusion from the nucleus and accumulation in dendritic granules and spine heads. Neurosci. Lett.

Beltran, A.S., Rivenbark, A.G., Richardson, B.T., Yuan, X., Quian, H., Hunt, J.P., Zimmerman, E., Graves, L.M., and Blancafort, P. (2011). Generation of tumor-initiating cells

by exogenous delivery of OCT4 transcription factor. Breast Cancer Res.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell.

Bertolotti, A., Lutz, Y., Heard, D.J., Chambon, P., and Toral, L. (1996). hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. EMBO J.

Bier, A., Oviedo-Landaverde, I., Zhao, J., Mamane, Y., Kandouz, M., and Batist, G. (2009). Connexin43 pseudogene in breast cancer cells offers a novel therapeutic target. Mol. Cancer Ther.

Bischof, J.M., Chiang, A.P., Scheetz, T.E., Stone, E.M., Casavant, T.L., Sheffield, V.C., and Braun, T.A. (2006). Genome-wide identification of pseudogenes capable of disease-causing gene conversion. Hum. Mutat.

Bodner, M., Castriilo, J.L., Theill, L.E., Deerinck, T., Ellisman, M., and Karin, M. (1988). The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell.

Boyer, L.A., Tong, I.L., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell.

Brosch, M., Saunders, G.I., Frankish, A., Collins, M.O., Yu, L., Wright, J., Verstraten, R., Adams, D.J., Harrow, J., Choudhary, J.S., et al. (2011). Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. Genome Res.

Brumbaugh, J., Hou, Z., Russell, J.D., Howden, S.E., Yu, P., Ledvina, A.R., Coon, J.J., and Thomson, J.A. (2012). Phosphorylation regulates human OCT4. Proc. Natl. Acad. Sci.

Carlile, M., Swan, D., Jackson, K., Preston-Fayers, K., Ballester, B., Flicek, P., and Werner, A. (2009). Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues. Nucleic Acids Res.

Carystinos, G.D., Bier, A., and Batist, G. (2001). The role of connexin-mediated cell-cell communication in breast cancer metastasis. J. Mammary Gland Biol. Neoplasia.

Cazalla, D., Yario, T., and Steitz, J.A. (2010). Down-regulation of a host {microRNA} by a {Herpesvirus} saimiri noncoding {RNA}. Science (80-.).

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell.

Chang, S.F., and Cheng, C.L. (1998). The suppression effect of DNA sequences within the C4A region on the transcription activity of human CYP21. In Endocrine Research, p.

Chen, L., Roake, C.M., Freund, A., Batista, P.J., Tian, S., Yin, Y.A., Gajera, C.R., Lin, S., Lee, B., Pech, M.F., et al. (2018). An Activity Switch in Human Telomerase Based on RNA Conformation and Shaped by TCAB1. Cell.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008a). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell.

Chen, Y.C., Hsu, H.S., Chen, Y.W., Tsai, T.H., How, C.K., Wang, C.Y., Hung, S.C., Chang, Y.L., Tsai, M.L., Lee, Y.Y., et al. (2008b). Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells. PLoS One.

Chen, Z., Wang, T., Cai, L., Su, C., Zhong, B., Lei, Y., and Xiang, A. (2012). Clinicopathological significance of non-small cell lung cancer with high prevalence of Oct-4 tumor cells. J. Exp. Clin. Cancer Res.

Cheng, L., Thomas, A., Roth, L.M., Zheng, W., Michael, H., and Karim, F.W.A. (2004). OCT4: A novel biomarker for dysgerminoma of the ovary. Am. J. Surg. Pathol.

Chiefari, E., Iiritano, S., Paonessa, F., Le Pera, I., Arcidiacono, B., Filocamo, M., Foti, D., Liebhaber, S.A., and Brunetti, A. (2010). Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. Nat. Commun.

Colombrita, C., Onesto, E., Megiorni, F., Pizzuti, A., Baralle, F.E., Buratti, E., Silani, V., and

Ratti, A. (2012). TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. J. Biol. Chem.

Comisso, E., Scarola, M., Rosso, M., Piazza, S., Marzinotto, S., Ciani, Y., Orsaria, M., Mariuzzi, L., Schneider, C., and Benetti, R. (2017). OCT4 controls mitotic stability and inactivates the RB tumor suppressor pathway to enhance ovarian cancer aggressiveness. Oncogene.

Cronin, C.N., Zhang, X., Thompson, D.A., and McIntire, W.S. (1998). cDNA cloning of two splice variants of a human copper-containing monoamine oxidase pseudogene containing a dimeric Alu repeat sequence. Gene.

Crozat, A., Åman, P., Mandahl, N., and Ron, D. (1993). Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature.

D'Errico, I., Gadaleta, G., and Saccone, C. (2004). Pseudogenes in metazoa: Origin and features. Briefings Funct. Genomics Proteomics.

Dean, M., Fojo, T., and Bates, S. (2005). Tumour stem cells and drug resistance. Nat. Rev. Cancer.

Deng, H., Gao, K., and Jankovic, J. (2014). The role of FUS gene variants in neurodegenerative diseases. Nat. Rev. Neurol.

Dichmann, D.S., and Harland, R.M. (2012). Fus/TLS orchestrates splicing of developmental regulators during gastrulation. Genes Dev.

Ding, W., Lin, L., Chen, B., and Dai, J. (2006). L1 elements, processed pseudogenes and retrogenes in mammalian genomes. IUBMB Life.

Dini Modigliani, S., Morlando, M., Errichelli, L., Sabatelli, M., and Bozzoni, I. (2014). An ALS-associated mutation in the FUS 3'2-UTR disrupts a microRNA-FUS regulatory circuitry. Nat. Commun.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells.

Nature.

Doolittle, W.F. (2013). Is junk DNA bunk? A critique of ENCODE. Proc. Natl. Acad. Sci.

Doolittle, F., Brunet, T.D.P., Linquist, S., and Gregory, T.R. (2014). Distinguishing between "Function" and "Effect" in genome biology. Genome Biol. Evol.

Dubbink, H.J., De Waal, L., Van Haperen, R., Verkaik, N.S., Trapman, J., and Romijn, J.C. (1998). The human prostate-specific transglutaminase gene (TGM4): Genomic organization, tissue-specific expression, and promoter characterization. Genomics.

Duret, L., Chureau, C., Samain, S., Weissanbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science (80-.).

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. PLoS One.

Elliman, S.J., Wu, I., and Kemp, D.M. (2006). Adult tissue-specific expression of a Dppa3-derived retrogene represents a postnatal transcript of pluripotent cell origin. J. Biol. Chem.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., et al. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat. Struct. Mol. Biol.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nat. Genet.

Ezkurdia, I., Del Pozo, A., Frankish, A., Rodriguez, J.M., Harrow, J., Ashman, K., Valencia, A., and Tress, M.L. (2012). Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. Mol. Biol. Evol.

Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St. Laurent, G., Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA

is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. Nat. Med.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: New players in cell differentiation and development. Nat. Rev. Genet.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol.

Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., and Chang, H.Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc.

Foord, S.M., Marks, B., Stolz, M., Bufflier, E., Fraser, N.J., and Lee, M.G. (1996). The structure of the prostaglandin EP4 receptor gene and related pseudogenes. Genomics.

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet.

von Frowein, J., Pagel, P., Kappler, R., von Schweinitz, D., Roscher, A., and Schmid, I. (2011). MicroRNA-492 is processed from the keratin 19 gene and up-regulated in metastatic hepatoblastoma. Hepatology.

Fujii, R. (2005). TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. J. Cell Sci.

Fujii, R., Okabe, S., Urushido, T., Inoue, K., Yoshimura, A., Tachibana, T., Nishikawa, T., Hicks, G.G., and Takumi, T. (2005). The RNA binding protein TLS is translocated to dendritic spines by mGluR5 activation and regulates spine morphology. Curr. Biol.

Ganster, C., Wernstedt, A., Kehrer-Sawatzki, H., Messiaen, L., Schmidt, K., Rahner, N., Heinimann, K., Fonatsch, C., Zschocke, J., and Wimmer, K. (2010). Functional PMS2 hybrid alleles containing a pseudogene-specific missense variant trace back to a single ancient intrachromosomal recombination event. Hum. Mutat.

Gardiner, M., Toth, R., Vandermoere, F., Morrice, N.A., and Rouse, J. (2008). Identification and characterization of FUS/TLS as a new target of ATM. Biochem. J.

Gerbino, V., Carrì, M.T., Cozzolino, M., and Achsel, T. (2013). Mislocalised FUS mutants stall spliceosomal snRNPs in the cytoplasm. Neurobiol. Dis.

Ghosh, A.K., Majumder, M., Steele, R., White, R.A., and Ray, R.B. (2001). A Novel 16-Kilodalton Cellular Protein Physically Interacts with and Antagonizes the Functional Activity of c-myc Promoter-Binding Protein 1. Mol. Cell. Biol.

Graur, D., Zheng, Y., Price, N., Azevedo, R.B.R., Zufall, R.A., and Elhaik, E. (2013). On the immortality of television sets: "Function" in the human genome according to the evolution-free gospel of encode. Genome Biol. Evol.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature.

Gu, P., LeMenuet, D., Chung, A.C.-K., Mancini, M., Wheeler, D.A., and Cooney, A.J. (2005). Orphan Nuclear Receptor GCNF Is Required for the Repression of Pluripotency Genes during Retinoic Acid-Induced Embryonic Stem Cell Differentiation. Mol. Cell. Biol.

Gu, S., Jin, L., Zhang, F., Sarnow, P., and Kay, M.A. (2009). Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. Nat. Struct. Mol. Biol.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature.

Hallier, M., Lerga, A., Barnache, S., Tavitian, A., and Moreau-Gachelin, F. (1998). The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. J. Biol. Chem.

Han, P., and Chang, C.P. (2015). Long non-coding RNA and chromatin remodeling. RNA Biol.

Han, Y.J., Ma, S.F., Yourek, G., Park, Y.-D., and Garcia, J.G.N. (2011). A transcribed pseudogene of MYLK promotes cell proliferation. FASEB J.

Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., and Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: An intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res.

Hawkins, P.G., and Morris, K. V. (2010). Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Transcription.

Hayashi, H., Arao, T., Togashi, Y., Kato, H., Fujita, Y., De Velasco, M.A., Kimura, H., Matsumoto, K., Tanaka, K., Okamoto, I., et al. (2015). The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. Oncogene.

He, W., Li, K., Wang, F., Qin, Y.R., and Fan, Q.X. (2012). Expression of OCT4 in human esophageal squamous cell carcinoma is significantly associated with poorer prognosis. World J. Gastroenterol.

Hicks, G.G., Singh, N., Nashabi, A., Mai, S., Bozek, G., Klewes, L., Arapovic, D., White, E.K., Koury, M.J., Oltz, E.M., et al. (2000). Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. Nat. Genet.

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell.

Hoell, J.I., Larsson, E., Runge, S., Nusbaum, J.D., Duggimpudi, S., Farazi, T.A., Hafner, M., Borkhardt, A., Sander, C., and Tuschl, T. (2011). RNA targets of wild-type and mutant FET family proteins. Nat. Struct. Mol. Biol.

Hogan, M.S., Parfitt, D.E., Zepeda-Mendoza, C.J., Shen, M.M., and Spector, D.L. (2015). Transient pairing of homologous Oct4 alleles accompanies the onset of embryonic stem cell differentiation. Cell Stem Cell.

Hu, L., McArthur, C., and Jaffe, R.B. (2010). Ovarian cancer stem-like side-population cells are tumourigenic and chemoresistant. Br. J. Cancer.

Hu, X., Yang, L., and Mo, Y.Y. (2018). Role of pseudogenes in tumorigenesis. Cancers (Basel).

Huang, P., Chen, J., Wang, L., Na, Y., Kaku, H., Ueki, H., Sasaki, K., Yamaguchi, K., Zhang, K., Saika, T., et al. (2012). Implications of transcriptional factor, OCT-4, in human bladder malignancy and tumor recurrence. Med. Oncol.

Huarte, M. (2013). LncRNAs have a say in protein translation. Cell Res.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. Nat. Rev. Genet.

Iko, Y., Kodama, T.S., Kasai, N., Oyama, T., Morita, E.H., Muto, T., Okumura, M., Fujii, R., Takumi, T., Tate, S.I., et al. (2004). Domain architectures and characterization of an RNA-binding protein, TLS. J. Biol. Chem.

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R., et al. (2013). Tandem stem-loops in roX RNAs act together to mediate X Chromosome dosage compensation in Drosophila. Mol. Cell.

Ilik, I.A., Maticzka, D., Georgiev, P., Gutierrez, N.M., Backofen, R., and Akhtar, A. (2017). A mutually exclusive stem-loop arrangement in roX2 RNA is essential for X-chromosome regulation in Drosophila. Genes Dev.

Indrieri, A., Grimaldi, C., Zucchelli, S., Tammaro, R., Gustincich, S., and Franco, B. (2016). Synthetic long non-coding RNAs [SINEUPs] rescue defective gene expression in vivo. Sci. Rep.

Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L., and Rosenfeld, M.G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell.

Ioffe, Y.J., Chiappinelli, K.B., Mutch, D.G., Zighelboim, I., and Goodfellow, P.J. (2012). Phosphatase and tensin homolog (PTEN) pseudogene expression in endometrial cancer: a conserved regulatory mechanism important in tumorigenesis? Gynecol. Oncol.

Iskow, R.C., Gokcumen, O., Abyzov, A., Malukiewicz, J., Zhu, Q., Sukumar, A.T., Pai, A.A., Mills, R.E., Habegger, L., Cusanovich, D.A., et al. (2012). Regulatory element copy number differences shape primate expression profiles. Proc. Natl. Acad. Sci.

Jang, H., Kim, T.W., Yoon, S., Choi, S.Y., Kang, T.W., Kim, S.Y., Kwon, Y.W., Cho, E.J., and Youn, H.D. (2012). O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. Cell Stem Cell.

Jeyapalan, Z., Deng, Z., Shatseva, T., Fang, L., He, C., and Yang, B.B. (2011). Expression of CD44 3′-untranslated region regulates endogenous microRNA functions in tumorigenesis and angiogenesis. Nucleic Acids Res.

Jez, M., Ambady, S., Kashpur, O., Grella, A., Malcuit, C., Vilner, L., Rozman, P., and Dominko, T. (2014). Expression and differentiation between OCT4A and its pseudogenes in human ESCs and differentiated adult somatic cells. PLoS One.

Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W.O., Corcoran, M., Grandér, D., and Morris, K. V. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nat. Struct. Mol. Biol.

Johnsson, P., Morris, K. V., and Grandér, D. (2014). Pseudogenes: A novel source of trans - acting antisense RNAs. Methods Mol. Biol.

Jóźwicki, W., Brożyna, A.A., and Siekiera, J. (2014). Expression of OCT4A: The first step to the next stage of urothelial bladder cancer progression. Int. J. Mol. Sci.

Kaltz, N., Funari, A., Hippauf, S., Delorme, B., Noël, D., Riminucci, M., Jacobs, V.R., Häupl, T., Jorgensen, C., Charbord, P., et al. (2008). In Vivo Osteoprogenitor Potency of Human Stromal Cells from Different Tissues Does Not Correlate with Expression of POU5F1 or Its Pseudogenes. Stem Cells.

Kandouz, M., Bier, A., Carystinos, G.D., Alaoui-Jamali, M.A., and Batist, G. (2004). Connexin43 pseudogene is expressed in tumor cells and inhibits growth. Oncogene.

Karoubi, G., Gugger, M., Schmid, R., and Dutly, A. (2009). OCT4 expression in human non-small cell lung cancer: implications for therapeutic intervention. Interact. Cardiovasc. Thorac. Surg.

Karreth, F.A., Tay, Y., Perna, D., Ala, U., Tan, S.M., Rust, A.G., Denicola, G., Webster, K.A., Weiss, D., Perez-Mancera, P.A., et al. (2011). In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell.

Karro, J.E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrrison, P., and Gerstein, M. (2007). Pseudogene.org: A comprehensive database and comparison platform for pseudogene annotation. Nucleic Acids Res.

Kastler, S., Honold, L., Luedeke, M., Kuefer, R., Möller, P., Hoegel, J., Vogel, W., Maier, C., and Assum, G. (2010). POU5F1P1, a putative cancer susceptibility gene, is overexpressed in prostatic carcinoma. Prostate.

Kellner, S., and Kikyo, N. (2010). Transcriptional regulation of the Oct4 gene, a master gene for pluripotency. Histol. Histopathol.

Kemler, I., Schreiber, E., Müller, M.M., Matthias, P., and Schaffner, W. (1989). Octamer transcription factors bind to two different sequence motifs of the immunoglobulin heavy chain promoter. EMBO J.

Khachane, A.N., and Harrison, P.M. (2009). Assessing the genomic evidence for conserved transcribed pseudogenes under selection. BMC Genomics.

Kim, R.-J., and Nam, J.-S. (2011). OCT4 Expression Enhances Features of Cancer Stem Cells in a Mouse Model of Breast Cancer. Lab. Anim. Res.

Kim, V.N., and Nam, J.W. (2006). Genomics of microRNA. Trends Genet.

Kim, J.B., Sebastiano, V., Wu, G., Araúzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., et al. (2009). Oct4-Induced Pluripotency in Adult Neural Stem Cells. Cell.

Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, et al. (2014). A draft map of the human proteome. Nature.

Kong, D., Su, G., Zha, L., Zhang, H., Xiang, J., Xu, W., Tang, Y., and Wang, Z. (2014). Coexpression of HMGA2 and Oct4 predicts an unfavorable prognosis in human gastric cancer. Med. Oncol.

Korneev, S., Park, J., and O'Shea, M. (1999). Neuronal Expression of Neural Nitric Oxide Synthase (nNOS) Protein Is Suppressed by an Antisense RNA Transcribed from an NOS Pseudogene. J. Neurosci.

Kuroda, M. (2000). Male sterility and enhanced radiation sensitivity in TLS-/- mice. EMBO J.

Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet.

Lagier-Tourenne, C., Polymenidou, M., Hutt, K.R., Vu, A.Q., Baughn, M., Huelga, S.C., Clutario, K.M., Ling, S.C., Liang, T.Y., Mazur, C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat. Neurosci.

Latifi, A., Luwor, R.B., Bilandzic, M., Nazaretian, S., Stenvers, K., Pyman, J., Zhu, H., Thompson, E.W., Quinn, M.A., Findlay, J.K., et al. (2012). Isolation and Characterization of Tumor Cells from the Ascites of Ovarian Cancer Patients: Molecular Phenotype of Chemoresistant Ovarian Tumors. PLoS One.

Law, W.J., Cann, K.L., and Hicks, G.G. (2006). TLS, EWS and TAF15: A model for transcriptional integration of gene expression. Briefings Funct. Genomics Proteomics.

Leder, P. (1982). The genetics of antibody diversity. Sci. Am.

Lee, J.T., and Bartolomei, M.S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell.

Lee, D.Y., Shatseva, T., Jeyapalan, Z., Du, W.W., Deng, Z., and Yang, B.B. (2009). A 3′-untranslated region (3′UTR) induces organ adhesion by regulating miR-199a* functions. PLoS One.

Lee, D.Y., Jeyapalan, Z., Fang, L., Yang, J., Zhang, Y., Yee, A.Y., Li, M., Du, W.W., Shatseva, T., and Yang, B.B. (2010). Expression of versican 3′-untranslated region modulates endogenous microrna functions. PLoS One.

Lee, H.H., Niu, D.M., Lin, R.W., Chan, P., and Lin, C.Y. (2002). Structural analysis of the chimeric CYP21P/CYP21 gene in steroid 21-hydroxylase deficiency. J. Hum. Genet.

Lee, S.C., Fang, L., Wang, C.H., Kahai, S., Deng, Z., and Yang, B.B. (2011). A non-coding transcript of nephronectin promotes osteoblast differentiation by modulating microRNA functions. FEBS Lett.

Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J., and Moreau-Gachelin, F. (2001). Identification of an RNA Binding Specificity for the Potential Splicing Factor TLS. J. Biol. Chem.

Li, N., Deng, W., Ma, J., Wei, B., Guo, K., Shen, W., Zhang, Y., and Luo, S. (2015). Prognostic evaluation of Nanog, Oct4, Sox2, PCNA, Ki67 and E-cadherin expression in gastric cancer. Med. Oncol.

Li, Z., Li, X., Li, C., Su, Y., Fang, W., Zhong, C., Ji, W., Zhang, Q., and Su, C. (2014). Transcription factor OCT4 promotes cell cycle progression by regulating CCND1 expression in esophageal carcinoma. Cancer Lett.

Libri, V., Helwak, A., Miesen, P., Santhakumar, D., Borger, J.G., Kudla, G., Grey, F., Tollervey, D., and Buck, A.H. (2012). Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. Proc. Natl. Acad. Sci.

Liedtke, S., Enczmann, J., Waclawczyk, S., Wernet, P., and Kögler, G. (2007). Oct4 and Its Pseudogenes Confuse Stem Cell Research. Cell Stem Cell.

Lin, H., Shabbir, A., Molnar, M., and Lee, T. (2007). Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. Biochem. Biophys. Res. Commun.

Lin, H., Sun, L.H., Han, W., He, T.Y., Xu, X.J., Cheng, K., Geng, C., Su, L.D., Wen, H., Wang, X.Y., et al. (2014). Knockdown of OCT4 suppresses the growth and invasion of pancreatic cancer cells through inhibition of the AKT pathway. Mol. Med. Rep.

Linn, D.E., Yang, X., Sun, F., Xie, Y., Chen, H., Jiang, R., Chen, H., Chumsri, S., Burger, A.M., and Qiu, Y. (2010). A role for OCT4 in tumor initiation of drug-resistant prostate cancer cells. Genes and Cancer.

Lister, N., Shevchenko, G., Walshe, J.L., Groen, J., Johnsson, P., Vidarsdóttir, L., Grander, D., Ataide, S.F., and Morris, K. V. (2017). The molecular dynamics of long noncoding RNA control of transcription in PTEN and its pseudogene. Proc. Natl. Acad. Sci.

Liu, T., Sun, B., Zhao, X., Li, Y., Gu, Q., Dong, X., and Liu, F. (2014). OCT4 expression and vasculogenic mimicry formation positively correlate with poor prognosis in human breast cancer. Int. J. Mol. Sci.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. Nat. Rev. Mol. Cell Biol.

Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engström, P.G., Lenhard, B., Aturaliya, R.N., Batalov, S., Beisel, K.W., et al. (2006). Transcript annotation in FANTOM3: Mouse gene catalog based on physical cDNAs. PLoS Genet.

Maenner, S., Müller, M., Fröhlich, J., Langer, D., and Becker, P.B. (2013). ATP-dependent roX RNA remodeling by the Helicase maleless enables specific association of MSL proteins. Mol. Cell.

Maher, B. (2012). The human encyclopaedia. Nature.

Malvezzi, M., Carioli, G., Rodriguez, T., Negri, E., and La Vecchia, C. (2016). Global trends and predictions in ovarian cancer mortality. Ann. Oncol.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science (80-.).

Marques, A.C., Tan, J., and Ponting, C.P. (2011). Wrangling for microRNAs provokes much crosstalk. Genome Biol.

Masuda, A., Takeda, J.I., Okuno, T., Okamoto, T., Ohkawara, B., Ito, M., Ishigaki, S., Sobue, G., and Ohno, K. (2015). Position-specific binding of FUS to nascent RNA regulates mRNA length. Genes Dev.

McDonell, L., and Drouin, G. (2012). The abundance of processed pseudogenes derived from glycolytic genes is correlated with their expression level. Genome.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature.

Mighell, A.J., Smith, N.R., Robinson, P.A., and Markham, A.F. (2000). Vertebrate pseudogenes. FEBS Lett.

Morlando, M., Dini Modigliani, S., Torrelli, G., Rosa, A., Di Carlo, V., Caffarelli, E., and Bozzoni, I. (2012). FUS stimulates microRNA biogenesis by facilitating co-transcriptional

Drosha recruitment. EMBO J.

Morohoshi, F., Ootsuka, Y., Arai, K., Ichikawa, H., Mitani, S., Munakata, N., and Ohki, M. (1998). Genomic structure of the human RBP56/hTAF(II)68 and FUS/TLS genes. Gene.

Morris, K. V. (2009). Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. Epigenetics.

Mounsey, A., Bauer, P., and Hope, I.A. (2002). Evidence suggesting that a fifth of annotated Caenorhabditis elegans genes may be pseudogenes. Genome Res.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol.

Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T., and Yamanaka, S. (2010). Promotion of direct reprogramming by transformation-deficient Myc. Proc. Natl. Acad. Sci.

Naumann, M., Pal, A., Goswami, A., Lojewski, X., Japtok, J., Vehlow, A., Naujock, M., Günther, R., Jin, M., Stanslowsky, N., et al. (2018). Impaired DNA damage response signaling by FUS-NLS mutations leads to neurodegeneration and FUS aggregate formation. Nat. Commun.

Ng, S.Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J.

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

Niwa, H., Miyazaki, J.I., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell.

Okamura, K., and Lai, E.C. (2008). Endogenous small interfering RNAs in animals. Nat. Rev. Mol. Cell Biol.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germ-line competent induced pluripotent stem cells. Nature.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. Science (80-.).

Olsen, M.A., and Schechter, L.E. (1999). Cloning, mRNA localization and evolutionary conservation of a human 5- HT7 receptor pseudogene. Gene.

Pain, D., Chirn, G.W., Strassel, C., and Kemp, D.M. (2005). Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. J. Biol. Chem.

Panagopoulos, I., Möller, E., Collin, A., and Mertens, F. (2008). The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1. Oncol. Rep.

Patrucco, L., Chiesa, A., Soluri, M.F., Fasolo, F., Takahashi, H., Carninci, P., Zucchelli, S., Santoro, C., Gustincich, S., Sblattero, D., et al. (2015). Engineering mammalian cell factories with SINEUP noncoding RNAs to improve translation of secreted proteins. Gene.

Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., et al. (2012). The GENCODE pseudogene resource. Genome Biol. *13*, R51.

Perrotti, D., Bonatti, S., Trotta, R., Martinez, R., Skorski, T., Salomoni, P., Grassilli, E., Iozzo, R. V., Cooper, D.R., and Calabretta, B. (1998). TLS/FUS, a pro-oncogene involved in multiple chromosomal translocations, is a novel regulator of BCR/ABL-mediated leukemogenesis. EMBO J.

Petermann, R., Mossier, B.M., Aryee, D.N.T., Khazak, V., Golemis, E.A., and Kovar, H. (1998). Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II. Oncogene.

Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., and Carter, D.R.F. (2011).

Pseudogenes: Pseudo-functional or key regulators in health and diseasě. RNA.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., et al. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol. Cell.

Poliseno, L. (2012). Pseudogenes: Newly discovered players in human cancer. Sci. Signal.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., and Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. Cell.

Popowski, M., Templeton, T.D., Lee, B.K., Rhee, C., Li, H., Miner, C., Dekker, J.D., Orlanski, S., Bergman, Y., Iyer, V.R., et al. (2014). Bright/Arid3A acts as a barrier to somatic cell reprogramming through direct regulation of Oct4, Sox2, and Nanog. Stem Cell Reports.

Poursani, E.M., Soltani, B.M., and Mowla, S.J. (2016). Differential expression of OCT4 pseudogenes in pluripotent and tumor cell lines. Cell J.

Powers, C.A., Mathur, M., Raaka, B.M., Ron, D., and Samuels, H.H. (1998). TLS (Translocated-in-Liposarcoma) Is a High-Affinity Interactor for Steroid, Thyroid Hormone, and Retinoid Receptors. Mol. Endocrinol.

Prasad, D.D., Ouchida, M., Lee, L., Rao, V.N., and Reddy, E.S. (1994). TLS/FUS fusion domain of TLS/FUS-erg chimeric protein resulting from the t(16;21) chromosomal translocation in human myeloid leukemia functions as a transcriptional activation domain. Oncogene.

Prensner, J.R., and Chinnaiyan, A.M. (2011). The emergence of lncRNAs in cancer biology. Cancer Discov.

Puget, N., Gad, S., Perrin-Vidoz, L., Sinilnikova, O.M., Stoppa-Lyonnet, D., Lenoir, G.M., and Mazoyer, S. (2002). Distinct BRCA1 Rearrangements Involving the BRCA1 Pseudogene

Suggest the Existence of a Recombination Hot Spot. Am. J. Hum. Genet.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell.

Ratti, A., and Buratti, E. (2016). Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. J. Neurochem.

Redshaw, Z., and Strain, A.J. (2010). Human haematopoietic stem cells express Oct4 pseudogenes and lack the ability to initiate Oct4 promoter-driven gene expression. J. Negat. Results Biomed.

Rijlaarsdam, M.A., Van Herk, H.A.D.M., Gillis, A.J.M., Stoop, H., Jenster, G., Martens, J., Van Leenders, G.J.L.H., Dinjens, W., Hoogland, A.M., Timmermans, M., et al. (2011). Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: Confirmation of OCT3/4 specificity for germ cell tumours. Br. J. Cancer.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem.

da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Mol. Cell.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of Nanog by OCT4 and SOX2. J. Biol. Chem.

Rogalla, P., Kazmierczak, B., Flohr, A.M., Hauke, S., and Bullerdiek, J. (2000). Back to the roots of a new exon - The molecular archaeology of a SP100 splice variant. Genomics.

Rogelj, B., Easton, L.E., Bogu, G.K., Stanton, L.W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., Haberman, N., et al. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. Sci. Rep.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirie, F., Rigby, P.W.J., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the

mammalian embryo. Nature.

Rungarunlert, S., Klincumhom, N., Tharasanit, T., Techakumphu, M., Pirity, M.K., and Dinnyes, A. (2013). Slow Turning Lateral Vessel Bioreactor Improves Embryoid Body Formation and Cardiogenic Differentiation of Mouse Embryonic Stem Cells. Cell. Reprogram.

La Russa, M.F., and Qi, L.S. (2015). The New State of the Art: Cas9 for Gene Activation and Repression. Mol. Cell. Biol.

Rutnam, Z.J., and Yang, B.B. (2012). The non-coding 3′ UTR of CD44 induces metastasis by regulating extracellular matrix functions. J. Cell Sci.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? Cell.

Samardzija, C., Quinn, M., Findlay, J.K., and Ahmed, N. (2012). Attributes of Oct4 in stem cell biology: Perspectives on cancer stem cells of the ovary. J. Ovarian Res.

Samardzija, C., Luwor, R.B., Volchek, M., Quinn, M.A., Findlay, J.K., and Ahmed, N. (2015). A critical role of Oct4A in mediating metastasis and disease-free survival in a mouse model of ovarian cancer. Mol. Cancer.

Sasidharan, R., and Gerstein, M. (2008). Genomics: Protein fossils live on as RNA. Nature.

Sass, G.L., Pannuti, A., and Lucchesi, J.C. (2003). Male-specific lethal complex of *Drosophila* targets activated regions of the X chromosome for chromatin remodeling. Proc. Natl. Acad. Sci.

Sasso, E., Latino, D., Froechlich, G., Succoio, M., Passariello, M., De Lorenzo, C., Nicosia, A., and Zambrano, N. (2018). A long non-coding SINEUP RNA boosts semi-stable production of fully human monoclonal antibodies in HEK293E cells. MAbs.

Sato, N., Kondo, M., and Arai, K. ichi (2006). The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. Biochem. Biophys. Res. Commun.

Saxe, J.P., Tomilin, A., Schöler, H.R., Plath, K., and Huang, J. (2009). Post-translational regulation of Oct4 transcriptional activity. PLoS One.

Scarola, M., Comisso, E., Pascolo, R., Chiaradia, R., Maria Marion, R., Schneider, C., Blasco, M.A., Schoeftner, S., and Benetti, R. (2015). Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat. Commun.

Schein, A., Zucchelli, S., Kauppinen, S., Gustincich, S., and Carninci, P. (2016). Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. Sci. Rep.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J.

Schoeftner, S., Scarola, M., Comisso, E., Schneider, C., and Benetti, R. (2013). An Oct4-pRb axis, controlled by miR-335, integrates stem cell self-renewal and cell cycle control. Stem Cells.

Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. Nature.

Schwartz, J.C., Ebmeier, C.C., Podell, E.R., Heimiller, J., Taatjes, D.J., and Cech, T.R. (2012). FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. Genes Dev.

Seitz, H. (2009). Redefining MicroRNA Targets. Curr. Biol.

Sisu, C., Pei, B., Leng, J., Frankish, A., Zhang, Y., Balasubramanian, S., Harte, R., Wang, D., Rutenberg-Schoenberg, M., Clark, W., et al. (2014). Comparative analysis of pseudogenes across three phyla. Proc. Natl. Acad. Sci.

Song, R., Hennig, G.W., Wu, Q., Jose, C., Zheng, H., and Yan, W. (2011). Male germ cells express abundant endogenous siRNAs. Proc. Natl. Acad. Sci.

Sorge, J., Gross, E., West, C., and Beutler, E. (1990). High level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with Gaucher disease. J. Clin. Invest.

Spelat, R., Ferro, F., and Curcio, F. (2012). Serine 111 phosphorylation regulates OCT4A protein subcellular distribution and degradation. J. Biol. Chem.

Sterneckert, J., Höing, S., and Schöler, H.R. (2012). Concise review: Oct4 and more: The reprogramming expressway. Stem Cells.

Sumazin, P., Yang, X., Chiu, H.S., Chung, W.J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J., et al. (2011). An extensive MicroRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell.

Sun, C., Orozco, O., Olson, D.L., Choi, E., Garber, E., Tizard, R., Szak, S., Sanicola, M., and Carulli, J.P. (2008). CRIPTO3, a presumed pseudogene, is expressed in cancer. Biochem. Biophys. Res. Commun.

Sun, S., Ling, S.C., Qiu, J., Albuquerque, C.P., Zhou, Y., Tokunaga, S., Li, H., Qiu, H., Bui, A., Yeo, G.W., et al. (2015). ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. Nat. Commun.

Suo, G., Han, J., Wang, X., Zhang, J., Zhao, Y., Zhao, Y., and Dai, J. (2005). Oct4 pseudogenes are transcribed in cancers. Biochem. Biophys. Res. Commun.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev.

Taft, R.J., Pheasant, M., and Mattick, J.S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays.

Takahama, K., Kino, K., Arai, S., Kurokawa, R., and Oyoshi, T. (2008). Identification of RNA binding specificity for the TET-family proteins. Nucleic Acids Symp. Ser.

Takahashi, H., Kozhuharova, A., Sharma, H., Hirose, M., Ohyama, T., Fasolo, F., Yamazaki, T., Cotella, D., Santoro, C., Zucchelli, S., et al. (2018). Identification of functional features of synthetic SINEUPs, antisense lncRNAs that specifically enhance protein translation. PLoS One.

Takahashi, K., Mitsui, K., and Yamanaka, S. (2003). Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. Nature.

Takahashi, K., Yamanaka, S., Zhang, Y., Li, Y., Feng, C., Li, X., Lin, L., Guo, L., Wang, H., Liu, C., et al. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature.

Tantin, D. (2013). Oct transcription factors in development and stem cells: insights and mechanisms. Development.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature.

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F., et al. (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell.

Tay, Y., Rinn, J., and Pandolfi, P.P. (2014). The multilayered complexity of ceRNA crosstalk and competition. Nature.

Tchnio, T., and Segal-bendirdjian, E. (1993). Generation of processed pseudogenes in murine cells. Nucleic Acids Res.

Trabesinger-Ruef, N., Jermann, T., Zankel, T., Durrant, B., Frank, G., and Benner, S.A. (1996). Pseudogenes in ribonuclease evolution: A source of new biomacromolecular function? FEBS Lett.

Troyanovsky, S.M., and Leube, R.E. (1994). Activation of the Silent Human Cytokeratin 17 Pseudogene???Promoter Region by Cryptic Enhancer Elements of the Cytokeratin 17 Gene. Eur. J. Biochem.

Tsai, L.L., Hu, F.W., Lee, S.S., Yu, C.H., Yu, C.C., and Chang, Y.C. (2014). Oct4 mediates tumor initiating properties in oral squamous cell carcinomas through the regulation of epithelial-mesenchymal transition. PLoS One.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E.,

and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science (80-.).

Tsuiji, H., Iguchi, Y., Furuya, A., Kataoka, A., Hatsuta, H., Atsuta, N., Tanaka, F., Hashizume, Y., Akatsu, H., Murayama, S., et al. (2013). Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. EMBO Mol. Med.

Udagawa, T., Fujioka, Y., Tanaka, M., Honda, D., Yokoi, S., Riku, Y., Ibi, D., Nagai, T., Yamada, K., Watanabe, H., et al. (2015). FUS regulates AMPA receptor function and FTLD/ALS-associated behaviour via GluA1 mRNA stabilization. Nat. Commun.

Uranishi, H., Tetsuka, T., Yamashita, M., Asamitsu, K., Shimizu, M., Itoh, M., and Okamoto, T. (2001). Involvement of the Pro-oncoprotein TLS (Translocated in Liposarcoma) in Nuclear Factor-κB p65-mediated Transcription as a Coactivator. J. Biol. Chem.

Vanin, E.F. (1984). Processed pseudogenes. Characteristics and evolution. BBA - Gene Struct. Expr.

Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. Proc. Natl. Acad. Sci.

Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. Nat. Rev. Drug Discov.

Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., Chen, N., Sun, F., and Fan, Q. (2010a). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res.

Wang, K., Sengupta, S., Magnani, L., Wilson, C.A., Henry, R.W., and Knott, J.G. (2010b). Brg1 is required for Cdx2-mediated repression of Oct4 expression in mouse blastocysts. PLoS One.

Wang, L., Guo, Z.Y., Zhang, R., Xin, B., Chen, R., Zhao, J., Wang, T., Wen, W.H., Jia, L.T., Yao, L.B., et al. (2013a). Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. Carcinogenesis.

Wang, W.Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., MacKenzie, I.R.A., Huang, E.J., and Tsai, L.H. (2013b). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat. Neurosci.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature.

Wang, X.Q., Ongkeko, W.M., Chen, L., Yang, Z.F., Lu, P., Chen, K.K., Lopez, J.P., Poon, R.T.P., and Fan, S.T. (2010c). Octamer 4 (Oct4) mediates chemotherapeutic drug resistance in liver cancer cells through a potential Oct4-AKT-ATP-binding cassette G2 pathway. Hepatology.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature.

Wei, C.-C., Nie, F.-Q., Jiang, L.-L., Chen, Q.-N., Chen, Z.-Y., Chen, X., Pan, X., Liu, Z.-L., Lu, B.-B., and Wang, Z.-X. (2017). The pseudogene DUXAP10 promotes an aggressive phenotype through binding with LSD1 and repressing LATS2 and RRAD in non small cell lung cancer. Oncotarget.

Werner, A., and Swan, D. (2010). What are natural antisense transcripts good for? Biochem. Soc. Trans.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature.

Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2008). c-Myc Is Dispensable for Direct Reprogramming of Mouse Fibroblasts. Cell Stem Cell.

Wezel, F., Pearson, J., Kirkwood, L.A., and Southgate, J. (2013). Differential expression of oct4 variants and pseudogenes in normal urothelium and urothelial cancer. Am. J. Pathol.

Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: MicroRNA biogenesis pathways and their regulation. Nat. Cell Biol.

World Cancer Research Funds (2018). Ovarian cancer statistics.

Wright, M.W., and Bruford, E.A. (2011). Naming "junk": Human non-protein coding RNA (ncRNA) gene nomenclature. Hum. Genomics.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet.

Xu, J., and Zhang, J. (2016). Are human translated pseudogenes functional'. Mol. Biol. Evol.

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. (2009). MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. Cell.

Yamazaki, T., Chen, S., Yu, Y., Yan, B., Haertlein, T.C., Carrasco, M.A., Tapia, J.C., Zhai, B., Das, R., Lalancette-Hebert, M., et al. (2012). FUS-SMN Protein Interactions Link the Motor Neuron Diseases ALS and SMA. Cell Rep.

Yang, L., Embree, L.J., and Hickstein, D.D. (2000). TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. Mol. Cell. Biol.

Yang, L., Gal, J., Chen, J., and Zhu, H. (2014). Self-assembled FUS binds active chromatin and regulates gene transcription. Proc. Natl. Acad. Sci.

Yeap, L.-S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. Epigenetics Chromatin.

Yin, J.Y., Tang, Q., Zhai, L. ling, Zhou, L. yu, Qian, J., Lin, J., Wen, X. mei, Zhou, J. dong, Zhang, Y. ying, Zhu, X. wen, et al. (2015). High expression of OCT4 is frequent and may cause undesirable treatment outcomes in patients with acute myeloid leukemia. Tumor Biol.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 Suppresses Target mRNA Translation. Mol. Cell.

Yu, Y., and Reed, R. (2015). FUS functions in coupling transcription to splicing by mediating

an interaction between RNAP II and U1 snRNP. Proc. Natl. Acad. Sci.

Yu, Y., Chi, B., Xia, W., Gangopadhyay, J., Yamazaki, T., Winkelbauer-Hurt, M.E., Yin, S., Eliasse, Y., Adams, E., Shaw, C.E., et al. (2015). U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. Nucleic Acids Res.

Yuan, P., Han, J., Guo, J., Orlov, Y.L., Huss, M., Loh, Y.H., Yaw, L.P., Robson, P., Lim, B., and Ng, H.H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev.

Zeimet, A.G., Reimer, D., Sopper, S., Boesch, M., Martowicz, A., Roessler, J., Wiedemair, A.M., Rumpold, H., Untergasser, G., Concin, N., et al. (2012). Ovarian cancer stem cells. Neoplasma.

Zeqiraj, E., and van Aalten, D.M.F. (2010). Pseudokinases-remnants of evolution or key allosteric regulators? Curr. Opin. Struct. Biol.

Zhang, Z., and Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev.

Zhang, J., Li, Y.-L., Zhou, C.-Y., Hu, Y.-T., and Chen, H.-Z. (2010a). Expression of octamer-4 in serous and mucinous ovarian carcinoma. J. Clin. Pathol. *63*, 879–883.

Zhang, X., Han, B., Huang, J., Zheng, B., Geng, Q., Aziz, F., and Dong, Q. (2010b). Prognostic significance of OCT4 expression in adenocarcinoma of the lung. Jpn. J. Clin. Oncol.

Zhang, Z.D., Frankish, A., Hunt, T., Harrow, J., and Gerstein, M. (2010c). Identification and analysis of unitary pseudogenes: Historic and contemporary gene losses in humans and other primates. Genome Biol.

Zhao, P., Liu, C., Xu, K., Zheng, S., Li, H., Xu, Y., Xu, A., Li, B., and Huang, P. (2012). [Expression of OCT4 protein in bladder cancer and its clinicopathological implications]. Nan Fang Yi Ke Da Xue Xue Bao.

Zhao, S., Yuan, Q., Hao, H., Guo, Y., Liu, S., Zhang, Y., Wang, J., Liu, H., Wang, F., Liu, K., et al. (2011). Expression of OCT4 pseudogenes in human tumours: Lessons from glioma and

breast carcinoma. J. Pathol.

Zheng, D., and Gerstein, M.B. (2007). The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? Trends Genet.

Zheng, D., Frankish, A., Baertsch, R., Kapranov, P., Reymond, A., Siew, W.C., Lu, Y., Denoeud, F., Antonarakis, S.E., Snyder, M., et al. (2007). Pseudogenes in the ENCODE regions: Consensus annotation, analysis of transcription, and evolution. Genome Res.

Zhou, Y., Liu, S., Liu, G., Öztürk, A., and Hicks, G.G. (2013). ALS-Associated FUS Mutations Result in Compromised FUS Alternative Splicing and Autoregulation. PLoS Genet.

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., and Ding, S. (2010). Reprogramming of human primary somatic cells by OCT4 and chemical compounds. Cell Stem Cell.

Zinszner, H., Sok, J., Immanuel, D., Yin, Y., and Ron, D. (1997). TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. J. Cell Sci.

Zou, M., Baitei, E.Y., Alzahrani, A.S., Al-Mohanna, F., Farid, N.R., Meyer, B., and Shi, Y. (2009). Oncogenic Activation of MAP Kinase by BRAF Pseudogene in Thyroid Tumors. Neoplasia.

Zucchelli, S., Cotella, D., Takahashi, H., Carrieri, C., Cimatti, L., Fasolo, F., Jones, M.H., Sblattero, D., Sanges, R., Santoro, C., et al. (2015b). SINEUPs: A new class of natural and synthetic antisense long non-coding RNAs that activate translation. RNA Biol.

Zucchelli, S., Fasolo, F., Russo, R., Cimatti, L., Patrucco, L., Takahashi, H., Jones, M.H., Santoro, C., Sblattero, D., Cotella, D., et al. (2015a). SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. Front. Cell. Neurosci.

Zucchelli, S., Patrucco, L., Persichetti, F., Gustincich, S., and Cotella, D. (2016). Engineering Translation in Mammalian Cell Factories to Increase Protein Yield: The Unexpected Use of Long Non-Coding SINEUP RNAs. Comput. Struct. Biotechnol. J.