UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

A long non-coding RNA regulating cellular reprogramming

Mestrado em Biologia Evolutiva e do Desenvolvimento

Sara Azevedo Barros

Dissertação orientada por:

Doutor Bruno Bernardes de Jesus

Professora Doutora Maria Gabriela Rodrigues

2016

This page intentionally left blank

Agradecimentos

Este foi um ano excelente.

Quero começar por agradecer ao Doutor Bruno Jesus, pela disponibilidade, por tudo que me ensinou e pelo apoio constante.
Ao Sérgio Marinho, por tudo que me ensinou e por todo o apoio ao longo do ano.
À professora Carmo Fonseca pelo apoio e ideias brilhantes.
Ao Jordi Garcia que, mesmo sem o saber, foi o principal responsável por eu ter decidido fazer este mestrado.
Agradeço também aos meus colegas do laboratório, Soraia e Afonso, Catarina, Vanessa, Tomás, Joana, Ana, Rita Drago, Rita Almeida e Zé, pela boa disposição, amizade e apoio constantes.
À professora Gabriela Rodrigues, tanto pelas interessantes aulas, como pela co-orientação deste trabalho.
Quero agradecer também à professora Sólveig Thorsteinsdóttir por tudo que me ensinou, por todo o apoio que sempre me dá e pelas fantásticas aulas.

À minha família, Manuela, Jorge, Berta e Vasco, por todo o apoio, amizade e por estarem sempre presentes, mesmo estando longe.
Um agradecimento especial todos os meus amigos. Eles sabem quem são.

Nota prévia

A presente dissertação encontra-se escrita em Inglês, pelo motivo de esta ser a língua-mãe da comunidade científica e por consequência possibilitar uma partilha mais abrangente. O design foi projectado de forma a facilitar a leitura e a análise deste estudo, seguindo uma estética clara. O formato é semelhante ao de uma publicação cientifica.

Este trabalho contribuiu para o manuscrito 'A long non-coding RNA antisense of Zeb2 enhances pluripotency and reprogramming', que será brevemente submetido a uma revista científica.

Resumo

A compreensão do processo de envelhecimento é fundamental para podermos contrariar e prevenir os efeitos adversos deste no declínio de células e tecidos. O envelhecimento é uma condição geral da vida. Com o passar do tempo, devido à acumulação de modificações genéticas e epigenéticas, as células estaminais tendem a perder a sua capacidade de proliferação e diferenciação, deixando de conseguir substituir tecidos danificados.

Em 2006, a descoberta de uma nova forma de reprogramação celular trouxe-nos a promessa da possibilidade de obtenção de células estaminais pluripotentes induzidas (células iPS, iPSC) a partir de células adultas do próprio indivíduo/paciente [34]. Nesta técnica, factores de transcrição ligados à pluripotência são introduzidos e expressos em células diferenciadas, levando-as a reprogramar. Contudo, embora este seja um processo muito promissor, há ainda questões que têm de ser afinadas, nomeadamente a introdução destes factores através de retro/lentivirus e o facto de alguns deles estarem relacionados com tumores. Até à data já se registaram alterações relevantes para a segurança desta técnica de reprogramação e, em 2014, as células iPS foram já utilizadas no tratamento de uma paciente com degeneração macular. No entanto, à medida que as células adultas envelhecem, tornam-se cada vez mais difíceis de reprogramar.

Desta forma, este projecto explora o interessante mundo dos RNAs longos não codificantes (conhecidos na lingua Inglesa como long non-coding RNAs - lncRNAs) e a sua aplicação na reprogramação celular. Temos um interesse particular em transcritos antisense naturais (NAT), transcritos da cadeia oposta à cadeia codificante, visto que são menos estudados do que outros RNAs não codificantes e tem a capacidade de controlar o seu gene correspondente. Para tal, seleccionamos o lncRNA Zeb2NAT, o transcrito antisense natural (NAT) do Zeb2 (Beltran et al., 2008). O primeiro intrão do gene Zeb2 contém um sitio de entrada de ribossoma interno (IRES - internal ribosomal entry site) que sofre *splicing* na ausência do Zeb2NAT. No entanto, a presença deste transcrito parece de alguma forma garantir que este intrão não sofra *splicing*, levando a que a proteína Zeb2 seja transcrita correctamente. A proteína do Zeb2 está envolvida na *down-regulation* da E-caderina após a transição epitélio-mesênquima, um facto de grande importância dado a transição oposta - transição mesênquima-epitélio - ser um requisito inicial para a reprogramação celular de fibroblastos.

A hipótese deste trabalho consiste na afirmação de que a manipulação do lncRNA Zeb2NAT aumenta a reprogramação celular. Para a análise da reprogramação celular foram utilizados fibroblastos adultos e embrionários retirados de ratinhos transgénicos reprogramáveis com uma cassete activada pela doxiciclina, que inclui uma sequência de

quatro factores de transcrição ligados à pluripotência - Oct4, Sox2, Klf4 e c-Myc (Takahashi & Yamanaka, 2006).

Os nossos resultados mostraram que o knockdown do Zeb2NAT utilizando LNA GapmeRs específicos, uma tecnologia livre de virus, levou a um aumento no número de colónias de células iPS formadas, especialmente em células envelhecidas. Mostramos também uma linha temporal da reprogramação celular, usando células com diferentes passagens, para mostrar o efeito do envelhecimento celular. De momento estamos a fazer a caracterização das células iPS resultantes através de um ensaio funcional.

Demonstrámos pela primeira vez que a manipulação de um lncRNA NAT pode aumentar a reprogramação de células envelhecidas, destacando a possível contribuição deste trabalho para o melhoramento de futuras terapias baseadas em células reprogramadas especificas de cada paciente. Este trabalho poderá também ajudar a estabelecer modelos celulares de doenças ligadas ao envelhecimento ou de pacientes mais velhos.

Palavras-chave: lncRNAs, envelhecimento, reprogramação celular, iPSC, Zeb2NAT.

Abstract

Understanding the aging process is critical to prevent and counteract the effects of age-related decline of cells and tissues. Throughout time, due to the accumulation of genetic and epigenetic changes, stem cells tend to lose their capacity to replace damaged tissues. A major breakthrough advance was the identification of a combination of four transcription factors [34], which forced expression revert somatic cells to stem-like pluripotent cells (named cellular reprogramming). However, as differentiated cells grow old, they become more and more difficult to be converted back to pluripotency. This project delves into the interesting world of long non-coding RNAs (lncRNAs) and its implications in cellular reprogramming. We are particularly interested in natural antisense transcripts (NATs), transcribed from the opposite strand of coding genes, since they are less studied than other non-coding transcripts and have the capacity of controlling its corresponding protein-coding gene. We selected the lncRNA Zeb2NAT, the natural antisense transcript of Zeb2 (Beltran et al., 2008). It has been previously described that the presence of Zeb2NAT allows Zeb2 to be translated correctly. Zeb2 is involved in the down-regulation of E-cadherin after epithelial-to-mesenchymal transition, a fact of great importance since the opposite transition, mesenchymal-to-epithelial transition, is an early event during cellular reprogramming of fibroblasts.

Our hypothesis states that targeting Zeb2NAT enhances reprogramming of aged cells. For the reprogramming assay we used embryonic and adult fibroblasts of a transgenic mouse model carrying a doxycycline-inducible cassette with the four Yamanaka's factors (Takahashi & Yamanaka, 2006). Our results show that the highly-efficient knockdown of Zeb2NAT with specific and virus-free LNA GapmeRs led to a significant increase in the number of induced pluripotent stem cells (iPS cells) colonies formed, especially in aged cells. We also documented a reprogramming timeline of cells with different passages, showing the effect of cellular aging. At the moment we are characterizing the resulting iPS cell colonies to assess true stemness features.

We demonstrated for the first time that a lncRNA NAT can enhance cellular reprogramming of aged cells, highlighting the novelty and possible contribution of this work to improve the development of future stem cell-based therapies using patient-specific cells. This will also allow to make disease models from elderly patients and to optimize culture conditions to expand stem cells.

Keywords: lncRNAs, aging, cellular reprogramming, iPSC, Zeb2NAT.

Table of Contents

Chapter 3

This page was intentionally left blank

Chapter 1

This page was intentionally left blank

1. Stem Cells

1.1 Overview

Stem cells research has proven to be an exciting and fascinating area of biology. If we think about how individuals are formed and in the diversity of cells each individual encloses, we can immediately deduce the existence of cells with the potential to differentiate in all kinds of tissues. From an embryo to an adult, it all starts with one cell, the zygote. Throughout development, it starts proliferating and, at some point, differentiating into all cell types in a very orchestrated way. From this process raised the first thought about the existence of primordial cells. It appeared in the theory of cells by Schleiden and Schwann (Schleiden, 1839; Schwann, 1847), from around mid 1800s, stating that cells are units of life and that they are able to generate other cells. A century passed until the self-renewal capacity of pluripotent stem cells was discovered. Studies made in teratomas, back in 1954, led to the conclusion that pluripotent stem cells can give rise both to quickly differentiating and undifferentiated cells (Stevens & Little, 1954). It was the first time stem cells were described as we know them today: undifferentiated cells with both the ability of generating all types of cells of an adult organism and the capacity of self-renewal. Stemness is the word used to describe these two aspects of stem cells.

Stem cells exist all over the body. They are responsible for the maintenance and repair of tissues and organs when needed. Stem cells can be easily found in blood , skin, muscles, in the brain and even in the fat tissue (Blanpain & Fuchs, 2006; Till & McCulloch, 1961; Weissman, Anderson, & Gage, 2001). This discovery allowed scientists to think about new regenerative therapies, such as bone marrow transplant (Gatti et al. 1968, Bortin, 1970).

There are many types of stem cells and in essence they are classified by their plasticity (potency) or by the way they are obtained. The plasticity refers to the capacity of a stem cell to differentiate into different cell types. The zygote (and its subsequent divisions until the morula) is the only animal stem cell with the capacity of generating all the cell types that constitute an individual – it is a totipotent stem cell, which means that it can give rise to embryonic and extra-embryonic tissues. Embryonic stem cells (ES cells/ ESC) are the most potent cells after the zygote. They are extracted from the inner cell mass or epiblast of the mammalian blastocyst and so they are capable of generating all types of cells of the three primary germ layers: ectoderm, mesoderm and endoderm (except extra-embryonic tissues) (Evans & Kaufman, 1981; Martin, 1981). ESCs are one type of the so-called pluripotent stem cells. As the development takes place, these pluripotent stem cells start to follow a path of differentiation and become more lineage committed. Within tissues there are stem cells with less plasticity, named multipotent stem cells. Examples of these multipotent stem cells (or adult stem cells) are the mesenchymal stem cells. They are only able to differentiate into cells from the mesodermal lineage, like adipocytes, fibroblasts, myocytes, osteocytes and chondrocytes (Prockop, 2007). The cells with lowest plasticity are the unipotent stem cells. They can only differentiate along one lineage.

After the establishment of the first stem cell culture, from mouse teratomas, it was observed that when these cells were injected in mice, they would form new teratomas - tumors with tissue from the three germ layers (Kahan & Ephrussi, 1970; Rosenthal, Wishnow, & Sato, 1970) similar to the inner cell mass of the blastocyst. This pioneer experiment also showed that when these stem cells were passed in culture, the capacity to form teratomas decreased drastically, denoting that stem cells also experience the aging process (view ‘Aging and senescence’ section).

1.2 Cellular reprogramming

If we consider Waddington’s epigenetic landscape (Waddington, 1957) in a stem cell perspective, we can imagine the marble rolling down the mountains as a cell. The top ball is an ESC, that starts differentiating as it roles down the mountains, trough valleys, losing its plasticity along the way (i.e. the differentiation process). This represents what naturally occurs and the unidirectional road was a biological dogma until recently. But what if we could make this ‘ball’ go uphill or even transpose a hill

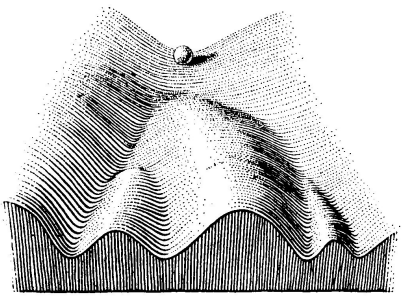

Figure 1: Waddington’s epigenetic landscape. From Waddington, 1957.

directly (Fig.1)? Known as transdifferentiation or direct reprogramming, this is a process in which a cell with a defined lineage, artificially turns into another cell type (as example a fibroblast into a cardiomyocyte (Ieda et al., 2010)). Back to the stem cell analogy, what if we can convert a differentiated cell back to pluripotency? This process is known as cellular reprogramming and does not occur naturally. Back in 1958, some years after Spemann showing the nucleus controls embryonic development (Spemann, 1938), John Gurdon was able to clone frogs by reprogramming the nuclei of adult frog cells after transferring them into frog eggs (Gurdon et al. 1958). This process is known as somatic cell nuclear transfer (SCNT) (Fig.2 A) and was the technology used to make Dolly the sheep (Campbell, McWhir, Ritchie, & Wilmut, 1996). The oocyte is the only cell able to revert the differentiated nuclei back to an embryonic state. SCNT can be used both for therapeutic and reproductive cloning, but has some limitations. Apart from being very inefficient, and that for itself is an important drawback, sometimes the nucleus do not reprogram correctly or completely, leading to different phenotypes. Besides, SCNT depends on very precise and expensive cell manipulations. In 2006, a new technology, named cellular reprogramming, was

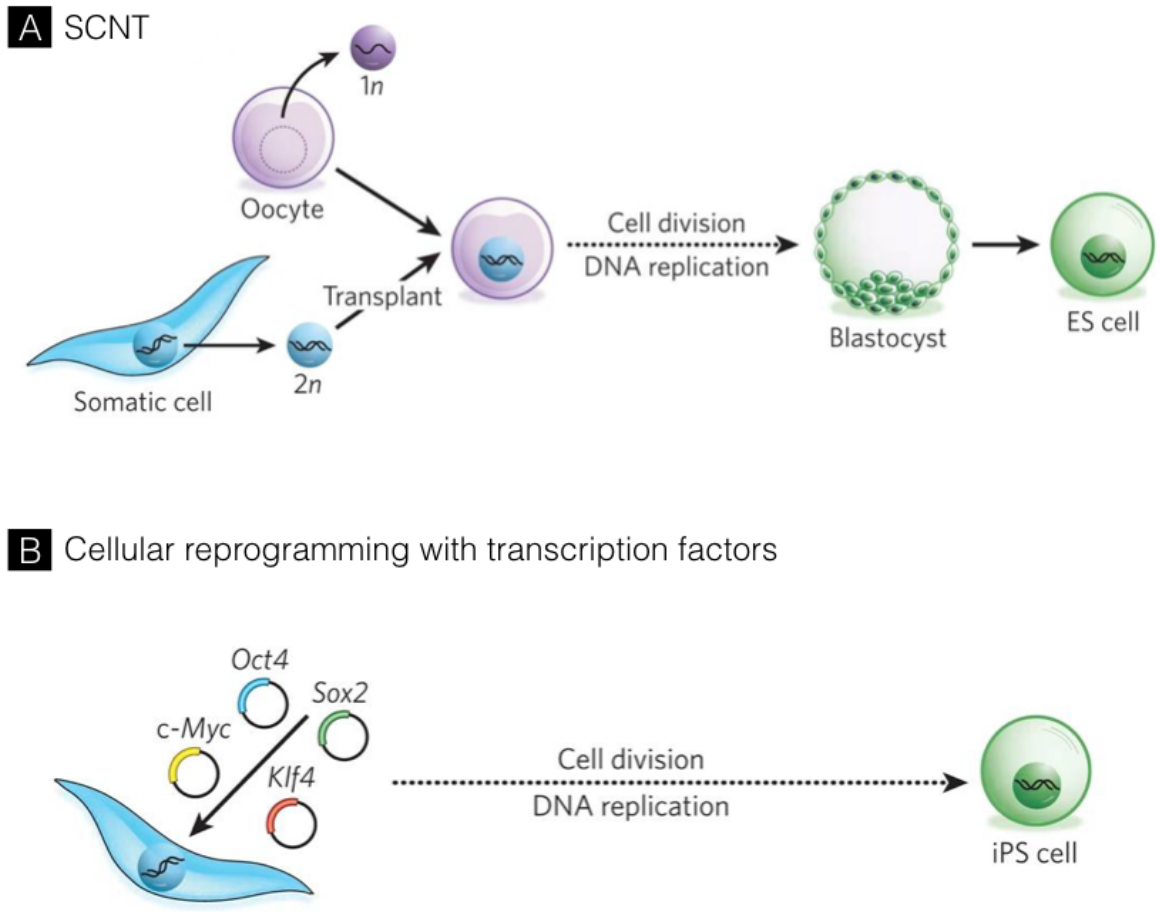

Figure 2: A Representation of somatic cell nuclear transfer, used for animal and therapeutic cloning; **B** Illustration of Yamanaka's cellular reprogramming using the ectopic expression of a certain cocktail of pluripotency transcription factors, the reprogrammed cells are called induced pluripotent stem cells (iPS cells). From Yamanaka, 2010.

described (Fig.2 B). Takahashi and Yamanaka found that some transcription factors were only active in ESC and not in differentiated cells. Using this knowledge, and following an extensive candidate approach, they showed that the ectopic expression of 4 out of 24 pre-selected transcription factors were able to convert differentiated somatic cells into induced pluripotent stem cells (iPS cells or iPSC) from mice (Takahashi & Yamanaka, 2006) and human cells (Takahashi et al., 2007) (Fig.2 B). These transcription factors are Oct3/4, Sox2, Klf4 and c-Myc (OSKM or Yamanaka factors) and were initially inserted in fibroblasts using retrovirus. The obtained iPSC are very similar to ESC in many aspects. The expression levels of many pluripotency markers, such as SSEA-1 and Oct4, are comparable. Both types of cells are pluripotent, with self-renewal capacities and with a similar morphology. As ESC, iPSC are also able to form teratomas and can contribute to chimeric mice embryos in germline transmission assays (Takahashi & Yamanaka, 2006). In 2012, Shinya Yamanaka and John Gurdon shared the Nobel Prize of Physiology or Medicine, because of their pioneer work in cellular reprogramming and somatic cell nuclear transfer, respectively.

1.3 The promise of induced Pluripotent Stem cells

Cellular reprogramming immediately enlightened the scientific community for countless biomedical applications. It gives us the exciting possibility to create patient-specific stem-like cells. These patient-specific reprogrammed cells - the so-called induced pluripotent stem cells (iPS cells or iPSC) - can then be differentiated into any type of cell or tissue to circumvent patient needs. They can also be either ‘corrected’ using genome editing technology such as CRISPR technology (Jinek et al., 2012) or used to make therapeutic drug screening (Robinton & Daley, 2012) and disease models.

Nevertheless, there are some problems associated with iPSCs reprogramming protocol. First, they can acquire a tumorigenic capacity. In fact, two of the Yamanaka factors are tumor-related and one is a tumor suppressor (Klf4). c-Myc is a proto-oncogene with a paradoxical function. It has an important role in mESC self-renewal (Cartwright et al., 2005), but in human it induces differentiation and apoptosis in hESC (Sumi et al. 2007). The second issue is the random and variable insertion of the OSKM by retrovirus. It has been described that each iPS colonies accommodated three to six integrations per factor (Takahashi et al., 2007) and that these retrovirus could be reactivated at any time (Okita et al., 2007). This urges to find new reprogramming cocktails and/or harmless ways for the manipulation of those transcription factors, especially especially if we consider the use of iPSC for human therapies.

During the last decade scientists have been fighting to solve these problems. Some even published fabricated results declaring the discovery of a new way to reprogram cells simply using a low-acid bath (Obokata, Sasai, et al., 2014; Obokata, Wakayama, et al., 2014). As the years passed, more reprogramming cocktails have been described. In 2008, it was shown that reprogramming can be achieved without c-Myc and Klf4, by the ectopic expression of Oct3/4, Sox2, Nanog and Lin28 in human fibroblasts (Tomioka et al., 2010), although Nanog has been previously described as a dispensable factor (Takahashi & Yamanaka, 2006). It has also been described that cells characterized by the endogenous expression of at least one reprogramming factor can reprogram by the induction of the other factors, excluding the one endogenously expressed (Utikal et al., 2009, Giorgetti et al., 2010). At the same time Kim et al. reprogrammed both mouse (J. B. Kim, Sebastiano, et al., 2009) and human neural stem cells (J. B. Kim, Greber, et al., 2009), that naturally express Sox2, only by over-expressing Oct3/4. To induce reprogramming using 'removable' episomal vectors seems one of the safest processes, since the vector is not integrated in the genome and does not request any viral infection (J. Yu et al., 2009), but its efficiency is a concern. There is also the possibility of using a combination of small molecules to reprogram cells (Hou et al., 2013), but this was never done with human cells (Lin & Wu, 2015). iPS technology is still evolving for a safer and better use in human therapies.

1.4 Aging and senescence

Understanding the aging process is crucial to prevent and counteract the age-associated decline of cells and tissues. Aging is a condition of life and is also linked with a vast number of diseases, namely cancer and Alzheimer's disease. Throughout time, due to the accumulation of genetic and epigenetic changes, stem cells tend to lose their capacity to replace damaged tissues. In culture, senescence (cellular aging) is aggravated with each passage, leading to decreased proliferation and differentiation or even cell death. Hayflick and Moorhead described this phenomenon for the first time in cultures of human fibroblasts (Hayflick & Moorhead, 1961). They defined three phases (Fig. 3): Phase I is a primary cell culture; Phase II regards the culture passages after phase I, where cells proliferate normally or are immortalized and multiply indefinitely; Phase III are senescent cells, characterized by absence of proliferation, changes in morphology and cell death. Senescence can be triggered by

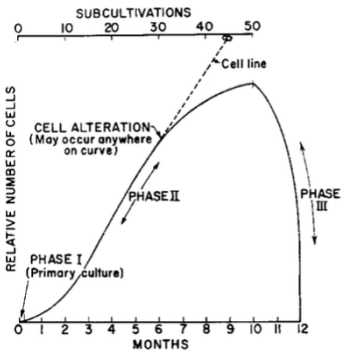

Figure 3: Representation of the three phases of senescence. From Hayflick and Moorhead, 1961.

endogenous or exogenous aspects such as telomere shortening or stress. As noted in previous sections, Evans and colleagues observed that when passed several times, ESCs ability to form teratomas decreased drastically (Evans & Kaufman, 1981). This was one of the first descriptions how aging could affect stem cells.

One of the biggest barriers to reprogram adult somatic cells is the barrier imposed by aging. Cellular reprogramming is a very inefficient process, especially in aged cells, the main target for therapies. Several studies were performed where reprogramming efficiency was compared between cells taken from groups of mice with different ages. Astonishingly, all younger groups reprogrammed significantly more efficiently than the older ones (Mahmoudi & Brunet, 2012), an observation favoring the idea that the capacity to reprogram is related to the division potential of a cell. Senescence was reported to decrease reprogramming's efficiency as a result of the up-regulation of p53 and others factors (Banito et al., 2009). Trying to counteract them will undoubtedly facilitate the process of reprogramming of aged cells. It has been shown that the manipulation of some proteins/ factors can counteract the effects of aging, enhancing reprogramming. A great example is the case of the down-regulation of the Ink4/Arf locus. This locus includes genes that encode the tumor suppressors p16, p19 and 15, that are up-regulated during organismal aging (H. Li et al., 2009). It has been shown than the down-regulation of this locus increases cellular reprogramming.

2. The ‘new’ RNA world

2.1 Overview

The Central Dogma of Molecular Biology states that a gene (DNA) could be transcribed as RNA and this RNA could then be translated into a protein (Crick, 1970). Assuming this, it was thought all complexity of eukaryotes was due only to the number of coding genes. In 2006, after several attempts (Lander, et al. 2001, Consortium et al. 2004), the sequence of the human genome was completed (Gregory et al., 2006). It allowed us to realize that the number of estimated human genes was quite lower than expected and only about 2% of the whole genome presented protein-coding regions. The fact that we have almost as many genes as smaller and less complex organisms such as a worm was overwhelming. Something more was accounting to the complexity distinguishing different organisms. A surprising from genome wide studies was the observation that most of the mammalian genome is transcribed (Birney et al., 2007; Carninci et al., 2005). How come only 2% of the genome can be directing the complexity of superior eukaryotes, and furthermore, why is the transcription machinery so pervasive transcribing regions without protein coding potential? The complexity could be partially answered with the presence of alternative splicing or post-translational modifications that will give rise to different isoforms of the same protein. This way, it is possible for organisms to have more different proteins than coding genes; however, these alone do not suffice. Transcription is a process that consumes a big amount of energy in a cell. Knowing that cells are very efficient in managing their energy, how come most of the genome is transcribed if it is apparently not needed? In 2012, the project ENCODE (de Souza, 2012) presented us a wider view of the genome, giving biochemical value to 80% of its non-coding part, the so-called junk DNA. It also contributed with new databases for the study of these non-coding transcripts (Bu et al., 2012).

2.2 Non-coding RNAs

Back in 1961, Jacob and Monod (Jacob & Monod, 1961) inferred for the first time the existence of RNAs capable of repressing the production of proteins. This is one of the functions of what we now call a non-coding RNA (ncRNA). Contrary to messenger RNA (mRNA), ncRNAs do not code for a protein nor are translated into one (Huttenhofer et al. 2005). There are several types of ncRNAs and its classification is still debatable and sometimes depends on the author (Fig.4). They were first discovered in 1965 (Holley et al., 1965). Starting by considering the difference between structural - housekeeping - and regulatory ncRNAs (Figure 4). Structural ncRNAs are responsible for normal and basic cell function. A good example is the use of non-coding transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) in transcription and translation. Following the same line, non-coding RNAs like telomerase RNA component (TERC), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) are also structural.

Figure 4: Classification of ncRNAs.

Regulatory ncRNAs are generally classified by size. Short ncRNAs (~18-200 nucleotides) are present in the cytoplasm and include small interference RNAs (siRNAs), Piwi interacting RNAs and the well-studied microRNAs (miRNAs). Long non-coding RNAs (lncRNAs) are mainly present in the nucleus. By definition they have more than 200 nucleotides and are less conserved than the miRNAs or the coding genes. At first, lncRNAs were described as junk non-functional species, a result from transcriptional noise (Struhl, 2007). Further and more recent studies surprisingly linked lncRNAs with key-regulatory processes of gene expression, acting both near the local of where the lncRNA was transcribed - in cis - and/or in other regions of the genome - in trans (Wilusz et al. 2009). They are involved in development (Blackshaw et al., 2004; Grote et al., 2013), epigenetic regulation (Mercer & Mattick, 2013), transcriptional interference, regulation of alternative splicing, imprinting, cancer (Wilusz et al., 2009) and some can be cleaved by Dicer and Drosha to form microRNAs (Cai et al. 2004; Lee et al., 2004). lncRNAs have also been described to control stemness and differentiation (Ghosal, Das, & Chakrabarti, 2013; Guttman et al., 2009; Guttman et al., 2011), and to be 'specific to a given species' (Wang et al., 2004). Welcome to the 'new' RNA world.

2.3 Long non-coding RNAs transcription and processing

Long non-coding RNAs represent the most transcribed ncRNAs in the genome, being some well conserved (Guttman et al., 2009). Nonetheless, they are the least studied and functionally described, being considered the genome's 'dark matter'. lncRNAs are mostly transcribed by RNA polymerase II (RNAP II) and can be polyadenylated. These transcripts can be stratified into many sub-types and are classified regarding the position they are transcribed when compared to the coding sequence. Following that logic, sense lncRNAs are transcribed from the same strand as the coding gene. lncRNAs can also be transcribed from the intron of a coding gene - intronic lncRNAs - and between genes - intergenic or long intergenic non-coding RNAs (lincRNAs). The least studied lncRNAs are the Natural Antisense Transcripts (NATs). Many of these NATs were found near known protein-coding genes (Yelin et al., 2003) and they are transcribed from the opposite strand (Fig.5). Since they overlap part of the coding sequence, NATs have been described to control the expression of its sense gene either by decreasing (Beltran et al., 2015; Hawkins & Morris, 2010; Katayama et al., 2005) or promoting it (Beltran et al., 2008). NATs have many functions and can act in cis or in trans (Hawkins & Morris, 2010). They can also act as epigenetic silencers (W. Yu et al., 2008) and chromatin modulators (Khalil et al., 2009).

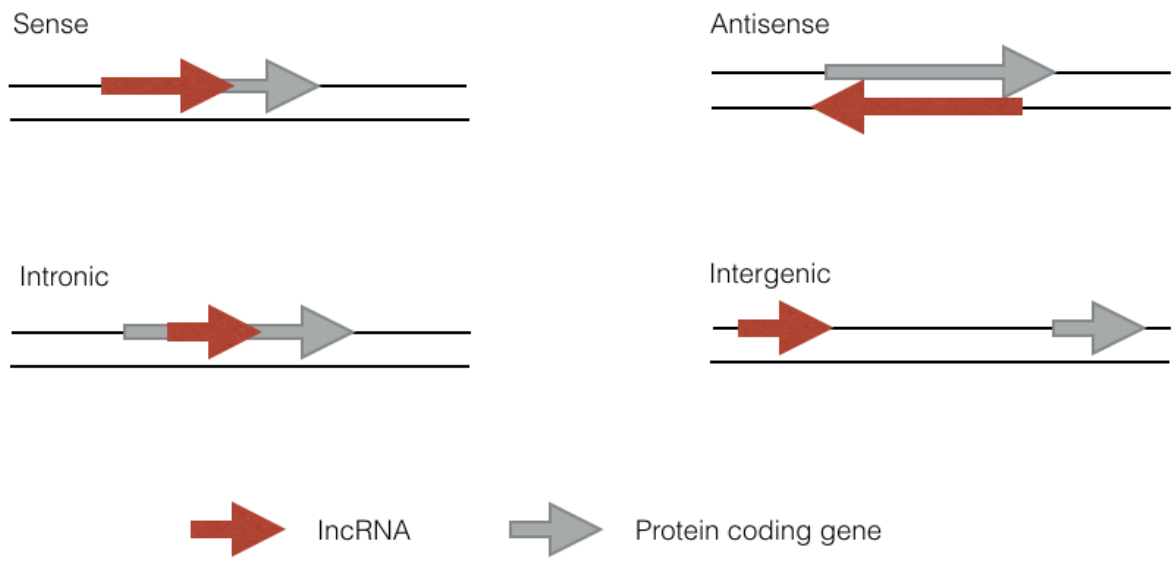

Figure 5: Illustration of four different types of lncRNAs: Sense, antisense (NAT), intronic and intergenic (lincRNAs).

2.4 LncRNAs in stemness and cellular reprogramming.

As seen in previous sections, lncRNAs are transcripts with very diverse functions. Hundreds of these transcripts are only expressed (D. H. Kim et al., 2015) or only repressed during cellular reprogramming (Beltran et al., 2008; Dinger et al., 2008). By controlling pluripotency factors like Oct4 and Nanog (Sheik Mohamed et al. 2010), some lncRNAs could be important players of the reprogramming process.

In this project we were particularly interested in a novel class of long non-coding RNAs, the antisense transcripts. As a model we focused on the Zeb2 (coding) / Zeb2NAT (lncRNA) locus. Zeb2NAT lncRNA is the natural antisense transcript (NAT) of the promoter region and the transcription start site (TSS) of Zeb2 (Zinc finger E-box-binding homeobox 2, also known as Smad Interacting Protein 1 - SIP1) (Beltran et al., 2008) and presents a good conservation between human and mouse (Nelles et al. 2003). This lncRNA controls the splicing of the first intron of Zeb2, by a mechanism not yet described, allowing the maintenance of this intron and the correct translation of the protein Zeb2 (Figure 6). Zeb2's protein is a transcriptional repressor of E-cadherin which is unregulated after epithelial-mesenchymal transition (EMT) (Vandewalle et al., 2005), allowing the maintenance of a mesenchymal phenotype. It has been shown that the opposite transition - mesenchymal-epithelial transition (MET) is one of the first steps required for fibroblasts to reprogram (R. Li et al., 2010). Another requirement for reprogramming is E-cadherin (R. Li et al., 2010), also linked with pluripotency of ESC (Chou et al., 2008).

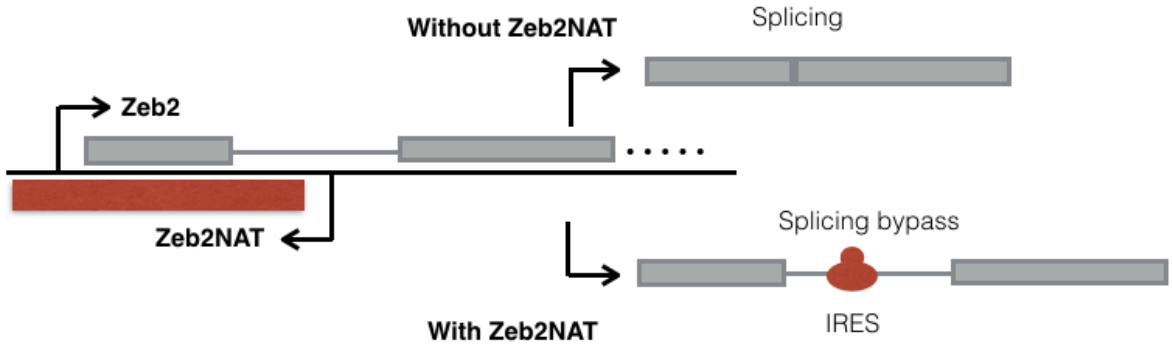

Figure 6: Representation of the interaction between Zeb2NAT and Zeb2. Zeb2NAT allows the maintenance of the intron of Ze2, which contains an IRES. From Beltran, 2008.

3. Objectives

The present work links aging with the exotic new world of long non-coding RNAs and the process of cellular reprogramming. Similarly to coding transcripts, we hypothesize that modulation of lncRNAs would affect the reprogramming process. In our particular case we were interested how the knockdown of Zeb2NAT would enhance reprogramming. Zeb2NAT, the natural antisense transcript of Zeb2, was previously described to be controlling the translation of Zeb2 (Beltran et al., 2008). Since Zeb2 is linked with the down-regulation of E-cadherin after epithelial mesenchymal transition, we wondered if the enhancement of the opposite transition - by the manipulation of Zeb2NAT - could counteract the dramatic effects of aging and senescence in cellular reprogramming. Besides, and in agreement with our hypothesis, previous unpublished results from our lab showed us that the knockdown of Zeb2NAT enhanced stemness of mESC. We used a novel knockdown strategy - LNA long GapmeRs - in order to achieve a highly-efficient targeting of our selected lncRNA. For reprogramming, we used cells from mice with a doxycycline-activated cassette with the four Yamanaka's factors.

Our main objectives were to:

- Establish the difference of reprogramming efficiencies between our embryonic and adult fibroblasts;
- Check the effectiveness of the knockdown of the lncRNA Zeb2NAT using LNA GapmeRs;
- Analyze mRNA and protein expression levels to ascertain if the knockdown of Zeb2NAT decreases the translation of Zeb2 and not the other way around;

- Analyze if the knockdown of Zeb2NAT, while inducing reprogramming of fibroblasts with different ages and passages, would enhance the reprogramming efficiency (by counting the number of iPS colonies formed);

- Characterize and test out the properties of our induced pluripotent stem cells.

I had a major role in the analysis and experimental work presented in this thesis (part of the experiments contained in Chapter 2/2.1 were performed by Bruno Jesus).

The funding for the work presented here was granted by Fundação para a Ciência e Tecnolocia, AXA and and Fundação AstraZeneca.

Chapter 2

This page intentionally left blank

1. Results and discussion

1.1 Expression levels of Zeb2NAT and Zeb2 in different cell types.

We started by measuring the expression levels of the long non-coding RNA Zeb2NAT in different types of cells, followed by Zeb2's first intron and one of Zeb2 exons. It was known from previous works that Zeb2 was up-regulated in mesenchymal cells and down-regulated in epithelium (Beltran et al., 2008). Since Zeb2NAT seems to be controlling Zeb2's correct translation, this transcript must follow the same patterns of Zeb2 i.e. highly expressed in mesenchyme and down-regulated in epithelial cells (Fig.7). Here, mesenchymal cells are represented by mouse embryonic fibroblasts (MEFs) and epithelial cells by ESC and iPSC. Expression of Zeb2, Zeb2NAT and Zeb2's intron are significantly increased in MEFs when comparing to ES and iPS cells. Zeb2NAT controls the splicing of the first intron of Zeb2. This intron has an internal ribosomal entry site (IRES). In the absence of Zeb2NAT this intron is spliced and not translated. However, in the presence of this NAT, the splicing of this intron do not occur, allowing the complete translation of the protein Zeb2. No significant difference was found between ESC and iPSC.

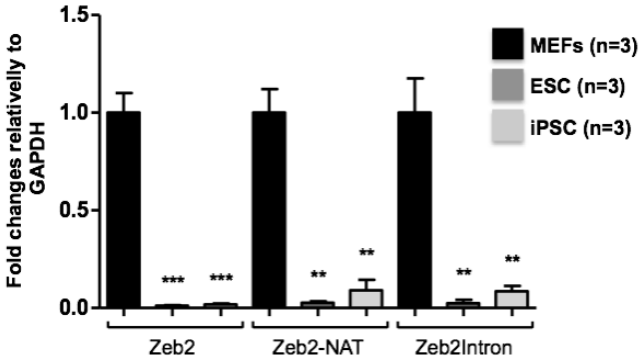

Figure 7: Fold change of the levels of expression of Zeb2, Zeb2NAT and Zeb2 intron were checked in MEFs, ESC and iPSC using qRT-PCR with specific primers. Normalized to MEFs in each case. Average and s.d. with unpaired two-tailed Student's t test (n=2) ** P<0,01, *** P<0,001.

1.2 Aging as a reprogramming barrier

As referred in previous sections, it is known that reprogramming efficiency decreases with aging. To confirm the impact of aging in the reprogramming efficiency of our cells, MEFs (E13,5) and fibroblasts (from 10 weeks old adult mouse) were compared. We were able to observe that two fold more iPS colonies (AP+ colonies) were present when MEFs were reprogrammed comparing to adult fibroblasts, using the same conditions (Figure 8). This comparison was made in cells from passage 3 (P3).

In the next sections we will address the severe effect of passages and freezing/thawing in cellular reprogramming, something previously reported in the literature to affect the reprogramming efficiency. Given that our hypothesis states that the knockdown of Zeb2NAT increases cellular reprogramming, we also wondered if this could be also validated in these two following situations: reverting reprogramming capacity of previously frozen fibroblasts and enhance reprogramming in p3 or more by manipulating Zeb2NAT (in the next sections).

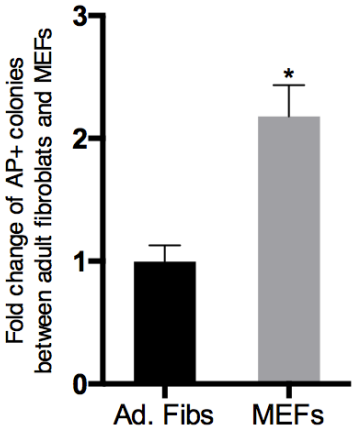

Figure 8: Fold change between reprogramming efficiency of mouse adult and embryonic fibroblasts (P3; n=2). Average and s.d. with unpaired two-tailed Student's t test. *P<0,05.

1.3 Testing LNA GapmeRs' knockdown efficiency

We performed the knockdown of Zeb2NAT and Zeb2 using LNA long GapmeRs. As observed in Figure 9 A and B, both knockdown's were highly efficient (p<0,01). This experiment also allowed us to confirm that, by doing the knockdown of Zeb2NAT, we indeed decreased the expression of Zeb2. The western blot performed goes in the same direction as the qRT-PCR. Targeting Zeb2NAT and Zeb2 decreases the production of the protein Zeb2. Also as expected, targeting Zeb2 did not affect significantly the expression on the lncRNA Zeb2NAT. The maintenance of Zeb2's first intron in the presence and absence of Zeb2NAT was also tested, but it turned out very variable and then, not reported here as further tests will be performed. We would expect that the transcription of this intron would decreased dramatically as we target Zeb2NAT. Although not visible or addressed here, a mechanism of compensation of Zeb2 may exist when we knockdown Zeb2NAT. That could be tested by looking at the nascent RNA. The real interaction of Zeb2NAT controlling the translation of Zeb2 is not very well described. Additionally, Zeb2NAT could be possibly regulating the expression of Zeb2 by a different mechanism, such as direct binding to the

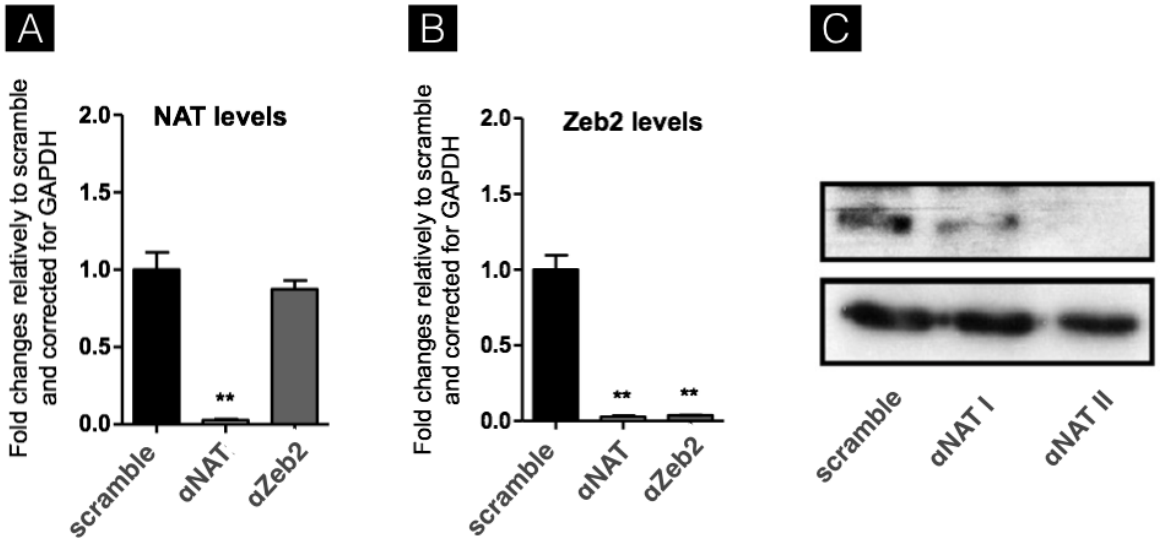

Figure 9: A) Fold change of the expression levels of Zeb2NAT between each treatment. **B)** Fold change of the expression levels of Zeb2 between each treatment. Average and s.d. with unpaired two-tailed Student's t test (n=2). ** P<0,01 **C)** Western blot showing Zeb2's decreased levels, caused by the knockdown of Zeb2NAT.

Zeb2 promoter. In order to look to what happens to Zeb2NAT, a Northern Blot should be performed.

1.4 Performing the knockdown of Zeb2NAT while inducing cellular reprogramming

Our hypothesis states that the knockdown of Zeb2NAT enhances cellular reprogramming. To test it we performed the knockdown of Zeb2NAT while inducing cellular reprogramming. Our experimental design is depicted in Fig. 10 A and consists in transfecting fibroblasts with LNA long GapmeRs (control (scramble), αNAT and αZeb2) in the first two days of the experiment, followed by the induction of cellular reprogramming at day 2. Using cells from i4F transgenic mice with a doxycycline (dox) inducible cassette containing OSKM (Fig.10 D), reprogramming was induced simply by the addition of dox to cell medium. Dox was introduced at day 2 to assure that the knockdown caused by the transfections was already happening. Otherwise, cells would start reprogramming as if they had no treatment.

To check if 'scramble', the LNA GapmeR negative control, was a good control, we compared it to cells with Lipofectamine - the transfection reagent - in the medium and cells with only medium (Fig.10 C). As expected, there was no significantly difference from the other two conditions. The scramble LNA GapmeR has no homology with any known sequence and that it has the same structure as other GapmeRs (Fig.10 B).

Instead of comparing the efficiency of reprogramming in percentage of cells reprogrammed, we decided to use a different approach by comparing the number of iPS

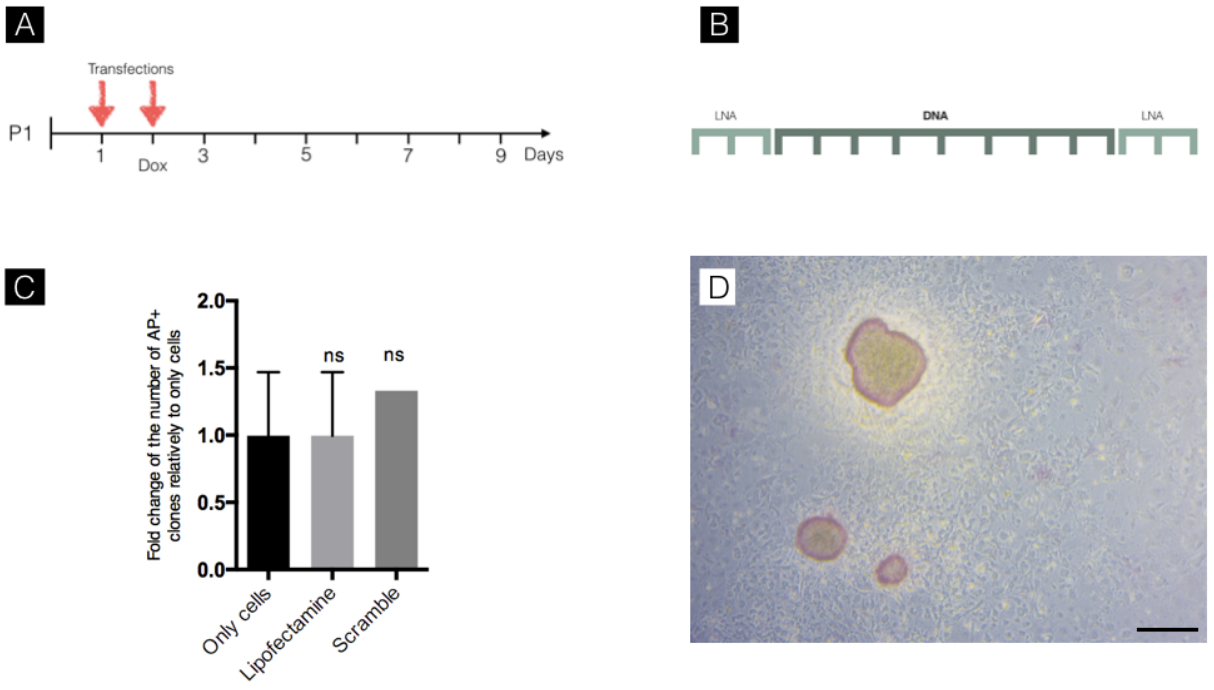

Figure 10: A: Diagram of the experimental design. **B:** Representation of a LNA GapmeRs: a strand of DNA specific for its target and two LNA ends conferring nuclease resistance.. **C:** Control analysis. Average and s.d. with unpaired two-tailed Student's t test (n=2). ns= non-significant. **D:** AP+ colonies (pink staining). Scale bar = 200µm.

cell colonies formed in each situation. To check if we could consider every appearing colony in our counting, we performed several alkaline phosphatase (AP) stainings. Every colony formed was positive for AP (Fig.10 D).

As mentioned before, there are many factors that can influence cellular reprogramming negatively. We wanted to compare reprogramming efficiencies of cells taken from mice with different ages, but we also took into account the different number of passages those cells had. This should always be done with caution, otherwise we would be making invalid comparisons without considering the dramatic effect of senescence. All reprogramming experiments were done with the same number of cells -50 000 per well of 6 well plate- and with a confluency of approximately 75%.

Using fresh fibroblasts from our transgenic mice (Fig.11 A), we compared the outcome of cellular reprogramming while down-regulating Zeb2NAT and Zeb2, in different passages of these cells. At passage one (P1), cells started changing morphology as early as day 5 (Fig. 11 C/D) and the first iPSC colonies were formed by the end of the first week (Fig11 E/F). Changes in morphology were more accentuated in αNAT (Fig.11 D) than in the control (Fig. 11 C). By the end of the first week, a difference in the number of colonies was already visible between treatments (Fig.11 G/F) and initially it was observed that αNAT colonies were more defined than scramble. αNAT had two times more colonies formed than the control (Fig.11 G/H). At week 3, this difference was more accentuated: both αNAT and

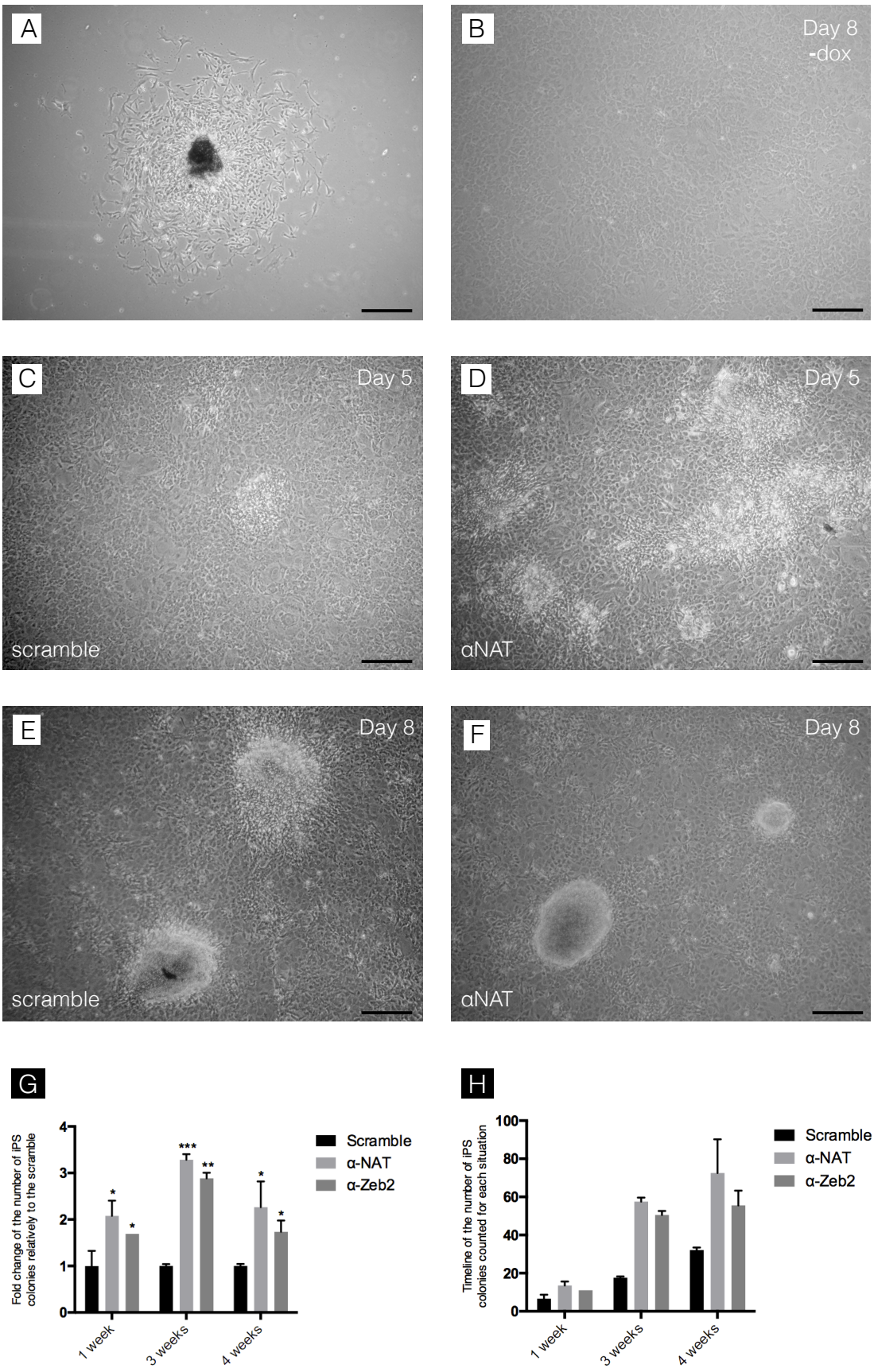

Figure 11: A: Fibroblasts migrating from an explant of mouse ear during the establishment of fresh fibroblast cultures. **B:** Doxycycline control (day 8). Dox was not added to the medium - no colonies were formed. **C:** Reprogramming's initial change of morphology at day 5 (scramble). **D:** Reprogramming's initial change of morphology at day 5 (αZeb2NAT). **E:** Scramble first colonies (day 8). **F:** αZeb2NAT first colonies (day 8), initially more defined than the control. **G:** P1 reprogramming fold change timeline. Normalized and statistically compared with each respective control (scramble) (n=2). Average and s.d. and unpaired one-tailed t test; *P<0,05, **P<0,01, ***P<0,001. **H:** Number of iPS colonies counted (n=2) in different times. Scale bars = 200μm

αZeb2 were forming around three times more colonies than the control (Fig.11 G/H). Week 3 represented the peak difference, in which αNAT formed a mean of 57,5 colonies, comparing to 17,5 in the control (Fig. 11 H).

The same experiment was performed in cells from passage 3 (P3). When P3 fibroblasts were plated, they did not show any visible signs of senescence. We first observed that the morphology of the fibroblasts started to change at the end of the second week and in a much lower number than P1 (Fig.12 A/B). As in P1, an increase in cellular reprogramming was also visible when targeting Zeb2NAT and Zeb2. The number of colonies of αZeb2 at 4 weeks was higher than the αNAT with an increased variability between the two replicates (Fig.12 B). Also, colonies were smaller and took more time to grow than αNAT and controls (not shown). Again, '3 weeks' represented the peak difference, this time αNAT formed 12 times more colonies than the control (scramble) (Fig.12 A), even though P3 globally gave rise to fewer colonies than P1 (Fig.12 C). The diferences between each treatment at passage 1 and 3 is significantly different in all treatments (scramble, αNAT and αZeb2), after 3 weeks

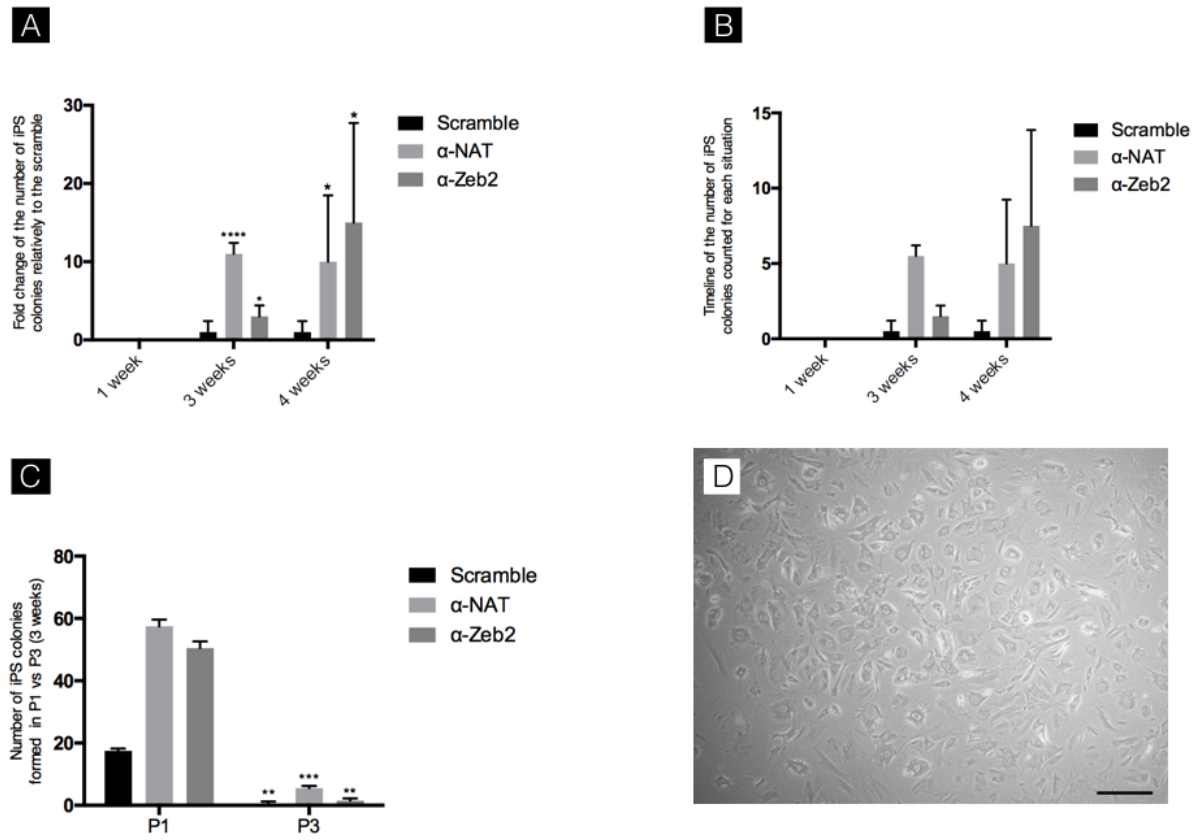

Figure 12: A: Timeline of P3 reprogramming fold change. Normalized and statistically compared with each respective control (scramble) (n=2). Average and s.d. and unpaired one-tailed t test. **B:** Number os iPS colonies counted in different times (n=2). **C:** Contrast between P1 and P3 reprogramming after 3 weeks. The statistical analysis was performed between each same condition in different passages. Average and s.d. and unpaired two-tailed t test (n=2). **D:** P6 fibroblasts with senescent morphology. *P<0,05, **P<0,01, ***P<0,001,****P<0,0001. Scale bar = 200μm

At passage 6 (P6), fibroblasts already displayed a senescent morphology (Fig.12 D). Cells looked bigger and spread and they practically and did not multiply. In this case, cellular reprogramming was not achieved. No reprogramming morphological changes were detected after one month.

The same experiment was performed in MEFs (P3) from our transgenic mice. This time we only did the knockdown of Zeb2NAT. There was a significantly difference between the number of colonies formed in αNAT than in the controls. At this time, this difference was six times higher than all the controls (Fig.13 A).

During the process of optimization of cellular reprogramming in our lab, we faced some problems with previously frozen cells. All the cells from our transgenic mice line that were stored in liquid nitrogen lost their capacity to reprogram. So, knowing that stress can constitute a barrier for cellular reprogramming, we decided to see if we could revert the reprogramming capacity of this cells. We used the above referenced experiment, in which we knockdown Zeb2NAT in P3 MEFs and adult fibroblasts previously stored in liquid nitrogen. As depicted in Fig.13 B, we were able to restore reprogramming capacity of MEFs by the manipulation of Zeb2NAT (n=2). However, no adult fibroblasts were able to reprogram after freezing/thawing.

Additionally, we conclude that starting with cells with a low confluency,, regardless (lass than 50%) the number of the passage or age, will not allow cells to reprogram (data not shown).

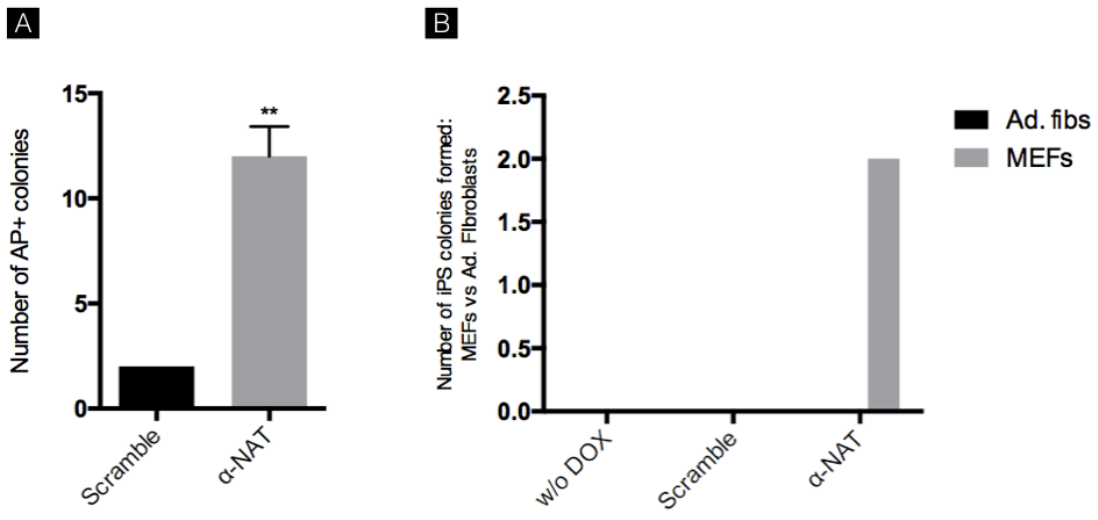

Figure 13: A: Number of AP+ colonies from MEFs P3 after two weeks. Average and s.d. and unpaired one-tailed t test (n=2), **P<0,01. **B:** Number of AP+ colonies after freezing/thawing of MEFs and adult fibroblasts, both P3. Only MEFs with the knockdown of Zeb2NAT were able to reprogram.

Dox was introduced at day 2 to assure the knockdown was already happening when inducing reprogramming. While optimizing the process, we checked what happened if we induced reprogramming before the knockdown of Zeb2NAT and we saw no difference in the quantity of colonies. In this case the absence of Zeb2NAT would possibly only being enhancing the stemness (the quality not quantity) of the cells that end up reprogramming, accordingly to previous results in the lab which showed that the knockdown of Zeb2NAT enhances stemness [unpublished].

It was very clear that targeting Zeb2NAT enhances reprogramming. Every time that we performed this experiment we got an increased amount of colonies in αNAT when comparing to the scramble. We checked the controls thoroughly to make sure we were indeed observing the effect of the knockdown of Zeb2NAT and Zeb2 and not the effect of the manufacturer's control (LNA GapmeR Negative control) acting in any way in cellular reprogramming (i.e. decreasing reprogramming). We consider that the appearance of colonies follow a Normal distribution. The experiments were performed in two replicates, but if we increased the number of replicates, the number of colonies would be very similar between each situation and tend to a number (mean).

This experiment not only allowed us to test our hypothesis, but also gave us the opportunity of documenting a timeline of cellular reprogramming with the variables of organismal aging and/or cellular aging. We would like to make more experiments with MEFs, namely reprogramming/knockdown of Zeb2NAT in P1 and P5/P6 cells, but that was not logistically possible. Regarding the outcome of reprogramming of MEFs and adult fibroblasts, we can only compare P3. As expected MEFs formed more iPS colonies and were able to reprogram even after freezing/thawing. Comparing to adult fibroblasts, MEFs are more undifferentiated cells with less epigenetic changes, which allow them to reprogram efficiently.

Now comparing the results between different passages of adult fibroblasts. Although there was significantly more colonies formed in P1 than in P3, the greatest difference between the proportion of colonies formed while knocking-down Zeb2NAT vs. scramble were in P3, with an impressive 12 fold increase (Fig.12 A). This difference in reprogramming seems to be increasing with the cellular aging. It seems that by performing the knockdown of Zeb2NAT we could rescue the reprograming capacity, especially of in aged cells. Nonetheless, the difference in the number of iPS cell clones between P1 and P3 after 3 weeks is very significant (Fig.13 C), showing, again, the severe effects of cellular aging at P3.

The knockdown of Zeb2 has also interesting results. It looks like that αZeb2 has a reprogramming delay of some days when comparing to αNAT (Fig.11 C/D). At P1, αZeb2 never reaches the same result as αNAT. At P3 there is a big leap from '3 weeks' to '4 weeks'

in which αZeb2 passes αNAT, yet there was a lot of variability in its replicates and the colonies, in this case, were very smaller than αNAT. Apart from this result, one could say that the down-regulation of Zeb2 is less effective in P3 than in P1 and otherwise for αNAT. We conclude that targeting Zeb2 nat is not as efective as targeting its lncRNA (Zeb2NAT).

1.5 Testing out the properties of our induced pluripotent stem cells.

Apart from the counting the colonies in each situation, we tried to ascertain if the colonies (αNAT, αZeb2 and scramble) would be able to grow and proliferate similarly under 2i conditions. 2i consists in two inhibitors - PD0325901 and CHIR99021 - of two pathways - MEK and GSK3, respectively - allowing stem cells to maintain stemness in culture. Media supplemented with 2i gives an advantage to the growth of undifferentiated cells like ES and iPS. Since dox is the activator of the transgene with the reprogramming factors, at this step it was removed to assure the cells were only expressing pluripotency factors endogenously. No other transfections were made at this stage.

For this experiment three colonies of scramble, αNAT and αZeb2 were picked and their cells were separated mechanically and cultured in each well (96 well plate) directly under 2i conditions. This way we could see if the cells we put under these conditions were already sufficiently undifferentiated and expressing by themselves the needed factors in order to survive and to form new colonies of iPSC. Also, we wanted to examine if any of the different situations (αNAT and αZeb2) would behave differently from the controls. At day 0 some fibroblasts were still visible, along with some aggregates and small round cells. colonies are

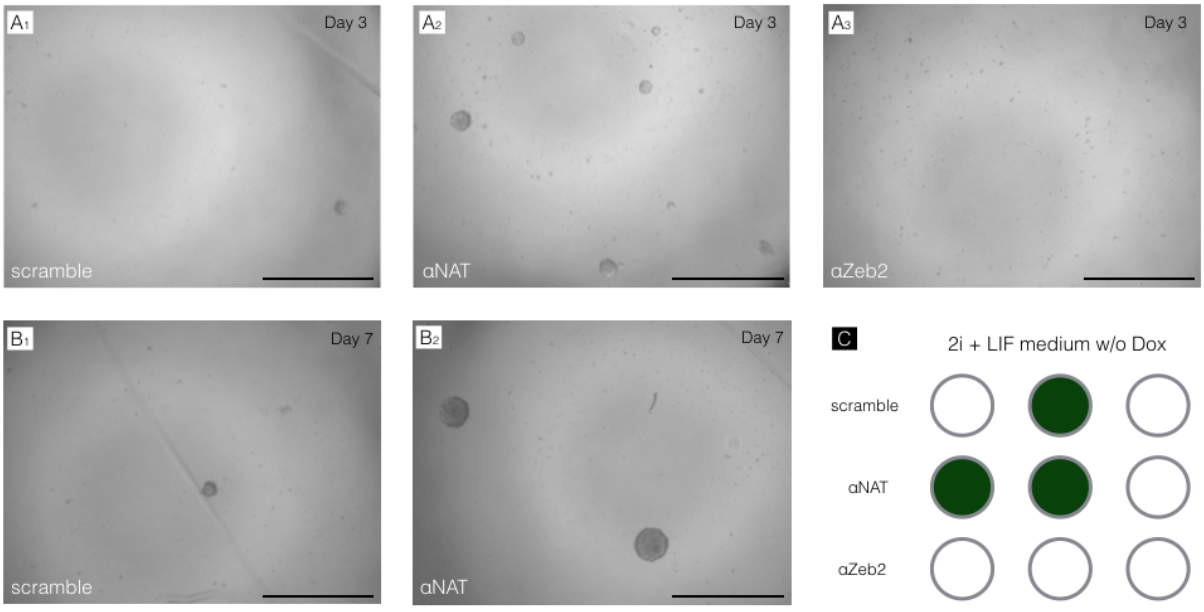

Figure 14: **A:** Three days after disaggregated colonies were passed to ESC medium supplemented with 2i. **B:** Eight days after disaggregated colonies were passed to ESC medium supplemented with 2i. **C:** Green represents the situations in which colonies survived. Scale barrs = 200µm

surrounded by fibroblasts, so picking them also means we will also pick some fibroblasts. At day 3, lots of cells died and some colonies were already visible. As the cells growth, some colonies and cells end up disappearing. Most of them should have been in an intermediate state, which were not sufficient to withstand 2i conditions for too long. At day 7 some colonies were still growing. Not all the colonies we picked were able to form new colonies. Figure 14 C shows how many of the three colonies picked for each situation ended up forming new ones. It was visible that more αNAT colonies were able to form new colonies than the control (scramble) and αZeb2 end up not forming a single colony. Also, we were able to see that more colonies were forming in aNAT than the controls, implying more αNAT cells survived 2i. This could mean that more cells were pluripotent with self-renewal capabilities, possibly representing true iPSC.

Despite the fact that all the colonies picked had the same morphology and size and that we can see a small pattern in this experiment, there might be some issues with the mechanical separation of cells that could be influencing. It is acknowledged that some cells are lost during the process and we did not performed a cell counting to assure we put the same number of cells in each well, because of the small size of the colonies. But we confirmed that all the wells had cells and aggregates in it and there were a similar number in each one of them after separating the cells from the first colonies. Nevertheless, it does not explain why no cells survived 2i in aZeb2. This experiment should be repeated with a greater number of replicates to obtain a stronger pattern. For this experiment we decided not to use cells from the first passage. We thought it might be easier to observe differences when cells were already presenting some barriers limiting reprogramming.

The same assay was performed, but this time we cultured the dissociated colonies with a layer of feeder cells (MEFs), also without doxycycline. In this experiment we could not observe a difference between any of the different conditions (αNAT, αZeb2 and control

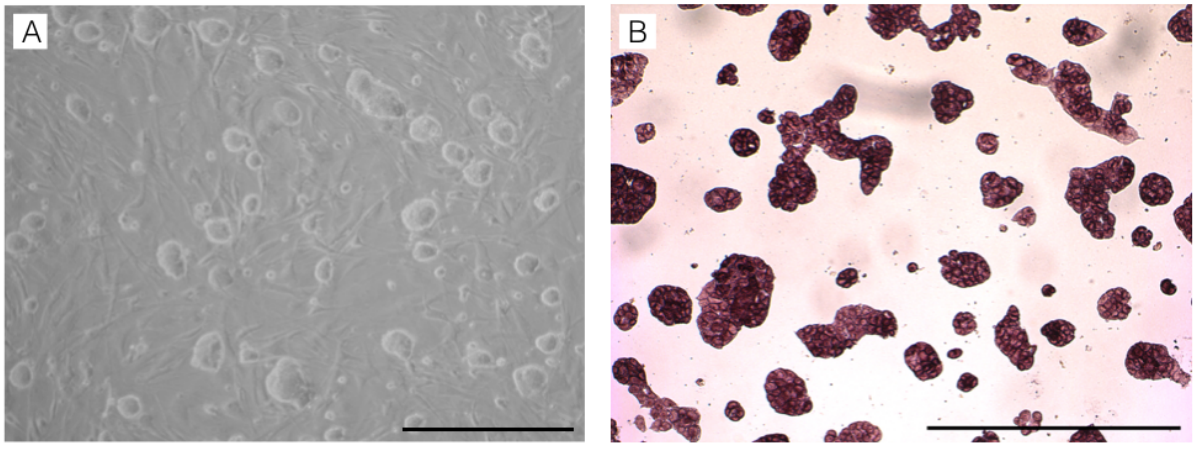

Figure 15: A: Isolated iPSC lines growing over a feeder layer. **B:** Isolated iPS cell lines grown in ESC medium with 2i expressing marked for alkaline phosphatase activity (pink = positive). Scale bars = 200µm

(scramble)) (Fig.15 A). Cells were very proliferative, they all formed new colonies in two days, as normally happens with ESC. We did this process several times, both by mechanically separating cells or using EDTA-trypsin. These cells were passed many times using trypsin and were adapted to 2i conditions with change of medium plus 2i in successive passages (Fig.15 B). They also resisted the freezing storage process and were positive for alkaline phosphatase (AP staining), a pluripotency marker for embryonic stem cells.

Given the amount of variables in the 2i experiment we only want to consider that passing the iPS cell colonies to feeders take less time for them to form new colonies, than when passed to 2i medium. Since there was no visible difference between using mechanical separation of the colonies or using trypsin when cells were seeded in a feeder layer, we can take that variable off the table, not considering this as what caused the delayed proliferation with 2i. Also, after adapting iPSC from feeders to 2i, cells proliferated much rapidly comparing to the cells passed directly to 2i, so we can consider that those cells could be near the iPSC state, but not yet there. Furthermore, the feeder layer is way less aggressive to cells, than 2i medium is, especially when cells are not fully pluripotent.

1.6 Characterization of our iPS cells: assessment of stemness.

After isolating iPSC lines (LNA and shRNA) and expanding them, we looked for the presence of specific pluripotency markers using immunocytochemistry, a technique used to visualize the location of the desired proteins in cells. Cells were stained with DAPI and

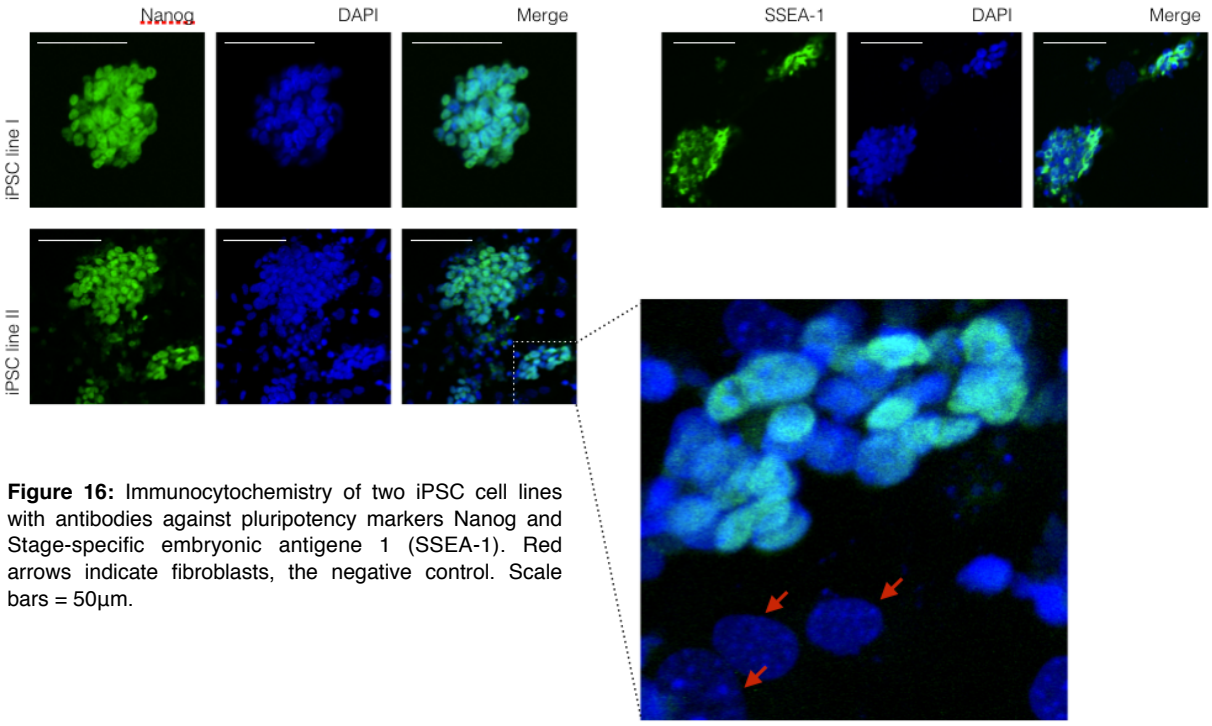

Figure 16: Immunocytochemistry of two iPSC cell lines with antibodies against pluripotency markers Nanog and Stage-specific embryonic antigene 1 (SSEA-1). Red arrows indicate fibroblasts, the negative control. Scale bars = 50µm.

antibodies against Nanog and SSEA-1. All iPSC lines expressed these pluripotency markers, comparing to negative controls (Fig.16). No substantial difference was observed between lines. This experiment was performed to assess stemness and not to compare stemness between each line. We expected to observe no substantial difference between lines, because at this stage all our colonies presented a consolidated stemness state. As a result, all cell lines should have the same potency and at this time the LNA GapmeRs would not be acting. To compare stemness between each line we would use the same number of cells.

Although our iPSC express AP, Nanog and SSEA-1, true stemness features can only be assessed by a teratoma formation assay and/or germ-line transmission. We are at the moment performing a teratoma formation assay with our iPSC lines.

2. Methods

Real-time quantitative PCR (RT-qPCR)

Total RNA extraction was performed using PureZOL RNA Isolation Reagent, following the instructions given by the manufacturer (Bio-Rad). Samples were treated with DNase, to avoid DNA contaminations, followed by reverse transcription using random primers (Roche). qPCR consists in quantifying the target DNA while it is being amplified. Here we used SYBRGreen, a reporter of the primers that links to the target, right after the denaturation of the double stranded DNA, emitting light when the hybridization takes place. The qPCR machine measures the amount of fluorescence emitted by SYBRgreen throughout the cycles. The result is given in threshold cycles (Ct), that corresponds to the number of cycles in which the amplification curve reaches the threshold defined by the fluorescence and the beginning of the exponential phase, defined automatically by the software (RT-PCR ViiA7 - Applied Biosystems). We used 384 well PCR plates and two replicates of each sample. House keepers mβ-actin and mGAPDH. Mastermix: SYBRgreen (BioRad) + Primers forward and reverse of the DNA tanger + water (RNase /DNase free). Temperature of the cycles is defined by the machine and the melting curve is obtained by the software. The data were analyzed with the 2^(-ΔΔ*Ct*) method (Yuan, Reed, Chen, & Stewart, 2006).

Primers used: mGAPDH-F, 5'-TTCACCACCATGGAGAAGGC-3'; mGAPDH-R: 5'-CCCTTTTGGCTCCACCCT-3'; Zeb2NAT-F, 5'-CTGGACCCCTCTACACCTCA-3'; Zeb2NAT-R, 5'-CCAATCCCTTCAGAGCAAAG-3'; Zeb2-F, 5'-CGTGGTGAACTATGACAAC-3'; Zeb2-R, 5'-GCTACAAAGAGGGCAGGAA-3'; Zeb2intron-F, 5'-CGTGTGCATTCCCTCATACG-3'; Zeb2intron-R, 5'-CTGTTTGGTGTGTTGCACTC-3';

Cell cultures

Fibroblasts were taken from C57BL6 i4F-B reprogrammable mice provided by M. Serrano [101]. This transgenic model has a doxycycline-induced cassette with the reprogramming factors Oct4, Sox2, Klf4 and c-Myc, inserted in tPPARg locus, and xTTA cassette inserted in the Rosa26 locus. MEFs were isolated from E12,5 - E13,5 i4F mouse embryos and cultured in DMEM supplemented with 10% of Fetal Bovine Serum and 1% of penicillin/streptomycin. Fibroblasts from ~10 weeks old i4F mice were obtained by culturing small ear explants in the same medium described previously for MEFs. For the reprogramming assays, a density of 50 000 fibroblasts were plated in each well of 6 well gelatinized-coated plates (day 0) in MEFs medium. As soon as the transfections started (day 1), MEFs medium was switched from DMEM supplemented 10% Knockout Serum Replacement, 1% penicillin/streptomycin, 1%NEAA, B-mercaptoethanol (0,1mM), 1% L-glu, 0,02% LIF. For the experiments in which reprogramming was not needed, fibroblasts from *wt* C57BL6 were used. All cells were handled and cultured under sterile conditions and incubated at 37ºC and 5% CO2.

Isolation of iPS colonies.

Colonies were detached from the plates and picked with the help of a needle and a micropipette, respectively. For mechanically dissociating the colonies, we used a syringe. For the chemical dissociation of colonies we used Trypsin-EDTA (0,25%) for 5 minutes at 37ºC. After dissociation, cells from each colony were plated in 96 well plates. After dissociation, cells from each colony were plated in 96 well gelatinized-coated plates and cultured in ESC medium + 2i (CHIR99021 (3 μM) and PD0325901 (1 μM)) or in a feeder layer in ESC medium.

Loss of function of Zeb2NAT and Zeb2 and cellular reprogramming.

Each knockdown was performed by transfecting specific LNA GapmeRs for each target -two for Zeb2NAT, two for Zeb2 and one for the negative control (scramble) - in MEFs and adult fibroblasts. For the transfection we use RNAiMAX Lipofectamine (Life technologies) and LNA GapmeRs (Exiqon). Transfection protocol was the following for each p35 plate (contained in 6 well plates): Make Solution A (125μl optiMEM + 1,5μl of each LNA 25nmol) and solution B (125μl optiMEM + 2μl lipofectamine), incubate at room temperature (RT) for 5 minutes, then

add solution A to B, incubate 20 minutes at RT and to the respective well, each with approximately 50 000 cells (p35, 6 well plates). In lipofectamine control, the LNA GapmeR of the solution A was replaced by lipofectamine. Transfection was performed in two consecutive days for each experiment, as depicted in Fig. 10 A. Cellular reprogramming was induced at day 2 by adding the transgene activator, doxycyline (1,5 mg/ml), to the medium. Medium was changed every 48 hours. Alkaline phosphatase staining was performed following the manufacturer's instructions (Milipore).

LNA GapmeRs.

High affinity strand specific antisense oligonucleotides used in the inhibition of RNA (i.e. lncRNA and mRNA). Each one is made out of DNA with specificity for the target, combined with two ends of LNA, which confers it nuclease resistance. LNA GapmeRs binds to its target forming DNA-RNA hybrids which are then degraded by RNaseH, conferring a highly efficient knockdown.

Specific LNA GapmeRs used:

- Scrambe: Negative control;
- αNAT: Zeb2NAT 6, Zeb2NAT12;
- αZeb2: Zeb2 6, Zeb2 12.

Western Blot:

Cells were lysed using Trypsin-EDTA (0,25%) and extracts were obtained using the RIPA buffer. Proteins were then quantified using Bradford protein assay. After denaturation, samples were loaded in a 8% SDS-PAGE gel (30% of acrylamide). After electrophoresis and gel transfer, the nitrocellulose membrane was blocked with 5% BSA and incubated with antibodies anti-Zeb2 (ABE573 1:500) and anti-Tubulin (Sigma 1:2000).

Immunocytochemistry

Our iPS cell lines were seeded and cultured in gelatinized-coated cover slips (0,1%). After fixing cells with paraformaldehyde (4%) for 30 minutes at RT, they were then permeabilized with Triton-X (0,2%) for 10 minutes and blocked with BSA (2%) for 45 minutes at RT. For Nanog detection we used a conjugated antibody anti-Nanog, Alexa fluor 488, 1:100 from eBioscience. For SSEA-1 we used an antibody anti-SSEA1, 1:100 from Millipore. As a nuclear stain we used DAPI. Photos were acquired using a Zeiss confocal microscope.

Statistical analysis.

The statistical method applied for the analysis of our results was Student's t test. We assume that, given the same conditions, colony's appearance represent a normal distribution. Since our hypothesis stated that the knockdown of Zeb2NAT enhanced cellular reprogramming, we analyzed the statistical significance of the data shown in Chapter 2: Figures 11 G, 12 A and 13 A with an unpaired one-tailed t-test. For the analysis, the null hypothesis (H_0) was: The knockdown of Zeb2NAT has no effect in cellular reprogramming (mean αNAT = control (scramble)); and the alternative hypothesis (Ha): The knockdown of Zeb2NAT enhances (has a positive effect) in cellular reprogramming (mean αNAT > scramble). In the case we used a two-tailed t-test, the alternative hypothesis would be: The knockdown of Zeb2NAT has an effect in cellular reprogramming (mean αNAT ≠ scramble). In this case we would be looking to the two extremities of the Gaussian distribution function, since there could be a negative or a positive effect. All the other data was analyzed with unpaired two-tailed t-test. Software used: Microsoft Excel and Graphpad Prism.

Chapter 3

This page intentionally left blank

Conclusion

The advance of technology during the last decades gave scientists the tools to make high-throughput sequencing more available and cheaper. This lead to a bigger knowledge of the genome, in particular to the identification of many non-coding RNAs. Non-coding RNAs were disregarded comparing to the protein-coding transcripts, and until recently, no function have been linked to this big family of transcripts. Apart from existing a big load of lncRNAs yet to be discovered in the vastness of our genome, many more are very far way from being functionally described, making lncRNAs a hot and interesting subject.

The work presented here enlightened us on the unknown world of long non-coding RNAs. It pointed out aging is not only controlled by the coding transcriptome, but also, and similarly, by the non-coding transcriptome. We were able to show for the first time that the manipulation of a non-coding RNA could contribute to counteract some of the severe effects of aging, particularly by attenuating the aged-imposed barriers limiting cellular reprogramming. Also, we highlight the existence of many variables when comparing reprogramming efficiency between different ages. Studying aging is a delicate process. Cellular and organismal aging are related, and since the first can aggravate the last, misleading conclusions can be made without taking it into account. The use of LNA GapmeRs as our way to target the lncRNA Zeb2NAT has proven to be an effective and straightforward methodology, with no need for the use of virus. Since Zeb2NAT is conserved in human and mouse, we could have tested human cells. However, using the reprogrammable mouse model was a handy tool that is not possible to replicate in human samples.

iPS technology holds an immense promise for regenerative medicine, drug screening and disease modeling. Therapies using patient's own cells will be the future of personalized

medicine. However, cellular reprogramming has yet to evolve in order to be completely safe for human therapies.

The existence of a colossal number of lncRNAs with unknown functions makes us predict that many other lncRNAs should be involved not only in aging and stemness, but also in many human diseases. Further studies of Zeb2NAT and other similar of divergent lncRNA are, thus, of the utmost importance. The lncRNA databases are updated at a daily basis, identifying novel possible targets with biological functions.

This project highlighted a novel role for non-coding RNAs in improving the development of future stem cell-based therapies either for patient-specific treatments or to disease modeling.

References

Abad, M., Mosteiro, L., Pantoja, C., Canamero, M., Rayon, T., Ors, I., . . . Serrano, M. (2013). Reprogramming in vivo produces teratomas and iPS cells with totipotency features. *Nature, 502*(7471), 340-345. doi:10.1038/nature12586

Banito, A., Rashid, S. T., Acosta, J. C., Li, S., Pereira, C. F., Geti, I., . . . Gil, J. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev, 23*(18), 2134-2139. doi:10.1101/gad.1811609

Beltran, M., Aparicio-Prat, E., Mazzolini, R., Millanes-Romero, A., Masso, P., Jenner, R. G., . . . de Herreros, A. G. (2015). Splicing of a non-coding antisense transcript controls LEF1 gene expression. *Nucleic Acids Res, 43*(12), 5785-5797. doi:10.1093/nar/gkv502

Beltran, M., Puig, I., Pena, C., Garcia, J. M., Alvarez, A. B., Pena, R., . . . de Herreros, A. G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev, 22*(6), 756-769. doi:10.1101/gad.455708

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., . . . de Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature, 447*(7146), 799-816. doi:10.1038/nature05874

Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W. P., . . . Cepko, C. L. (2004). Genomic analysis of mouse retinal development. *PLoS Biol, 2*(9), E247. doi:10.1371/journal.pbio.0020247

Blanpain, C., & Fuchs, E. (2006). Epidermal stem cells of the skin. *Annu Rev Cell Dev Biol, 22*, 339-373. doi:10.1146/annurev.cellbio.22.010305.104357

Bortin, M. M. (1970). A compendium of reported human bone marrow transplants. *Transplantation, 9*(6), 571-587.

Bu, D., Yu, K., Sun, S., Xie, C., Skogerbo, G., Miao, R., . . . Zhao, Y. (2012). NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res, 40*(Database issue), D210-215. doi:10.1093/nar/gkr1175

Cai, X., Hagedorn, C. H., & Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA, 10*(12), 1957-1966. doi:10.1261/rna.7135204

Campbell, K. H., McWhir, J., Ritchie, W. A., & Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature, 380*(6569), 64-66. doi:10.1038/380064a0

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., . . . Hayashizaki, Y. (2005). The transcriptional landscape of the mammalian genome. *Science, 309*(5740), 1559-1563. doi:10.1126/science.1112014

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., & Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development, 132*(5), 885-896. doi:10.1242/dev.01670

Chou, Y. F., Chen, H. H., Eijpe, M., Yabuuchi, A., Chenoweth, J. G., Tesar, P., . . . Geijsen, N. (2008). The growth factor environment defines distinct pluripotent ground states in novel blastocyst-derived stem cells. *Cell, 135*(3), 449-461. doi:10.1016/j.cell.2008.08.035

Crick, F. (1970). Central dogma of molecular biology. *Nature, 227*(5258), 561-563.

de Souza, N. (2012). The ENCODE project. *Nat Methods, 9*(11), 1046.

Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., . . . Mattick, J. S. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res, 18*(9), 1433-1445. doi:10.1101/gr.078378.108

Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature, 292*(5819), 154-156.

Gatti, R. A., Meuwissen, H. J., Allen, H. D., Hong, R., & Good, R. A. (1968). Immunological reconstitution of sex-linked lymphopenic immunological deficiency. *Lancet, 2*(7583), 1366-1369.

Ghosal, S., Das, S., & Chakrabarti, J. (2013). Long noncoding RNAs: new players in the molecular mechanism for maintenance and differentiation of pluripotent stem cells. *Stem Cells Dev, 22*(16), 2240-2253. doi:10.1089/scd.2013.0014

Giorgetti, A., Montserrat, N., Rodriguez-Piza, I., Azqueta, C., Veiga, A., & Izpisua Belmonte, J. C. (2010). Generation of induced pluripotent stem cells from human cord blood cells with only two factors: Oct4 and Sox2. *Nat Protoc, 5*(4), 811-820. doi:10.1038/nprot.2010.16

Gregory, S. G., Barlow, K. F., McLay, K. E., Kaul, R., Swarbreck, D., Dunham, A., . . . Prigmore, E. (2006). The DNA sequence and biological annotation of human chromosome 1. *Nature, 441*(7091), 315-321. doi:10.1038/nature04727

Grote, P., Wittler, L., Hendrix, D., Koch, F., Wahrisch, S., Beisaw, A., . . . Herrmann, B. G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell, 24*(2), 206-214. doi:10.1016/j.devcel.2012.12.012

Gurdon, J. B., Elsdale, T. R., & Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature, 182*(4627), 64-65.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., . . . Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature, 458*(7235), 223-227. doi:10.1038/nature07672

Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., . . . Lander, E. S. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature, 477*(7364), 295-300. doi:10.1038/nature10398

Hawkins, P. G., & Morris, K. V. (2010). Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription, 1*(3), 165-175. doi:10.4161/trns.1.3.13332

Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res, 25*, 585-621.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., . . . Zamir, A. (1965). STRUCTURE OF A RIBONUCLEIC ACID. *Science, 147*(3664), 1462-1465.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., . . . Deng, H. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science, 341*(6146), 651-654. doi:10.1126/science.1239278

Huttenhofer, A., Schattner, P., & Polacek, N. (2005). Non-coding RNAs: hope or hype? *Trends Genet, 21*(5), 289-297. doi:10.1016/j.tig.2005.03.007

Ieda, M., Fu, J. D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B. G., & Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell, 142*(3), 375-386. doi:10.1016/j.cell.2010.07.002

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol, 3*, 318-356.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821. doi:10.1126/science.1225829

Kahan, B. W., & Ephrussi, B. (1970). Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *J Natl Cancer Inst, 44*(5), 1015-1036.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., . . . Consortium, F. (2005). Antisense transcription in the mammalian transcriptome. *Science, 309*(5740), 1564-1566. doi:10.1126/science.1112009

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., . . . Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A, 106*(28), 11667-11672. doi: 10.1073/pnas.0904715106

Kim, D. H., Marinov, G. K., Pepke, S., Singer, Z. S., He, P., Williams, B., . . . Wold, B. J. (2015). Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell, 16*(1), 88-101. doi:10.1016/j.stem.2014.11.005

Kim, J. B., Greber, B., Arauzo-Bravo, M. J., Meyer, J., Park, K. I., Zaehres, H., & Scholer, H. R. (2009). Direct reprogramming of human neural stem cells by OCT4. *Nature, 461*(7264), 649-643. doi:10.1038/nature08436

Kim, J. B., Sebastiano, V., Wu, G., Arauzo-Bravo, M. J., Sasse, P., Gentile, L., . . . Scholer, H. R. (2009). Oct4-induced pluripotency in adult neural stem cells. *Cell, 136*(3), 411-419. doi: 10.1016/j.cell.2009.01.023

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J, 23*(20), 4051-4060. doi:10.1038/sj.emboj.7600385

Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Canamero, M., . . . Serrano, M. (2009). The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature, 460*(7259), 1136-1139. doi:10.1038/nature08290

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., . . . Pei, D. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell, 7*(1), 51-63. doi:10.1016/j.stem.2010.04.014

Lin, T., & Wu, S. (2015). Reprogramming with Small Molecules instead of Exogenous Transcription Factors. *Stem Cells Int, 2015*, 794632. doi:10.1155/2015/794632

Mahmoudi, S., & Brunet, A. (2012). Aging and reprogramming: a two-way street. *Curr Opin Cell Biol, 24*(6), 744-756. doi:10.1016/j.ceb.2012.10.004

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A, 78*(12), 7634-7638.

Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol, 20*(3), 300-307. doi:10.1038/nsmb.2480

Nelles, L., Van de Putte, T., van Grunsven, L., Huylebroeck, D., & Verschueren, K. (2003). Organization of the mouse Zfhx1b gene encoding the two-handed zinc finger repressor Smad-interacting protein-1. *Genomics, 82*(4), 460-469.

Obokata, H., Sasai, Y., Niwa, H., Kadota, M., Andrabi, M., Takata, N., . . . Wakayama, T. (2014). Bidirectional developmental potential in reprogrammed cells with acquired pluripotency. *Nature, 505*(7485), 676-680. doi:10.1038/nature12969

Obokata, H., Wakayama, T., Sasai, Y., Kojima, K., Vacanti, M. P., Niwa, H., . . . Vacanti, C. A. (2014). Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature, 505*(7485), 641-647. doi:10.1038/nature12968

Okita, K., Ichisaka, T., & Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature, 448*(7151), 313-317. doi:10.1038/nature05934

Prockop, D. J. (2007). "Stemness" does not explain the repair of many tissues by mesenchymal stem/multipotent stromal cells (MSCs). *Clin Pharmacol Ther, 82*(3), 241-243. doi:10.1038/sj.clpt.6100313

Robinton, D. A., & Daley, G. Q. (2012). The promise of induced pluripotent stem cells in research and therapy. *Nature, 481*(7381), 295-305. doi:10.1038/nature10761

Rosenthal, M. D., Wishnow, R. M., & Sato, G. H. (1970). In vitro growth and differetiation of clonal populations of multipotential mouse clls derived from a transplantable testicular teratocarcinoma. *J Natl Cancer Inst, 44*(5), 1001-1014.

Schleiden, M. J. (1839). Contributions to Phytogenesis.

Schwann, T. H. (1993). Microscopial researches into the accordance in the structure and growth of animals and plants. 1847. *Obes Res, 1*(5), 408-418.

Sheik Mohamed, J., Gaughwin, P. M., Lim, B., Robson, P., & Lipovich, L. (2010). Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA, 16*(2), 324-337. doi:10.1261/rna.1441510

Spemann, H. (1938). Embryonic Development and Induction. *Yale University Press, New Haven*.

Stevens, L. C., & Little, C. C. (1954). Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc Natl Acad Sci U S A, 40*(11), 1080-1087.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol, 14*(2), 103-105. doi:10.1038/nsmb0207-103

Sumi, T., Tsuneyoshi, N., Nakatsuji, N., & Suemori, H. (2007). Apoptosis and differentiation of human embryonic stem cells induced by sustained activation of c-Myc. *Oncogene, 26*(38), 5564-5576. doi:10.1038/sj.onc.1210353

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell, 131*(5), 861-872. doi:10.1016/j.cell.2007.11.019

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi: 10.1016/j.cell.2006.07.024

Till, J. E., & McCulloch, E. A. (2012). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. 1961. *Radiat Res, 178*(2), Av3-7.

Tomioka, I., Maeda, T., Shimada, H., Kawai, K., Okada, Y., Igarashi, H., . . . Okano, H. (2010). Generating induced pluripotent stem cells from common marmoset (Callithrix jacchus) fetal liver cells using defined factors, including Lin28. *Genes Cells, 15*(9), 959-969. doi:10.1111/j.1365-2443.2010.01437.x

Utikal, J., Maherali, N., Kulalert, W., & Hochedlinger, K. (2009). Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci, 122*(Pt 19), 3502-3510. doi:10.1242/jcs.054783

Vandewalle, C., Comijn, J., De Craene, B., Vermassen, P., Bruyneel, E., Andersen, H., . . . Berx, G. (2005). SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell-cell junctions. *Nucleic Acids Res, 33*(20), 6566-6578. doi:10.1093/nar/gki965

Waddington, C. H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology.

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., . . . Wong, G. K. (2004). Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. *Nature, 431*(7010), 1 p following 757; discussion following 757.

Weissman, I. L., Anderson, D. J., & Gage, F. (2001). Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu Rev Cell Dev Biol, 17*, 387-403. doi:10.1146/annurev.cellbio.17.1.387

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev, 23*(13), 1494-1504. doi:10.1101/gad.1800909

Yamanaka, S., & Blau, H. M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. *Nature, 465*(7299), 704-712. doi:10.1038/nature09229

Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., . . . Rotman, G. (2003). Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol, 21*(4), 379-386. doi:10.1038/nbt808

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, II, & Thomson, J. A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. *Science, 324*(5928), 797-801. doi:10.1126/science.1172482

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P., & Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature, 451*(7175), 202-206. doi:10.1038/nature06468

Yuan, J. S., Reed, A., Chen, F., & Stewart, C. N., Jr. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics, 7*, 85. doi:10.1186/1471-2105-7-85