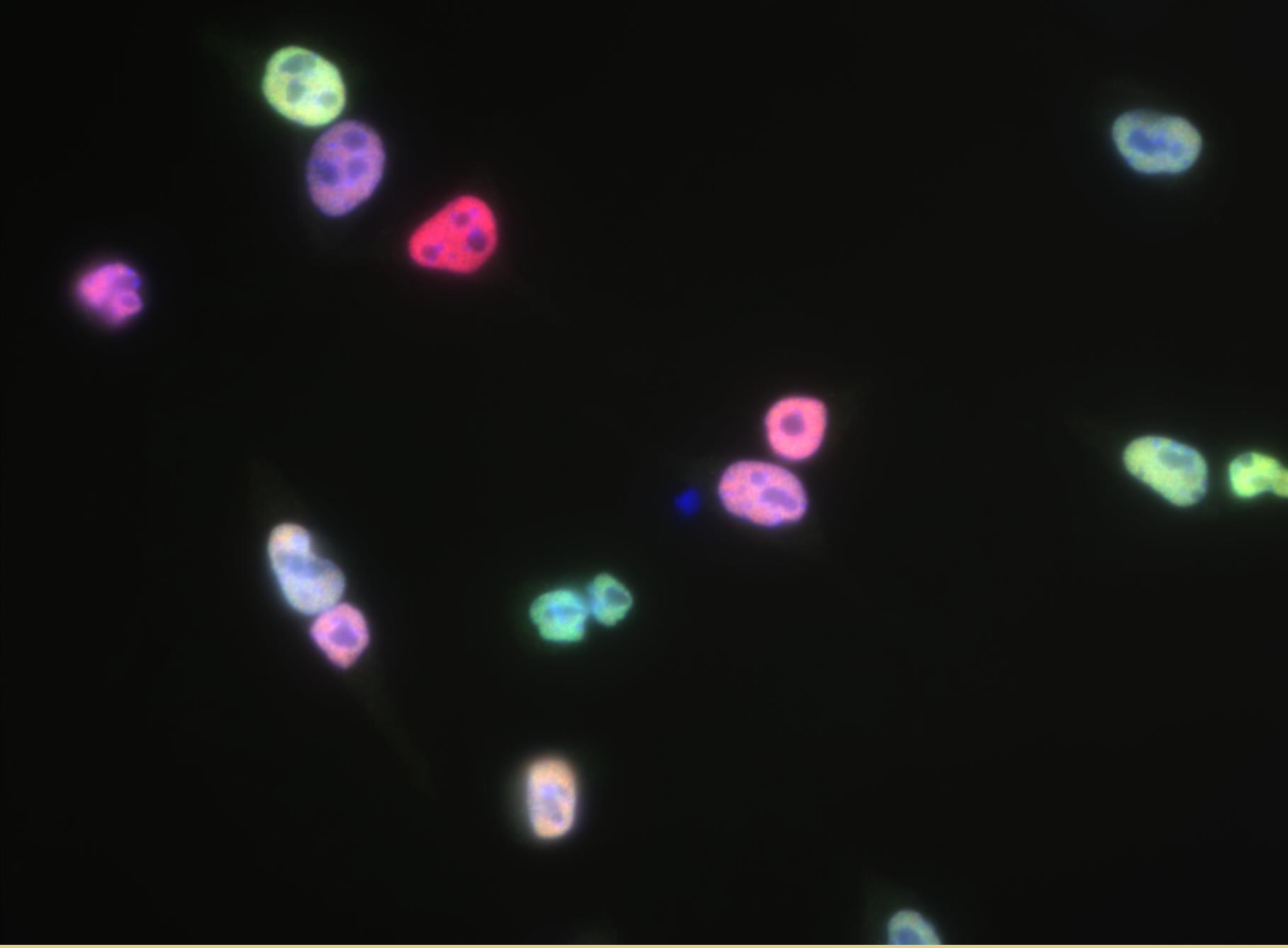

Mariana Rama Pedro Alves

Unravelling the function of Oct4 via its acute depletion in embryonic stem cells

Tese de Mestrado em Biologia Celular e Molecular, realizada sob a orientação científica do Doutor José Silva (WT-MRC Cambridge Stem Cell Institute, Universidade de Cambridge, Reino Unido) e orientação interina do Professor Doutor Carlos Duarte (Centro de Neurociências e Biologia Celular da Universidade de Coimbra), e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra (DCV-FCTUC).

Junho de 2017

· U C ·

UNIVERSIDADE DE COIMBRA

Unravelling the function of Oct4 via its acute depletion in embryonic stem cells

by
Mariana Rama Pedro Alves

Thesis presented to the Department of Life Sciences of the Faculty of Sciences and Technology of the University of Coimbra (DCV-FCTUC), for the MSc degree in Cellular and Molecular Biology, under supervision of Dr. José Silva (Wellcome Trust – MRC Cambridge Stem Cell Institute, University of Cambridge, United Kingdom), and the internal supervision of Dr. Carlos Duarte (University of Coimbra, and Center for Neurosciences and Cell Biology of University of Coimbra). This work was conducted at the Wellcome Trust – MRC Stem Cell Institute, University of Cambridge, United Kingdom.

Tese apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra (DCV-FCTUC), para obtenção do grau de Mestre na especialidade Biologia Celular e Molecular, realizada sob a orientação científica do Doutor José Silva (Wellcome Trust – MRC Cambridge Stem Cell Institute, Universidade de Cambridge, Reino Unido), e orientação interina do Professor Doutor Carlos Duarte (Universidade de Coimbra e Centro de Neurociências e Biologia Celular da Universidade de Coimbra). Este trabalho foi desenvolvido no Wellcome Trust – MRC Stem Cell Institute, Universidade de Cambridge, Reino Unido.

Cover image:

Mouse embryonic stem cells stained for pluripotency factors Oct4 (green), Klf4 (red), and Nanog (white), and nuclei (blue).

The study presented in this dissertation was supported by an Erasmus + studentship.
Dr. José Silva is a recipient of a Wellcome Trust Senior Research Fellowship.
The research in the Silva Lab is supported by a Core Support Grant from the Wellcome Trust and the Medical Research Council to the WT-MRC Cambridge Stem Cell Institute.

The thesis entitled "Unravelling the function of Oct4 via its acute depletion in embryonic stem cells" is my own original work and corresponds to the final version before the defense. Any work done in collaboration is specifically stated in the text.

A tese intitulada "Estudo da função do Oct4 através da eliminação rápida e eficaz da sua proteína em células estaminais embrionárias" é de minha autoria e corresponde à versão final antes da defesa. Qualquer experiência feita em colaboração encontra-se especificamente identificada no texto.

Declaração de confidencialidade | Confidentiality clause

Esta tese de mestrado contém dados confidenciais do Silva Lab – Universidade de Cambridge. Este trabalho poderá apenas ser acessível aos arguentes e membros autorizados do júri. Qualquer publicação e duplicação deste tese de mestrado – mesmo que apenas em parte – é proibida.

A acessibilidade deste trabalho por terceiros requer a permissão explícita da autora Mariana Rama Alves e do Doutor José Silva.

Ao assinar esta declaração, Mariana Rama Pedro Alves, a candidata a exame de mestrado, deseja que este trabalho tenha o seu acesso ao Estudo Geral embargado até que o manuscrito que pode incluir partes dos dados originais desta tese seja aceite para publicação. A autora informará o Estudo Geral aquando de tal acontecimento para o acesso passar a ser livre.

This MSc thesis contains confidential data of the Silva Lab–University of Cambridge. This work may only be available to the first and second reviewers and authorized members of the board of examiners. Any publication and duplication of this MSc thesis – even in part – is prohibited.

An inspection of this work by third parties requires the explicit permission of the author Mariana Alves and Dr. José Silva.

By signing this clause, Mariana Rama Pedro Alves, the MSc exam candidate, wishes this work to remain in access banned to the General Study of the University of Coimbra until the manuscript which can include some parts of the original data of this thesis is accepted for publication. The author will inform the General Study upon that event so that the holding is lifted.

Cambridge, 19.06.2017

A Mestranda/ The MSc student,

_____

Para a Avó Nana | To my grandma, Fernanda

“ Deste modo, durante muito tempo julguei que esta terra onde não nasci fosse tudo o que havia no mundo. (…) Nuto, que ao contrário de mim nunca se afastou de Salto, diz que para resistir neste vale é preciso não o abandonar nunca. “

La luna e i falò, Cesare Pavese

“ On my return home, it occurred to me, in 1837, that something might perhaps be made out on this question by patiently accumulating and reflecting on all sorts of facts which could possibly have any bearing on it. “

On the origin of the species, Charles Darwin

Acknowledgements

I obviously must start by thanking José. For letting me spend 8 weeks in your lab when I was just a young undergrad with everything to prove, showing me what science is like at its most competitive level and believing I can make part of it, for letting me come back after those weeks with much to improve, for the useful feedback, focus on structure and big picture, and for tirelessly helping me spread my wings onto my next chapter, including 10pm existential phone calls. I am and always will be truly grateful.

Secondly, I would like to thank Lawrence. I appreciate the fact that looking after and collaborating with someone for a whole year while being a PhD student with its own hurdles is an undoubtedly tough task. Thank you for always explaining concepts with such brilliance and with the same patience regardless of how many times I had asked you the same question before.

I would like to thank the Silva Lab members. Beza for the sharing of expertise; Charlotte for January's companionship and IF tips; Elsa for eye-opening and arm-holding; Hannah for the huggles, sharpest scientific comments, for well-being consciousness and for being the first to teach me what a properly controlled experiment should look like; Katie for the useful discussions, for making me part of the wonderful stem cell exchange project, finding beating bodies and for always laughing at my jokes; Tim for setting high standards; Sergey for making me less afraid in the lab at weird hours, Rodrigo for helping me arrive here in the first place, my summer buddy Saeed, Alex for always keeping in touch, and Yael for sensible advice.

I would like to point out the help of Christophe, Celine and George, be it regarding experiments, macros, mock interviews, general discussions, always there and also making TC much more fun. It is always a plus to have insights from different angles, reminding us to be more critical of our own work. Your support was very important for the success of this work, but most importantly I want to thank you for your friendship. "The Quest stands upon the edge of a knife. Stray but a little, and it will fail, to the ruin of all. Yet hope remains while the Company is true.".

Dearest Timo, thank you for your friendship and for reminding me that working hard rhymes with playing hard as well. Juergen and Sarra, thank you for your useful insights.

I would like to thank the Genome, Tissue Culture and Microscopy facilities of the Cambridge Stem Cell Institute. A special shout-out to Pippa, Becky and Gill for all their support, you make public engagement so much easier and I learned a lot from you.

It is only fair to also mention all the friendly faces around the SCI who contributed to lighten the working days, especially Roy, and including Catherine, Hyunki, Magda, and Seungmin. I would like to acknowledge also all the SCI staff who makes the Institute run.

I would like to thank Erasmus+ for partially funding this research year abroad.

I would like to thank everyone who gave their useful opinion on this dissertation and Katy Marshall for the epic histone images.

Rebecca, you rising star, thank you for showing me all the mystical undergrad-cambridge that there is to see, just how johnians live it well, and sharing so much magic with me. David, from climbing up a ladder to save my locked room to welcoming me in yours and River's house when I arrived, danke schön. Eva, chegaste mesmo no fim mas foste uma irmã e uma mãe! Gonçalo obrigada pela sintonia, Marisa thank you since the start, Anja thank you for always checking up. Andy, Lydia, Silwya, Tomás, Joana, Alyssa and Chiara, thank you for being so always so kind.

This MSc journey did not consist only of this intense scientific year but of the year before as well, which I would like to acknowledge as well.

Even when we raise our bars higher for the future, it is important to remember where we came from and the roads we've been in. Thank you, Professor Francisco, the first door I had open, and also Fi. Thank you, Carlos Faro, for your advices that led me to Cambridge.

Kim, I am so grateful for those 6 weeks of intense professional and personal growth in your lab, and for all your help afterwards as well. Unforgettable weeks which include Shiro, Stine, Marianne, Veronica, Guillem, Svetlana, Pawel, Troels, Ditte, Jordi and Maria. Karen, for not letting me give up on one of the most useful skills in life. Constança, por seres casa e estrela. Copenhagen, the best city of the world, thank for the wonderful breeze, the tan, the proper cycle ways, jordbaerkage, the Danes teaching how life should be lived and for setting proper standards for any city I move in to.

Thank you, Professor Carlos Duarte, for being my internal supervisor and for the help along the way. Thank you, Professora Emília Duarte, for allowing MBCM students the flexibility to choose their own path these two years. Miguel and all the people who embarked in the Funky Virus project. Tadeja, Jemal, Michele, Muluken, Nika, Mara, Alena and Silvia, the ones making me miss home. Pechincha, Pereirinha e Marques, por serem as colegas de faculdade que nunca tive. Joana Saraiva, pelo costume essencial.

People who inspire us play a very important role in our lives. Renata Gomes and Catarina Gomes, I am very lucky to be your friend besides being your fan. João André, obrigada por me teres guiado no caminho da filosofia e das flores do Sr. Ibrahim. Joca, mais que um amigo, um modelo. Cristina Branco, por cantar que saber aqui estar é bom, mas saber não estar também. Moreover, all the scientists I have met during my science communication path have left a little bit of their magic dust with me, especially Martin Chalfie, Oliver Smithies, Marina Cortês, Heather Douglas and António Damásio.

I want to thank all my friends, recent and old, the ones who visited, the ones who didn't, who write and call and stay and show there is no mountain high enough, thank you for your patience and your love.

João, se um dia alguém perguntar por mim, dizes o que disseste quando me ensinaste a andar de bicicleta: que devagarinho, posso voltar a aprender.

To Inês, my sister, agradeço-te sempre pela sabedoria inerente aos teus mais dez anos, a risota, e a adição este ano da banda sonora de menina em Cambridge invernoso.

To my parents, who miss me and always supported me and without whom this would be impossible. Aos meus pais, que têm saudades e sempre me apoiaram e sem os quais nada disto seria possível.

Finally, this thesis is dedicated to my grandma, Fernanda Rama (Avó Nana), a woman ahead of her time, who I grew up with, listening to the stories of her time studying in this city of lights, not knowing that by chance I would end up here as well, crossing Silver Street – and coincidently enough Copenhagen as well; who taught me English between so many other things and who never liked me studying after midnight (good I did not write next to her).

Abstract

Naïve pluripotency is a cell state acquired in mammalian development during epiblast formation, in the pre-implantation embryo. It is characterized by the unbiased potential to make all lineages of the adult organism and can be captured in *in vitro* culture as embryonic stem cells. Oct4, one of the archetypal reprogramming factors, regulates the naïve network and lineage commitment in a defined-level manner. Naïve pluripotent stem cells cannot maintain their identity without Oct4. Although this factor has been extensively studied, the mechanism behind its action is still poorly understood. This is mainly due to the lack of a system to delete Oct4 protein rapidly. In this study, a novel system is used for the depletion of Oct4 protein in a rapid and efficient manner in mouse embryonic stem cells. I characterize this system for the first time, demonstrating depletion of Oct4 by 2h without exhibiting an Oct4-low state, and use it to explore the roles of Oct4 in the transcriptional and epigenetic regulation of naïve identity. I observed that Oct4 depletion leads to immediate changes in histone modifications and non-coding RNA expression prior to the appearance of pronounced transcriptional changes. Moreover, transcription of pluripotency genes can occur in the absence of Oct4 protein in on short timescales. In the long term, cells follow a trajectory toward TE identity upon loss of Oct4, but the increase in expression of TE-related genes does not occur earlier than 24h after Oct4 depletion. Overall, in this work I describe for the first time a direct molecular mechanism of the action of Oct4 in naïve pluripotency maintenance, shedding light in the mystery of how this crucial factor acts to instruct cell identity.

Keywords: *Oct4, Naïve pluripotency, Loss of function, Transcriptional wiring, Epigenetic regulation, Histone modifications, Mouse embryonic stem cells*

Sumário

A pluripotência naïve é um estado celular que ocorre durante o desenvolvimento dos mamíferos aquando da formação do epiblasto do embrião pré-implantação. Este estado caracteriza-se pela capacidade, sem qualquer viés, de originar todas as linhagens celulares do organismo adulto e pode ser capturado em cultura *in vitro* na forma de células estaminais embrionárias. O Oct4, um dos principais factores de reprogramação, regula a identidade naïve e a transição para diferenciação, de forma dependente do seu nível de expressão. As células estaminais pluripotentes naïve são incapazes de manter a sua identidade na ausência do Oct4. Embora este factor tenha sido amplamente estudado, os mecanismos subjacentes à sua acção são ainda desconhecidos. Este facto pode ser explicado, principalmente, pela inexistência, até à data, de um sistema capaz de eliminar rapidamente a proteína Oct4. Neste estudo, é usado um novo sistema para induzir a degradação da proteína Oct4 de forma rápida e eficiente em células estaminais embrionárias de ratinhos. Este sistema é caracterizado pela primeira vez neste trabalho, demonstrando a degradação do Oct4 ao fim de 2h, sem passar por um estado de reduzida expressão de Oct4, e é usado para explorar o papel do Oct4 na regulação transcricional e epigenética da identidade naïve. Foi possível observar que a degradação do Oct4 induz alterações imediatas nas modificações de histonas e na expressão de RNA não codificante, que ocorrem previamente ao aparecimento de mudanças acentuadas na transcrição. Além disso, a transcrição de genes marcadores de pluripotência pode ocorrer na ausência da proteína Oct4 a curto prazo. A longo prazo, as células seguem uma trajetória em direcção à identidade de trofectoderme, mas o aumento na expressão de genes associados à trofectoderme não ocorre antes das 24h após a degradação do Oct4. Em conclusão, neste trabalho descrevo pela primeira vez um mecanismo molecular directo da acção do Oct4 na manutenção da pluripotência naïve, ajudando a clarificar de que forma este factor crucial contribui para estabelecer a identidade celular.

Palavras-chave: *Oct4, Pluripotência naïve, Perda de função, Rede transcripcional, Regulação epigenética, Modificações de Histonas, Células estaminais embrionárias de ratinho*

Contents

List of Figures and Tables

Abbreviations most commonly used

2iL: medium of dual inhibition of Mek/Erk and GSK3, supplemented with LIF

cDNA: complementary DNA

ChIP: chromatin immunoprecipitation

DAPI: 4',6-diamidino-2-phenylindole

DOX: doxycycline

EB: embryoid body

ESC: embryonic stem cell

h: hour(s)

ICM: inner cell mass

LIF: leukemia inhibitory factor

PCR: polymerase chain reaction

qPCR: quantitative polymerase chain reaction

RNA-seq: high-throughput sequencing of RNA

S: second(s)

SL: serum-based media, supplemented with LIF

TE: trophectoderm

TSC: trophoblast stem cell

WT: wild type

CHAPTER 1

Introduction

1.1 The ground state of development

"Let us suppose the mind to be, as we say, white paper, void of all characters" (Locke, 1689). In 1689, John Locke, reviving the aristotelic "tabula rasa" hypothesis, proposed that mind at birth was like a tabula rasa, which would be furnished through experience. A same state of blank slate would be discovered in mouse embryonic cells and described four centuries later as naïve pluripotency (Nichols & Smith, 2009). Naïve pluripotent cells can be found in the pre-implantation mouse embryo (corresponding to the naïve epiblast at embryonic day 4.5), and are characterized by their unbiased capacity to make all cell lineages of the adult organism, including the germ line, while also having the capacity to self-renew. Female naïve pluripotent stem cells still have two active X chromosomes. This state is lost upon implantation, with the priming of the cells for lineage commitment (Nichols & Smith, 2009). The transient naïve cell state can be captured *in vitro* by culturing embryonic stem cells (ESCs) (Evans & Kaufman, 1981; Martin, 1981) or by reprogramming somatic cells to induced pluripotency (Takahashi & Yamanaka, 2006), and maintained in defined culture conditions, which provide external cues to control pluripotency and self-renewal. *In vitro*, these cells exhibit small dome-shape morphology.

1.1.2 The two layers of naïve identity

Central in mouse ESC self-renewal is the cytokine Leukemia inhibitory factor (LIF), identified almost 30 years ago (Smith *et al*., 1988; Williams *et al*., 1988). This contributes to ESC maintenance mainly through Janus activated kinase (JAK)-dependent signal transducer and activator of transcription 3 (STAT3) activation (Niwa *et al*., 1998; Matsuda *et al*., 1999), whose downstream targets are various factors of the pluripotent network, but mainly Klf4, Klf5 and Tfcp2l1 (Bourillot *et al*., 2009; Martello *et al*., 2013; Niwa *et al*., 2009; Ye *et al*., 2013). LIF also activates the PI3K-Akt signalling pathway (Niwa *et al*., 1998), which is described to positively regulate the expression of pluripotency factors Nanog and Tbx3 (Niwa *et al*., 2009; Watanabe *et al*., 2006). Beneficial external cues do not include only positive regulators of pluripotency, but also negative regulators of differentiation. Such is the case for BMP4, a factor present in serum (often added to ESC media) which inhibits neural differentiation through the Smad-Id signalling pathway (Ying *et al*., 2003a). LIF blocks differentiation into non-neural lineages (Tropepe *et al*., 2001; Ying *et al*., 2003b), thus acting in synergy with BMP4 to prevent differentiation into any lineage. These two external inputs together are sufficient for self-renewal maintenance and are found in Serum/LIF (SL) medium conditions (described in the Methods chapter).

20 years after the identification of LIF, an optimized serum-free media condition to maintain the naïve state was created and named 2i medium (Ying *et al*., 2008), standing for two inhibitors, one of the Fgf/Mek/Erk signalling pathway and one of the GSK3 pathway. The inhibition of the GSK3 pathway leads to repression of pluripotent repressor Tcf3 (also known as Tcf7l1) through β-catenin activation (Wray *et al*., 2011) and increases the viability and biosynthetic capacity of cells (Ying *et al*., 2008). The inhibition of the Fgf/Mek/Erk pathway leads to a reduced propensity to exit from pluripotency into any lineage (Kunath *et al*., 2007). 2i medium is sufficient to maintain self-renewal, but the addition of LIF enhances this potential (Ying *et al.*, 2008).

It is important to take note that external physical cues also play a role in the regulation of the naïve cell identity. It has been described that soft gel substrates, unlike the very stiff mainstream tissue culture plastic typically used, enhance ESC self-renewal, bypassing some of the chemical needs described above (Chowdhury *et al*., 2010).

The nature of a cell is modulated by many diverse factors, such as transcriptional control and environmental cues. Starting with 24 candidates, Takahashi and Yamanaka were able to narrow down a combination of four, capable of acting as a reprogramming cocktail: Oct4, Sox2, c-Myc, and Klf4. The possibility of generating induced-pluripotent stem cells (iPSCs) from both human and mouse somatic cells by ectopic expression of only four factors accentuates the importance and influence of transcriptional wiring in cell state/identity, by the activation and/or repression of certain gene expression programmes. The establishment and maintenance of the naïve cell identity are achieved by a synergistic orchestration of the extrinsic stimulation described before and a potent internal transcriptional circuitry, a phenomenon as wonderful as it is challenging to study. Klf4, Klf5, Tfcp2l1, Nanog and Tbx3 have been noted above as pluripotency factors important in the naïve identity circuitry. Other members of this network include Oct4 (Nichols *et al*., 1998), Sox2 (Avilion *et al*., 2003), Rex1 (Masui *et al*., 2008), Esrrb (Luo *et al*., 1997), Klf2 (Kuo *et al*., 1997), Nr0b1 (Niakan *et al*., 2006), and others. Oct4, Nanog and Sox2 are considered the core transcription factors (Kim *et al*., 2008; Niwa *et al*., 2007; reviewed in Nichols & Smith, 2012), which specify naïve cell identity by occupying regulatory regions to either silence or enhance the expression of genes (Jaenisch & Young, 2008). These three core factors, together with the rest of the vast network of pluripotency factors cooperate to sustain naïve identity.

Naïve cells maintained in SL conditions exhibit heterogeneous expression of several pluripotency genes, such as PGC7/Stella (Hayashi *et al*., 2008), Klf4, Klf5, Tbx3 (Reynolds *et al*., 2012), Rex1 (Toyooka *et al*., 2008), Esrrb (Festuccia *et al*., 2012) and Nanog (Chambers *et al*., 2007). This heterogeneity in Nanog is observable in the form of interconvertible Nanog-high and Nanog-low populations (up to 20% of the cells) present which differ in the stability of their

naïve network, observed by their different rates of spontaneous differentiation (Chambers *et al*., 2007). On the other hand, 2i/LIF (2iL) conditions strongly promote homogeneous expression of naïve pluripotency genes (Wray *et al*., 2010), proving to be the optimal condition for mESC culture. These differences in the transcriptional circuitry and cell identity depending on the culture condition show how important the interaction between internal and external identity regulation is.

1.1.3 A third layer of identity control

There is another layer of regulation, namely epigenetic regulation, adding even more complexity but also more beauty to the nature of this transient identity. The chromatin of ESCs is substantially different from that of cells which have exited pluripotency. ESCs display global DNA hypomethylation (Leitch *et al*., 2013) and exhibit a "hyperdynamic", largely euchromatic chromatin state (Meshorer *et al*., 2006; Zhu *et al*., 2013; Wu *et al*., 2016). This is lightly packed and more accessible, which allows for a more adaptable nature (Lessard & Crabtree, 2010) suiting the transient nature of the naïve state *in vivo*. This plasticity allows for the quick resolution of epigenetic changes into transcriptional ones when cells start to commit to a specific lineage (Efroni *et al*., 2008). Another characteristic which permits this flexibility is the presence of bivalent chromatin domains: regulatory regions of genes populated by both active (acetylation of H3 and H4; methylation of H3K4) and repressive histone marks (trimethylation of H3K27) (Bernstein *et al*., 2006; Mikkelsen *et al*., 2007; Azurara *et al*., 2006), schematically represented in Figure 1.1. Genes marked in this way are typically related to lineage commitment and are silenced in the naïve state; however, this bivalent histone marking keeps them in a "poised" state – ready to be activated if the repressive mark is displaced, or to be kept repressed, depending on which fate the cell will enter when exiting pluripotency (Bernstein *et al*., 2006). Moreover, pressure for active gene expression is exerted by the methylation of H3K4 (Koche *et al*., 2011) and acetylation of H3K27 (Creyghton *et al*., 2010), including in vast enhancer regions of pluripotency-related genes, regions which are also bound by transcription factors Oct4, Nanog, Sox2, Esrrb and Klf4 (Whyte *et al*., 2013).

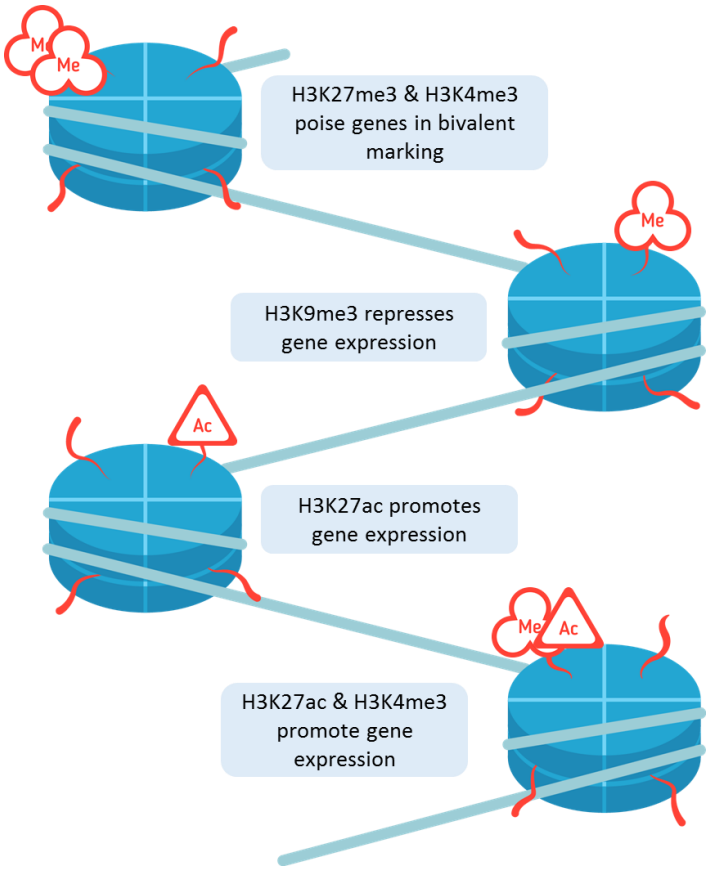

Figure 1.1 | Chromatin and histone modifications
Scheme representing nucleosomes around the double-stranded DNA (here represented as a single strand). Histone tails are shown as orange and histone modifications are included, such as acetylation (triangle) or trimethylation (shamrock-like structure). The upper left histone tail is trimethylated twice in two different places, representing bivalent marking by H3K4me3 and H3K27me3. Figure adaptation using images designed by Katy Marshall.

Several chromatin modifiers in naïve pluripotency are recruited by the pluripotent transcription factors (Young, 2011), and important ones include Trithorax group Set/MLL complexes (related to H3K4 methylation) (Ang *et al*., 2011), Polycomb group PRC1 and PRC2 complexes (related to H3K27me3) (Pasini *et al*., 2007; Chamberlain *et al*., 2008; Leeb *et al*., 2010), Jmjd2c (H3K9me3 demethylase) (Loh *et al*., 2007), and SetDB1 (H3K9 methyltransferase). Notably, SetDB1 is responsible for repressing several genes (Bilodeau *et al*., 2009), making it an important factor in preventing differentiation to the trophectoderm (TE) lineage (Dodge *et al*., 2004; Lohnmann *et al*., 2010).

TE hallmarks include flat morphology and expression of Cdx2, Eomes, Gata3 (activated by Tead4) (Nishioka *et al*., 2009; Ralston *et al*., 2010; Yagi *et al*., 2007). This lineage, which will contribute to part of the placenta, is specified in the first fate decision made by the totipotent zygote. At around the 8-16 cell stage, the indistinguishable cells, called blastomeres, compact. The outside cells will follow the TE route, while the cells inside will form the inner cell mass (ICM). This will

then segregate into the hypoblast/primitive endoderm (which will give rise to extraembryonic tissue) and the naïve pluripotent epiblast which will form the embryo proper (Sutherland *et al*., 1990).

Although the discussion until now has been focused on mouse development, ESCs have been derived from several other mammals, including humans (Thomson *et al*., 1995; Li *et al.,* 2008; Thomson *et al*., 1998); however, in this case the cells exhibit similarities with mouse cells derived from the post-implantation stage (Thomson *et al*., 1998; Tesar *et al*., 2007). Furthermore, the well-defined and easily manipulable mouse system can serve as an invaluable tool to understand the nature of the pluripotent identity, and the mechanisms behind the action of many of the factors involved (genetic, epigenetic, chemical, physical), a lot of which remains a mystery. This work will focus on one of the core pluripotency factors, Oct4, frequently described as the most important (Martello & Smith, 2014; Nichols & Smith, 2012), for several reasons which will be detailed below.

1.2 Oct4, the octamer binding protein

Oct4 (also known as OCT3, OTF3, or OTF4) is a transcription factor belonging to Class V of the POU-domain family (Ryan & Rosenfeld, 1997; Holland *et al*., 2007). In mouse, of its 352 amino acids (Okamoto *et al*., 1990; Rosner *et al*., 1990; Scholer *et al*., 1990), 150 consist of a POU-specific domain, a short variable linker region and a POU homeodomain, all responsible for the binding of Oct4 to the genome (Ruvkun & Finney, 1991). The crystal structure of the Oct4-DNA complex is represented in Figure 1.2. This protein is encoded by Pou5f1, localized within or immediately adjacent to the major histocompatibility complex (MHC) region in human chromosome 6 and mouse chromosome 17 (Scholer *et al*., 1990). Moreover, the human and mouse Oct4 proteins share 87% amino acid identity between them (Takeda *et al*.,1992).

Oct stands for octamer binding protein: Oct4 binds the ATGCAAAT motif (or its inverse), as well as the AT-rich sequence TTAAAATTCA (Okamoto *et al*., 1990), in enhancer and/or promoter regions to activate transcription (Petryniak *et al*., 1990). Oct4 can also have a repressive activity, whether by forming complexes with other factors (Niwa *et al*., 2005; Guo *et al*., 2002; Liang *et al*., 2008) or by recruiting epigenetic regulators (Yuan *et al*., 2009; Yeap *et al*., 2009). It is also possible that the binding of Oct4 to the genome can prevent other factors from binding to the same locus by occluding that space, which would impact gene expression. Like the rest of the POU-domain family, Oct4 takes advantage of the flexibility conferred by its bipartite DNA-binding domain, being able to bind DNA motifs with different spacing and/or orientation and exhibiting temporal and spatial specificity on gene regulation, due to its ability to partner with different cofactors (Ryan & Rosenfeld, 1997; Verriizer *et al*., 1991). Some studies suggest that

Oct4 is a pioneer factor (King & Klose, 2017; Chen & Dent, 2014) –a transcription factor which can access compact chromatin without the help of co-factors or chromatin modifiers, binding the genome before other factors (Zaret & Carroll, 2011).

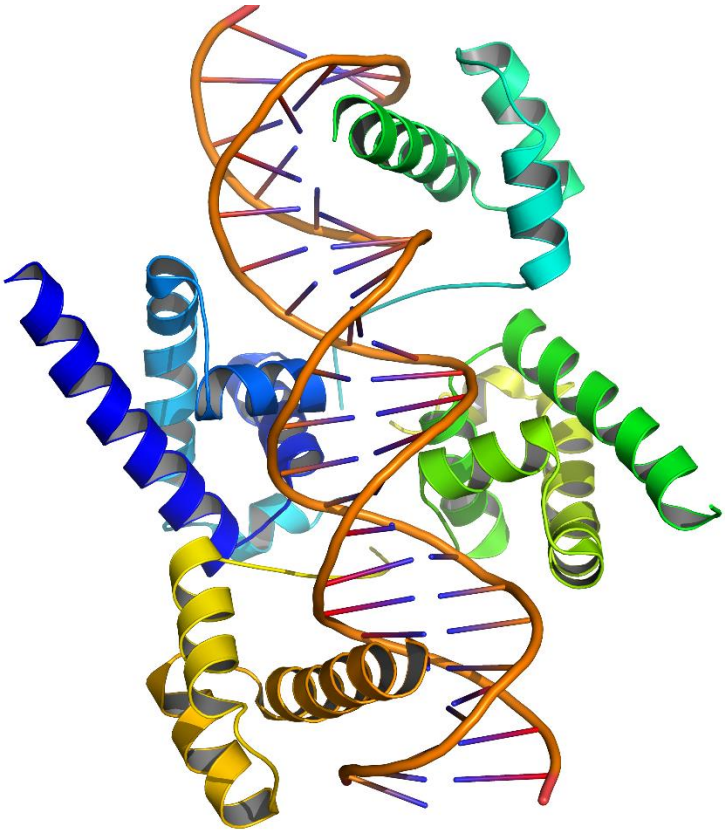

Figure 1.2 | Structure of Oct4-DNA complex
Crystal structure of an Oct4 (light blue)-DNA (grey) complex at 2.8 Angstrom resolution. Protein Data Bank ID: 3L1P

1.3 Oct4, the preeminent pluripotency factor

Originally, Oct4 was identified in ESCs and embryonal carcinoma cells (ECs) (Scholer *et al*., 1989; Lenardo *et al*., 1989; Okamoto *et al*., 1990, Scholer *et al*. 1990). *In vitro*, this protein is expressed in ECs, ESCs and embryonic germ cells (EGCs). *In vivo*, it is initially expressed in the oocyte as a maternal factor and continues to be expressed following zygotic genome activation, uniformly at the morula stage. After that, Oct4 becomes restricted to ICM, having a transient peak of higher expression in primitive endoderm. Finally, Oct4 will be confined to primordial germ cells (PGCs) (Yeom *et al*., 1991; Palmieri *et al*., 1994; Rosner *et al*., 1990; DeVeale *et al*., 2013; Downs, 2008; Ovitt & SchÖler,1998; Osorno *et al*. 2012; Thomson *et al*. 2011).

When comparing the profiles of pluripotent and germline cells with somatic cells, Oct4 is one of the few factors to stand out (Kim *et al*., 2009), and it was the first transcription factor to be labelled as pluripotency regulator (Okamoto *et al*., 1990), currently considered a master contributor for the self-renewal of pluripotent cells (Figure 1.3). Loss-of-function experiments,

in vivo and *in vitro*, strengthened the theory that Oct4 was critical for pluripotency: Oct4-null embryos give rise to non-pluripotent ICM cells stalled without progression to epiblast/hypoblast specification, resulting in death before implantation (Nichols *et al*., 1998) and loss of this protein in the germ cell lineage leads to apoptosis of PGCs (Kehler *et al*., 2004). Furthermore, naïve pluripotent cells differentiate into TE and primitive endoderm when Oct4 is repressed (Niwa *et al*., 2000; Hay *et al*., 2004; Niwa *et al*., 1998). Moreover, induction of ECSC differentiation by retinoic acid leads to a decrease in Oct4 expression (Okamoto *et al*., 1990, Lenardo *et al*., 1989; Scholer *et al*., 1989). But while it would be expected that Oct4 overexpression would enhance the pluripotent identity, as is the case for pluripotency factors Nanog, Klf2, Esrrb, Klf4, Tbx3 and Tfcp2l1 (Martello *et al*., 2013; Chambers *et al*., 2003; Festuccia *et al*., 2012; Hall *et al*., 2009; Niwa *et al*., 2009), an increase in the protein levels of Oct4 leads to ESC differentiation into primitive endoderm and mesoderm (Niwa *et al*., 2000). In fact, the levels of Oct4 which benefit pluripotency are tightly confined (Radzisheuskaya *et al*., 2013). It is also of note that PGCs express Oct4 but are unipotent, although they can give rise to pluripotent embryonic germ cells *in vitro* in culture conditions supplied with three growth factors (Matsui *et al*., 1992; Resnick *et al*., 1992). Interestingly, low levels of this transcription factor do not impair self-renewal, but are detrimental to differentiation of ESCs (Karwacki-Neisius *et al*. 2013; Radzisheuskaya *et al*. 2013).

Despite the fact that both loss-of-function experiments (Avilion *et al*., 2003; Masui *et al*., 2007) and *in vitro* overexpression experiments (Kopp *et al*., 2008) show a similar output when performed with pluripotency factor Sox2, the Sox2-null phenotype in ESCs can be rescued by ectopic expression of Oct4 to wild-type levels (Masui *et al*., 2007), suggesting that Sox2 is a mere activator of Oct4 (Radzisheuskaya & Silva, 2014). Thereby, and even though its role is complex and not precisely understood, Oct4 is considered the preeminent pluripotency factor. But how does Oct4 act? In ESCs, it is believed that Oct4 participates in the auto-regulatory pluripotent network among other factors such as Nanog and Sox2, its function being to activate genes which contribute to this pluripotent state and repress differentiation-related ones (Bilodeau *et al*., 2009; Boyer *et al*., 2005; Boyer *et al*., 2006; Chen *et al*., 2008; Kim *et al*., 2008; Loh *et al*., 2006; Pasini *et al*., 2010). Because of its involvement with lineage commitment *in vivo* (which will be expanded below), others believe that Oct4 is a differentiation driver but its action is blocked by the actions of other pluripotency factors, such as Nanog and Klf4 (Loh & Lim,2011). To understand more thoroughly how Oct4 does act, its relationship with its targets, partners and regulators need to be dissected.

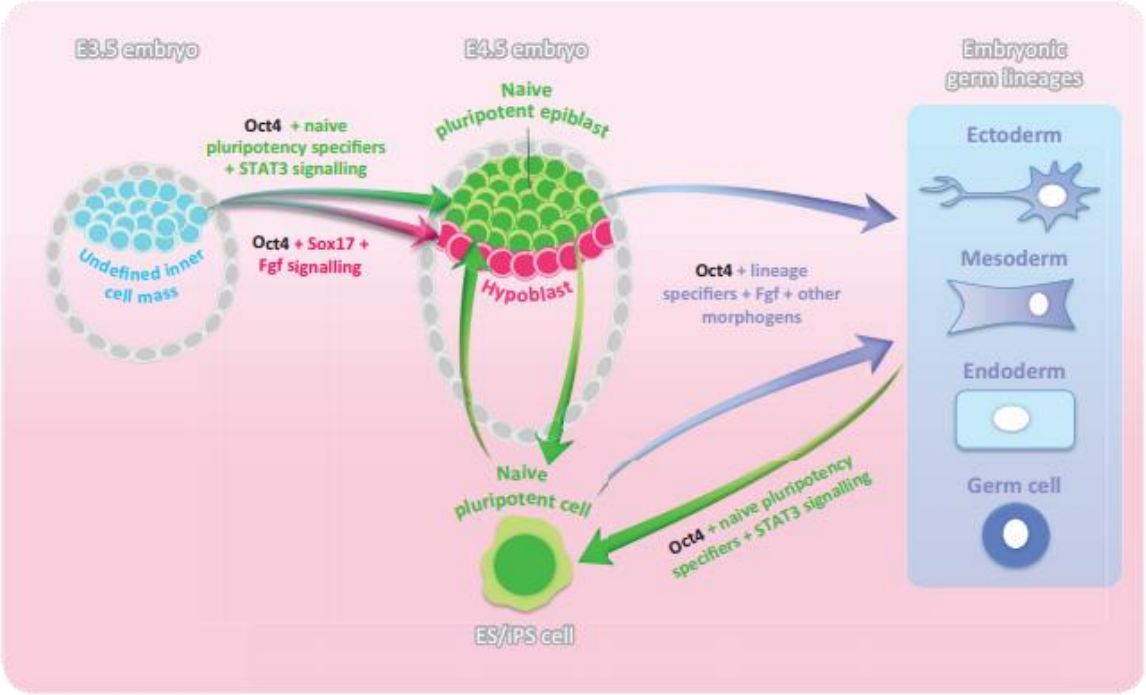

Figure 1.3 | Oct4's regulation of naïve pluripotency and cell state transitions
Scheme from Radzisheuskaya *et al*., 2013 reinforcing the centrality of Oct4, being involved in every direction of cell fate decision at this early stage of development: the formation of the naïve epiblast, the regulation of the hypoblast, the exit from pluripotency, and the reprogramming route from differentiated to induced pluripotent stem cell.

1.4 Behind a great transcription factor, there are always other molecules

So far, a nearly endless list of genes and proteins, both in human and mouse ESCs, has been shown to be possibly regulated by or interacting with Oct4, respectively (Boyer *et al*., 2005; Kim *et al*., 2008; Chen *et al*., 2008; Loh *et al*., 2006; Matoba *et al*., 2006; Pardo *et al*.,2010; Ding *et al*., 2012; van de Berg *et al*.,2010). These include direct targets, partners and regulators. But most of its putative targets do not display an essential role in ESCs – for example proteins involved in basal DNA-processes such as DNA replication, recombination and repair (Pardo *et al*.,2010). Some interpret this as Oct4 acting either "through a select group within these targets" or through a "highly redundant network" (Hall *et al*., 2009). The first suggestion implies that only a restricted group within the multitude of Oct4 targets is meaningful to concretize its function and role, and the second interpretation suggests a redundancy between a variable group of genes and proteins.

Between the multitude of studies on direct physical interactors, there is a high discrepancy in the target list obtained, due mostly to the difference between the techniques used to acquire them, with the highest confidence interactors including members of the chromatin-modifying repressor complexes NuRD (Nucleosome Remodeling and Deacetylase) and SWI/SNF (SWItch/Sucrose Non-Fermentable) (van den Berg *et al*., 2010; Ding *et al*., 2012; Liang *et al*.,

2008; Pardo *et al*., 2010; Wang *et al*., 2006). It is hypothesized that, in ESCs, Oct4 recruits these transcriptional repressors to lineage commitment loci to prevent differentiation (Pesce & Schöler, 2001). Other examples of co-repressive partners are Cdx2 (Niwa *et al*., 2005) and FoxD3 (Guo *et al*., 2002). The short variable linker between the two DNA-binding domains of Oct4 was proposed to function as an interface for protein-protein interactions and be important in the recruitment of epigenetic regulators to Oct4's target genes in reprogramming, as point mutations in this structure impair reprogramming (Esch *et al*., 2013). A study in mouse cells with a mutant linker, comparing its interactome to three different published Oct4 interactomes (van den Berg *et al*., 2010; Ding *et al*., 2012; Pardo *et al*., 2010), identified Chd4 and Smarca4 as Oct4 interactors, as their binding is reduced upon impairment of the linker integrity (Esch *et al*., 2013). Chd4 and Smarca4 are both helicases, the former belonging to the NuRD complex, and the latter belonging to the BAF remodeling complex. Oct4 was also described to interact with H3K9 methyltransferase Eset in trophoblast-lineage silencing in ESCs (Yuan *et al*., 2009), and to mediate nucleosome demethylation and deletion with involvement of histone lysine demethylase JmJd1c recruitment (Shakya *et al*., 2015). Recently, Oct4 was shown to recruit BRG1, a chromatin remodeler, in order to access its targets in inaccessible chromatin (King & Klose, 2017), reinforcing the idea of its activity as a pioneer factor.

Given the complexity of a cell's gene expression programme, particularly in a state in which it is poised for differentiation in many different directions and prone to undergo changes throughout development, it is not surprising that a master factor like Oct4 would need more than direct gene activation/repression activity to perform its duty. Thus, interacting not only with other transcription factors but also with epigenetic regulators such as chromatin modifiers. In fact, it is interesting to observe the tight relationship between the transcriptional and epigenetic machinery.

A well-known transcription factor-partner of Oct4 is the HMG-box transcription factor Sox2 (Ambrosetti *et al*. 1997). The two factors heterodimerize (Pardo *et al*. 2010; van den Berg *et al*., 2010) and bind DNA together, being described to affect the gene expression profile of mouse ESCs (Botquin *et al*., 1998; Nishimoto *et al*., 1999; Yuan *et al*., 1995; Chen *et al*. 2008). As described above, both present the same phenotypes in functional experiments in ESCs (Avilion *et al*., 2003; Masui *et al*., 2007). Sox2's essential activity is suggested by some to only involve of activating Oct4 (Radzisheuskaya & Silva, 2014), while others label Oct4 and Sox2 as "partner molecular rheostats" which balance their expression to direct the course of embryogenesis (Rizzino & Wuebben, 2016). Examples of the synergy between these two molecules are their concerted activation of Fgf-4 by binding its enhancer (Ambrosetti *et al*., 2000; Yuan *et al*., 1995),

their co-binding of UTF1 enhancer (Nishimoto *et al*., 2005) and shared regulation of Nanog (Rodda *et al*., 2005).

The first protein to be described as a functional interactor with Oct4 was adenoviral E1A protein (Schöler, 1991). This protein and other similar proteins mediate distance transactivation by Oct4 – a feature unique to embryonic cells (Pesce & Schöler, 2001). Transactivation is the process of the Oct4 protein recruiting molecules to stabilize the basal transcription machinery, in order to bridge together enhancers and promoters, to ease its task of gene expression regulation (Lim *et al*., 2009; Brehm *et al*., 1997; Vigano *et al*., 1996). This transactivation can occur via the binding of factors and consequent conformational changes at each one of Oct4's termini independently – the amino domain or the carboxy domain (Imagawa *et al*., 1991; Schöler *et al*., 1991; Vigano & Staudt, 1996; Brehm *et al*., 1997) – due to their richness in proline residues (Mermod *et al*., 1989).

According to Hammachi *et al*., Oct4 functions in promoting pluripotency solely by gene activation (Hammachi *et al*., 2012). But it may be argued that the results are directed by the experiment itself, since the targets of Oct4 should be the same regardless of which artificial construct is controlling its activity. Considering this, artificially restricting Oct4's activity to an activator or repressor biases the readouts of this study. In agreement with the authors, the presence of Oct4 at non-expressed genes does not necessarily mean that it is repressing them, as these genes could be non-functional in the context when Oct4 is bound to them. It would be more sensible to consider that Oct4 may act as an activator or as a repressor depending on context, as it has been associated with complexes acting in either manner (Pan *et al*., 2002; Pardo *et al*., 2010; Ang *et al*., 2011; Bilodeau *et al*., 2009; Pasini *et al*., 2010; Yeap *et al*., 2009; Yuan *et al*., 2009; Boyer *et al*., 2006). Moreover, it appears that Oct4 can have a dual activity for a same target, depending on its levels, as has been reported with Rex-1 at Ben-Shushan *et al*., 1998.

More studies need to be performed to further dissect the complex network surrounding Oct4, and efforts to connect the several interactomes in order to find common answers would also be helpful. Maybe the study of Oct4 homologs in other species could also benefit our understanding of its action, since conserved mechanisms usually have a strong functional correlation.

1.4.2 A master regulator needs regulation itself

Although the processes by which Oct4 is expressed could benefit from further study, it is known that this gene is regulated by two enhancers linked to, and located upstream of, the promoter (Figure 1.4). These are the proximal enhancer, 1.2kb away, and the distal enhancer, 2 kb away

(Yeom *et al*., 1996; Minucci *et al*., 1996). Interestingly, these two enhancers act not in synergy but in a switch-like manner – when one is on, the other is off, correlating with cell's identity: ESCs exhibit an active distal enhancer and, following implantation, there is a switch to proximal enhancer activation, directing Oct4's expression in the primitive ectoderm (Yeom *et al*., 1996; Ovitt & Schöler,1998). Further evidence sustaining the association of distal/proximal duality with the undifferentiated/differentiated phenotype was the observation that Oct4's repression by retinoic acid is mediated by the proximal enhancer (Okazawa *et al*., 1991). These regulatory elements can be, in turn, regulated themselves. During gastrulation, there is a wave of *de novo* methylation in embryonic somatic cells (Jaenisch, 1997), in which methylation of these enhancers may occur, accompanying extinction of Oct4 gene activity (Ben-Shushan *et al*., 1993), with protection of PGCs from this process by their segregation into extraembryonic mesoderm (Pesce *et al*., 1998).

Oct4 can also undergo post-translational modifications, such as phosphorylation (Rosfjord *et al*., 1995; Brehm *et al*., 1997), sumoylation (Wei *et al*., 2007) or the addition of O-linked N-acetylglucosamine (O-GlcNAc) (Webster *et al*., 2009). These post-translational modifications do not necessarily affect the overall DNA-binding activity of Oct4 (Rosfjord *et al*., 1995), but evidence suggests a correlation between these events and the modulation of specific regions within this protein (Brehm *et al*., 1997).

Overall, both transcriptional and post-translational regulation of Oct4 are controlled by a complex network of players and processes. This complex regulation environment can be one of the contributors or explanations to the complexity of the functions and actions of Oct4. The specificity of Oct4's functions dependent on cellular and signalling context could be a behavior consequence of its dynamic regulation, which can also be specific to the context. Having already discussed the function of Oct4 in pluripotency, it is time now to focus on the other side of the coin: the role of Oct4 in differentiation.

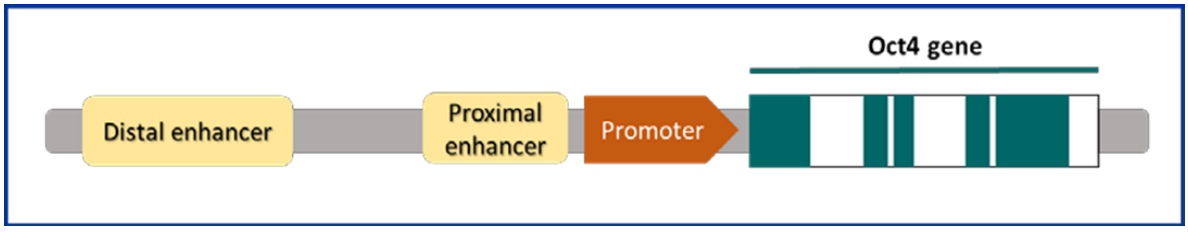

Figure 1.4 | Genomic structure of Oct4
The gene that encodes Oct4 is constituted by five exons (in green), and is regulated by two enhancers linked to a single promoter.

1.5 Oct4 in lineage commitment

Jack of all trades, master of all as well, Oct4 regulates not only pluripotency acquisition but also its loss. As already mentioned, Niwa and colleagues showed in 2000 that differentiation into primitive endoderm and mesoderm can by driven by an increase of Oct4 protein levels as slight as two-fold over ESC level. The fate promoted can change in response to external cues, as enhanced Oct4 levels can contribute to differentiation into various lineages, including neural or cardiac, as long as the conditions are favorable (Shimozaki *et al*., 2003; Zeineddine *et al*., 2006; Thomson *et al*., 2011). In mouse development, Oct4 was confirmed to be necessary for primitive endoderm development (Frum *et al*., 2013), and essential for maintenance of cell proliferation and viability within the primitive streak (DeVeale *et al*. 2013). It was observed from single-cell gene expression analysis that primitive endoderm differentiation requires an ESC level of Oct4 (Kurimoto *et al*., 2006; Guo *et al*., 2010), and that to carry out this process Oct4 may switch partners from Sox2 to Sox17 (Askoy *et al*., 2013). Regarding partnerships, Sox2 is reported to have a direct opposite role to Oct4 regarding mesoderm versus neural ectoderm commitment (Thomson *et al*., 2011). The interaction with the BMP4 pathway was also described to specify commitment to distinct fates, depending on both BMP4 and Oct4 levels (Wang *et al*., 2012). Whereas high levels of both factors in human ESCs lead to commitment into mesoderm, combination of low Oct4 with low, medium or high BMP4 can specify, neuroectoderm, embryonic ectoderm or primitive endoderm, respectively. How Oct4 controls these processes is still unclear. Once again, Oct4 proves to have dual functions, apparently being able to activate genes responsible for differentiation (Frum *et al*., 2013), (Askoy *et al*., 2013) and to repress pluripotency genes (Radzisheuskaya *et al*., 2013).

To add even more complexity to the situation and simultaneously prove how powerful this transcription factor is, Oct4 is also reported to serve as a direct bridge between two distinct differentiated states, through the process of transdifferentiation. This has been seen for transformation of mouse fibroblasts into cardiomyocytes (Efe *et al*., 2011), neural stem cells (Their *et al*., 2012) and neural progenitors (Kim *et al*., 2011). In these cases, expression of Oct4 was induced along with the expression of the other three canonical reprogramming factors. Interestingly, for the transdifferentiation into neural stem cells, Oct4's activity was limited to the initial phase. It has been also described that ectopic expression of Oct4, coupled with a specific cytokine treatment, can transdifferentiate human fibroblasts into multipotent haematopoietic progenitors (Szabo *et al*., 2010).

1.6 Oct4 in induced pluripotency

Reprogramming is a fascinating biological bridge between differentiation and pluripotency, and can be achieved through pluripotent nuclear transfer into somatic cells (Wakayama *et al*., 2006), fusion of somatic with pluripotent cells (Ogle *et al*., 2004) or the ectopic expression of defined factors (Takahashi & Yamanaka, 2006) in a somatic cell. Somatic cell reprogramming to pluripotency occurs through a complex multi-step mechanism, which has been thoroughly studied for the last 10 years, since its discovery. Starting with 24 candidate genes, Takahashi and Yamanaka were able to narrow down to a combination of four genes as a reprogramming cocktail: Oct4, Sox2, c-Myc, and Klf4. Although the process of transcription factor-mediated reprogramming is still far from being completely dissected, it is known that Oct4 is an essential player. This has been proven by the demonstration that Oct4 is the only of the four canonical Yamanaka factors which cannot be replaced by a family member (Nakagawa *et al*., 2008; Jiang *et al*., 2008) and by the fact that exogenous Oct4 or its endogenous activation are necessary for reprogramming (Radzisheuskaya & Silva, 2014). In addition to this, Oct4 is sufficient to induce pluripotency, though not as efficiently or rapidly as traditional combinations of factors (Kim *et al*., 2009a; Kim *et al*., 2009b; Wu *et al*., 2011; Tsai *et al*., 2011; Yuan *et al*., 2011; Li *et al*., 2011; Zhu *et al*., 2010). Moreover, this essential function is conserved evolutionarily (Rossello *et al*., 2013; Lu *et al*., 2012; Tapia *et al*., 2012), and understanding it can shed light on its function in pluripotency itself.

Due to the role of Oct4 in pluripotency maintenance, which has been described, one could think that its function in reprogramming would solely be to activate the pluripotent network, and that all other events would come due to that step. Oct4 also has its role in differentiation induction, which may appear contradictory to its ability to reprogram differentiated cells to pluripotency. But activation of the pluripotent network is only acquired fully in the late phase of reprogramming, with various other processes needing to occur before.

Starting at the initiation phase, Oct4 is reported to be involved in its hallmark, MET, by inhibiting the opposite process – epithelial-to-mesenchymal transition (EMT) (Li *et al*., 2010) – partly in partnership with Sox2 (Wang *et al*., 2013). Interestingly, EMT inhibition is achieved in synergy with BMP signalling (Samavarchi-Tehrani, *et al*., 2010), the influence of which on cell fate in Oct4-dependent lineage commitment has been described above. Oct4 gene activation is also thought to contribute to the dedifferentiation/identity repression of the somatic cells (Shimazaki, 1993; Kumar *et al*., 2012), by acting as a pioneer factor, accessing compact chromatin and loosening it (Soufi *et al*., 2012), binding the genome promiscuously – meaning that ectopic Oct4 binds to different regions from endogenous Oct4 in ESCs. Further highlighting the importance of these mechanisms and of Oct4's action is the fact that H3K9 demethylases

Jmjd1a and Jmjd2c, which are recruited by Oct4 for nucleosome deletion and demethylation, show increased expression with the progress of reprogramming (Takahashi & Yamanaka, 2006), while their absence leads to failure to obtain iPSCs efficiently (Nichols *et al*., 1998).

The non-redundancy of Oct4 in reprogramming has been related to its short variable linker region (Radzisheuskaya & Silva, 2014), mutation of which impairs reprogramming capacity. I have explained before that this is involved in recruitment of helicases from chromatin remodellers NuRD and BAF complexes, reinforcing the connection of Oct4 with epigenetic regulation, which certainly requires further study. The activation of endogenous Sox2 represents the initiation of the final stages of reprogramming, which is followed by full core pluripotency circuitry activation. The aforementioned non-redundancy of function between Oct4 and its family members and the fact that only Oct4 can heterodimerize with Sox2 could lead to the idea that an Oct4-Sox2 interaction would be the key for the non-redundancy and hence the key for Oct4's crucial role in reprogramming. Expectedly, Oct4 and Sox2 are the most frequent pair of interactors out of the four canonical actors of reprogramming (Chen *et al*., 2008). It is possible that the cruciality of Sox2 activation to initiate the late phase could be due to its suggested role as mere activator of Oct4. In that case, what would be crucial in reality would be the activation of endogenous Oct4 (as opposed to retroviral Oct4) and, therefore, the activation of Sox2 before to concretize that action. The activation of endogenous Oct4 is essential as one of the last reprogramming steps is the silencing of retroviral gene expression. Loh and Lim proposed in 2011 that Oct4 and Sox2 cooperate in inhibiting each other and counterbalancing the lineage specifications they would induce alone – the mesendodermal lineage in the Oct4 case and the ectodermal lineage in the case of Sox2. This "seesaw reprogramming model" has its flaws, one of those being the proposed narrow flexibility of Oct4/Sox2 level balance, which in reality can be shifted in ESCs without repercussions in terms of lineage commitment. Without further expansion on the discussion of this model, it is worth mentioning that, as already described, Sox2 null ESCs can be rescued by Oct4, and Oct4 bypasses this factor in other reprogramming strategies as well, such as the heterokaryon-based approach (Pereira *et al.*, 2008). This observation does not fit this model as Oct4 alone would be expected to induce mesendoderm. The greatest proof of principle behind this model is the replaceability of the canonical factors by lineage specifiers (Shu *et al*., 2013), but lineage specifiers have been proposed to succeed to their capacity as pioneer factors or by endogenous gene upregulation. It is likely that this leads to endogenous activation of Oct4, as Oct4's action is not replaceable in reprogramming. In this late phase there is also upregulation of extracellular matrix and cell adhesion proteins, and Oct4 may be also acting here via E-cadherin. E-cadherin was shown to

be a downstream effector of Oct4 in the acquisition of pluripotent adhesion characteristics (Livigni *et al*., 2013).

When the cells become induced pluripotent stem cells (iPSCs), their pluripotency network is equivalent to that of an ESC, and they exhibit similar Oct4 binding configuration (Sridharan *et al*., 2009). To enter pluripotency the levels of Oct4 need to be ESC-like but after establishment cells can be maintained with low levels of Oct4 (Radzisheuskaya *et al*. 2013; Carey *et al*., 2011), suggesting that at this point Oct4's duty is solely to maintain pluripotency, just as in ESCs. In addition to the earlier evidence regarding the action of Oct4 in pluripotency maintenance, Oct4 also displays a lack of redundancy between its family members with regards to pluripotency maintenance (Niwa *et al*., 2002), and its ESC self-renewal capacities are conserved between most of its homologs (Niwa *et al*., 2008; Morrison *et al*., 2006).

The role of Oct4 in reprogramming discussed is summarized in Figure 1.6 below.

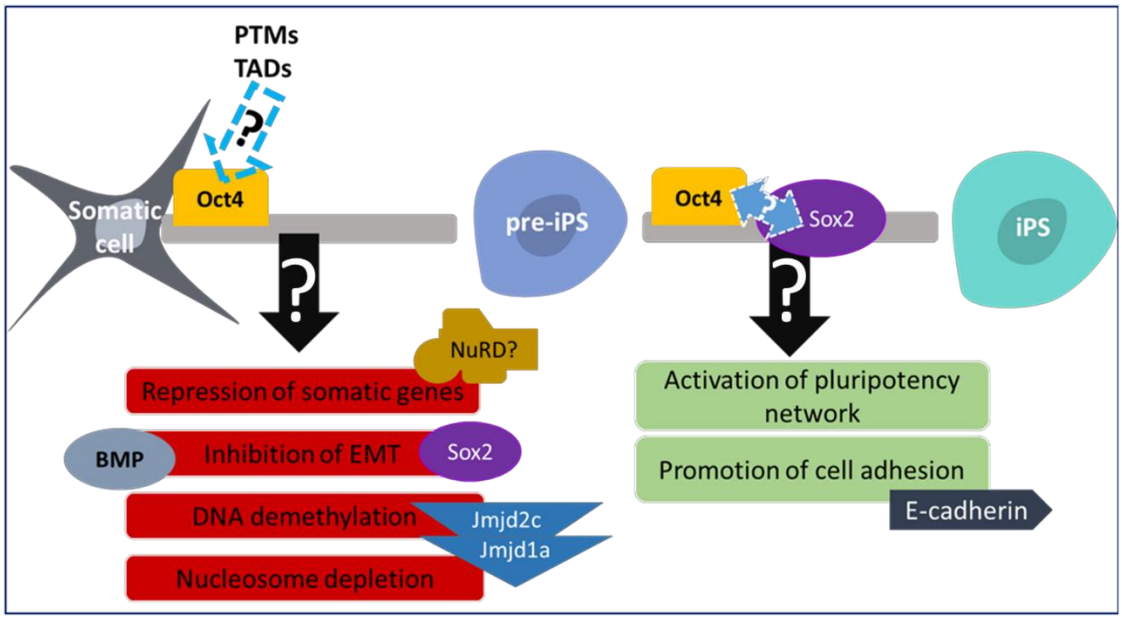

Figure 1.6| Oct4 in reprogramming
Summary of the proposed roles of Oct4 during reprogramming, described and analysed in the previous section, and for which the mechanisms of action are still unknown. Possible partners and regulators for each function are presented.

CHAPTER 2

Purpose of the study

The plasticity of a cell's identity choice is fascinating and has provided a lot of material for scientists to work on for the last decades. It is reliant upon a complex network of transcriptional machinery, epigenetic regulation and environmental signals. Although Oct4 has been identified as a key player, we lack comprehensive understanding of its mechanism of action. The complex nature of Oct4 does not facilitate such investigation, as it is capable of activating or repressing genes with spatial and temporal specificity, partner with endless molecules, and even influence the epigenetic landscape. Adding another layer of complexity, Oct4 can be regulated by its two enhancers differentially and by post-translational modifications. Oct4 is described to have dual functions, being essential for both pluripotency maintenance and determination of lineage specification. Importantly, it was found recently that Oct4 governs cell state transitions in a defined-level manner. In addition, Oct4 is necessary and irreplaceable for the conversion of a somatic cell into a pluripotent one. The most intriguing of all cell states is naïve pluripotency, because it confers the unbiased potential to become any cell of the adult lineages. To further dissect the role of Oct4 in the maintenance of this identity, loss-of-function studies have been performed. These have shown that when Oct4 is repressed in naïve pluripotent cells, they differentiate into TE and primitive endoderm. However, these studies use systems that are slow in the repression of Oct4, thereby only shedding light on late and possibly indirect effects of Oct4's loss, confounding the effort to unveil the mechanism behind its action. Another feature of these systems of slow deletion that compromises the assessment of direct effects is that the cells will transiently pass through an Oct4-low state, which will boost pluripotency, thus not allowing the observation of effects linked directly to the loss of Oct4 protein. Thus, many doors to understanding Oct4's control of naïve identity remained unopened, with innumerable and diverse questions including its connection to chromatin regulation or the molecules it interacts with. These include the following questions: Which are the most direct transcriptional targets of Oct4? Does Oct4 promote the maintenance of certain histone marking to the genome? What is the primary role of Oct4 – acting directly on transcription or acting on a more regulatory level? How fast does the naïve network collapse after Oct4 loss?

The study of the direct consequences of the loss of Oct4 is invaluable to better understand the its function and the network of interactors in which it behaves, and that is what I propose to discuss in this MSc thesis, using the O4AID system previously established by Lawrence Bates (unpublished) – described in Figure 1.6 – consisting of an auxin-based approach (Nishimura *et al*., 2009) to rapidly degrade Oct4 protein from ESCs. In this work I characterized this system for the first time and investigated the impact of Oct4 protein depletion in both the transcriptional

and the epigenetic landscapes of ESCs, looking at histone modifications, in particular. The literature hints at the epigenome being a major part of naïve identity regulation, possibly collaborating with this transcription factor in naïve identity maintenance, given the dramatic difference between epigenetic characteristics of ESCs and differentiated cells. By assessing the immediate epigenetic and transcriptional changes upon acute Oct4 protein degradation, a better understanding of how this protein acts will be accomplished – both in its transcription factor and pluripotency regulator functions – while simultaneously shedding light on the mystery of cell identity and fate decision. As such, the findings from this research could ultimately contribute to a variety of applications, such as disease modelling and pluripotent stem cell-based therapies.

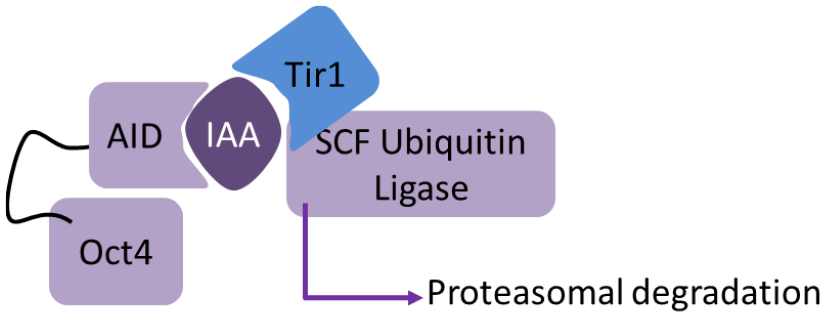

Figure 2 | Auxin-based protein degradation system used to rapid deplete Oct4
Scheme of the auxin-based degron system used to delete Oct4 protein in a rapid fashion. On the left, there is transgenic Oct4 expressed as a fusion protein with an auxin-binding domain. On the right, there is transgenic F-box protein Tir1. F-box proteins confer specificity to the SCF ubiquitination complex in terms of which proteins are targeted to degradation. Tir1 is only expressed in plants and specifically recognizes auxin binding domains only when auxin is bound. Thereby, with the presence of both transgenes, Oct4 is targeted to proteasomal degradation once auxin (IAA) is added.

CHAPTER 3

Materials & Methods

3.1 Cell Culture

Assistance was provided by the WT-MRC Stem Cell Institute Tissue Culture Facility, including providing quality controlled reagents and cell culture training. The facility is composed by Kamila Bulzcak, Jean Thompson and Sally Lees, its Manager.

3.1.1 Cell lines:

Three different established mouse embryonic stem cell (ESCs) lines were used: wild type (WT) E14 Tg2a line, the ZhbTc4.1 line (Niwa *et al*., 2000) and ZhbTc4.1 Oct4AID-Hyg Tir1-Blast (ZhbTO4AID), generated by Lawrence Bates (unpublished).

3.1.2 Culture media

Cells were maintained in either Serum/LIF (SL) conditions (GMEM (Cat. No.:G5154, Sigma-Aldrich) containing 10% Fetal Bovine Serum (Cat. No.:FCS-SA, LabTech), 1x non-essential amino acid (NEAA) (Cat. No.:M11-003, A&E Scientific), 1x Penicillin/Streptomycin (Pen/Strep) (Cat. No.:P11-010, A&E Scientific), 1x sodium pyruvate (Cat. No.:S11-003, A&E Scientific), 0.1 mM 2-mercaptoethanol (Cat.No.:31350-010, Life Technologies), 2 mM L-glutamine (Cat.No.:25030-024, Life Technologies), supplemented with 20 ng/ml murine leukemia inhibitory factor (LIF; Department of Biochemistry, University of Cambridge) or N2B27/2i/LIF conditions (DMEM/F12 (Cat.No.:21331-020, Life Technologies) and Neurobasal (Cat.No.:21103-049, Life Technologies) in 1:1 ratio, 1x Pen/Strep, 0.1 mM 2-mercaptoethanol, 2 mM L-glutamine, 1:200 N2 (WT-MRC Stem Cell Institute) and 1:100 B27 supplement (Cat.No.:17504-044, Life Technologies), supplemented with 20 ng/ml LIF and 2i inhibitors (Biotechnology Center TU Dresden, Stewart Lab): 3 µM CHIR99021 and 1 µM PD0325901.

ZhbTO4AID were maintained in blasticidin (BLAST; Cat. No.:A11139, Life Technologies) at 0.02 ug/ml and doxycycline (DOX; Cat. No.: 198955, MPBiomedicals) at 1 ug/ml, to only express the AID-tagged Oct4 protein.

3.1.3 Routine cell culture manipulations

All cells were maintained in a humidified Sanyo incubator (MCO-18M) at 37°C and 7% CO2 and manipulated in a BioMAT Class II Microbiological Safety Cabinet (Thermo Scientific), using sterile technique.
All cells were cultured on plastic pre-coated with 0.1% gelatin (Cat.No.:G1890, Sigma-Aldrich).
Upon reaching 80% confluency, cells were split in a 1:4-10 ratio. Cells were counted using a Vi-CELL Cell Viability Analyser (Beckman Coulter) when needed. Dissociation for passaging was performed with TrypLE express (TrE; Cat.No.:15090046, Life Technologies), followed by dilution with GMEM and centrifugation at 300 g for 3 minutes to remove the dissociating agent and replating at the required density in fresh media.
All assays were performed the day after the cells were passaged and plated according to the experiments' needs.
Freezing was performed in cryovials, in media supplemented with 10% DMSO (Cat. No.: A3372,0100, AppliChem), for -80°C freezer or liquid nitrogen tank storage.

3.1.4 Oct4 depletion/repression assays

500 µM Indole-3-Acetic Acid (IAA; Cat. No.: I5148, Sigma-Aldrich) was added to the culture media to induce Oct4 protein degradation in ZhbTO4AIDs.
1 ug/ml DOX was added to the culture media to induce Oct4 transcript repression ZhbTc4.1s.

3.1.5 Cell differentiation

For embryoid body (EB) differentiation 1.5×10^6 cells were plated in 10 cm low-attachment dishes in SL conditions lacking LIF supplement (Sw/oL). RNA was collected at days 3, 5, and 7. At day 7, EBs were re-plated on 0.1% gelatin-coated dishes and the obtained outgrowths were analyzed for differentiation into beating cardiac cells. Assistance with the analysis for differentiation into beating cardiac cells was provided by Katie Tremble.
For TE differentiation 1×10^4 cells were plated in 6-well plates, and Oct4 protein was degraded or repressed as stated in 5.1.4.

3.1.6 Live cell imaging

Images of live cells were acquired using inverted Olympus IX51 microscope and Leica DFC310 FX colour camera, videos were acquired using Leica DMI4000B microscope and Leica DFC310FX camera. Images and videos were processed and analysed with Leica software and Fiji-ImageJ2 (Schindelin, J. *et al*., 2012).

3.1.7 Plasmid transfection

For the Tir-1 indirect interaction experiment, plasmids were transfected using Lipofectamine 2000 with either a TetO empty vector (EV) or a CAG Tir1 plasmid, and selected with BLAST.

3.2 Western Blot

3.2.1 Protein extraction and sample preparation

For total protein extracts, cell pellets were incubated for 30mins on ice with RIPA buffer (50mM Tris (Cat. No.:101734734, Sigma-Aldrich) at pH 8, 150mM NaCl (Cat. No.:31434-1KG-R, Honeywell),1% Triton X-100 (Cat. No.:T8787, Sigma-Aldrich), 0.1% SDS (Cat. No.: BP1311-1, Fisher Scientific), 1 mM EDTA (Cat. No.: 15575-038, Life Technologies)) supplemented with proteinase inhibitors (Cat. No.: 05 892 970 001, Roche), and sonicated with Bioruptor200 (Diagenode) at high frequency, alternating 30 seconds on/off for 3 minutes. For histone extracts, cell pellets were incubated for 30mins on ice with acid extraction lysis buffer (20mM Tris-HCl (Cat. No.:1001485627, Sigma-Aldrich) at pH 8, 150mM NaCl, 1mM EDTA at pH 8, 0.5% NP-40 (Cat. No.:101505971, Sigma-Aldrich), 1.5mM PMSF (Cat. No.:P7626, Sigma-Aldrich)), followed by incubation for 10mins at room temperature (RT) with 0.1M HCl.
Protein was quantified using BCA protein assay kit (Cat. No.: 23227, Thermo Scientific). Denaturation of protein extracts was performed at 95°C, using Bolt LDS Sample Buffer (Cat. No.: B0007, Life Technologies) and Bolt Reducing Agent (Cat. No.: B0009, Life Technologies).

3.2.2 Electrophoresis and transfer

Bolt 10% Bis-Tris gels (Cat. No.: NW00102BOX, Thermo Scientific), in a Novex Bolt Minigel Tank (Life Technologies), were used for SDS-PAGE electrophoresis, with the same amount of protein being loaded between different samples. SeeBlue Puls2 Prestained Standard (Cat. No.: LC5925, Thermo Scientific) was used to reference protein molecular weights. Protein transfer was performed using iBlot2 system (Life Technologies), with iBlot Transfer Stacks (Cat. No.: IB23002, Thermo Scientific).

3.2.3 Immunoreactivity

The resulting membranes with protein transferred were blocked using Blocking Solution (5% milk (Cat. No.: 70166, Sigma-Aldrich) in Wash Buffer (0.1% tween-20 (Cat. No.: P1379, Sigma-Aldrich) in PBS (Cat. No: D8537, Sigma-Aldrich)) for 1 hour rolling at RT. Following blocking, membranes were incubated with primary antibody overnight (o/n) at 4°C. The next day, membranes were washed with wash buffer in three 15 min steps, followed by 1h incubation with secondary HRP-conjugated antibody, rolling at RT. Then, membranes were again washed

with wash buffer in three 15 min. The antibodies used were diluted in blocking solution and are listed in the following table.

Table 3.2 | Antibodies used for Western Blot

Antibodies used for Western Blot					
Target	**Species**	**Type**	**Concentration**	**Dilution**	**Company and Cat. No.**
Primary antibodies					
Oct4	Rabbit	Monoclonal	-	1:2000	Cell Signaling Technologies, 83932
Sox2	Rat	Monoclonal	0.5 mg/ml	1:5000	Ebioscience, 14-9811
Nanog	Rabbit	Polyclonal	1 mg/mL	1:2000	Bethyl, A300-397
Esrrb	Mouse	Monoclonal	1 mg/mL	1:1000	Perseus Proteomics
H3K9me3	Rabbit	Polyclonal	1 mg/mL	1:1000	Abcam, ab8898
α-Tubulin	mouse	Monoclonal	1 mg/mL	1:10000	Abcam, ab7291
Total H3	Rabbit	Polyclonal	0.7 mg/mL	1:20000	Abcam, ab1791
Secondary antibodies - conjugated to HRP					
Rabbit	Donkey	Polyclonal	-	1:10000	GE Healthcare, NA934VS
Rat	Goat	Polyclonal	-	1:5000	GE Healthcare, NA935VS
Mouse	Sheep	Polyclonal	-	1:10000	GE Healthcare, NA931VS

The developing of membranes on X-ray films was performed using the ECL Prime detection Kit (Cat. No.: RPN2232, GE Healthcare).

Membrane re-probing with another primary antibody was carried out after washing with wash buffer and blocking. In case of Total H3 reprobing, the washes were performed with mild stripping buffer (15g/L glycine (Cat.No.: 1002284371, Sigma-Aldrich), 1g/L SDS, 1% tween-20) at pH 2.2.

3.3 ImmunoFluorescence

5.3.1 Slide Preparation

5×10^5 cells were added to each pre-gelatinized slide and left to settle for 4-6hours before performing the experiment. After the respective inductions, cells were fixed for 10 mins in 4% paraformaldehyde (PFA, Cat.No.: P/0840/53, Thermo Scientific), and transferred to PBS for storage at 4°C. Lawrence Bates provided assistance during the above described step.

3.3.2 Immunostaining

Fixed cells were permeabilised for 30mins, at RT, in permeabilisation buffer (1% Bovine Serum Albumin (Cat.No.: 1002173737, Sigma-Aldrich), 0.55% Triton X-100, in PBS), followed by blocking for 30mins at RT, in blocking buffer (1% BSA, 0.2% Triton X-100, in PBS), to reduce non-specific antibody binding. After blocking, cells were incubated with primary antibody o/n at 4°C. Then, cells were washed with blocking buffer and incubated with secondary antibody for 1hour at RT, followed by washing again and finally the slides were mounted using Vectashield+DAPI (Cat.No.:H-1200, Vector Laboratories), and covered with a cover glass. The antibodies used were diluted in blocking buffer and are listed in the following table.

Table 3.3 | Antibodies used for ImmunoFluorescence

Antibodies used for ImmunoFluorescence					
Target	**Species**	**Type**	**Concentration**	**Dilution**	**Company and Cat. No.**
Primary antibodies					
Oct4 (N-19)	Goat	Polyclonal	-	1:500	Santa Cruz, sc-8628
Oct4	Rabbit	Monoclonal	-	1:300	Cell Signaling Technologies, 83932
Sox2	Rat	Monoclonal	0.5 mg/ml	1:200	Ebioscience, 14-9811
Nanog	Rat	Monoclonal	0.2 mg/ml	1:300	Ebioscience, 14-5761-80
Klf4	Goat	Polyclonal	0.2 mg/ml	1:300	R&D Systems Inc, AF3158
H3K9me3	Rabbit	Polyclonal	1 mg/mL	1:500	Abcam, ab8898
AID	Mouse	Monoclonal	1 mg/mL	1:200	Medical and Biological Labs (MBL), M214-3
Secondary antibodies					
Rabbit| 488nm	Goat	IgG	2 mg/ml	1:1000	ThermoFisher, A11034
Rabbit| 555 nm	Goat	IgG	2 mg/ml	1:1000	ThermoFisher, A32732
Goat| 488nm	Chicken	IgG	2 mg/ml	1:1000	ThermoFisher, A21467
Goat| 555nm	Rabbit	IgG	2 mg/ml	1:1000	ThermoFisher, A21431
Rat| 647nm	Goat	IgG	2 mg/ml	1:1000	ThermoFisher, A21247
Mouse| 488nm	Donkey	IgG	2 mg/ml	1:1000	ThermoFisher, A21202

5.3.3 Imaging and analysis

Images were acquired with a Zeiss Apotome microscope using a 63X objective. Peter Humphreys, Imaging Facility Manager of the WT-MRC Stem Cell Institute, provided assistance with microscopy. Images were processed and analysed with Fiji-ImageJ2.

3.4 RNA isolation

Isolation of total RNA was performed with the RNeasy Mini Kit (Cat. No.: 74104, QIAGEN), according to manufacturer's protocol for Animal Cells – using QIAshredder columns (Cat. No.: 79656, QIAGEN), using 350 µL of RLT, and performing the DNAse (Cat. No.: 79254, QIAGEN) digest step. RNA concentration was measured with a Nanodrop100 spectrometer (Thermo Scientific).

3.5 RNA sequencing

3.5.1 Library preparation and RNA sequencing

Ribosomal RNA was depleted from 5 µg of total RNA using Ribo-Zero capture probes (Illumina). 10-100ng of depleted RNA was converted into stranded NGS libraries with the NEXTflex Rapid directional RNA-seq kit (#5138-08, Bioo Scientific). Purification of reaction products at each step was performed with Ampure XP paramagnetic beads (Beckman Coulter). Library quantity, size distribution and molarity was assessed by Qubit and the HS DNA assay on the 2100 Bioanalyzer (Agilent). All described above was performed by Dr. Maike Paramor and Joaquin Martinez Herrera, of the Genomic Facility of the WT-MRC Stem Cell Institute. Finally, 150 base pair-end format sequencing was performed on the Illumina HiSeq 4000 at Cancer Research UK Cambridge Institute.

3.5.2 RNA sequencing analysis

Sequencing reads were processed to remove adapters using trim galore (Babraham) and were aligned to the mm10 mouse genome build (UCSC) with Tophat (CCB John Hopkins University). Data was normalized to library size. All described above was performed by Lawrence Bates. The following analysis was performed by Mariana Alves, with the exception for the PCA plot and the list of significant changes hits obtained, generated by Lawrence Bates.

3.6 RNA relative expression quantification

3.6.1 cDNA synthesis

Total RNA reverse-transcription was performed with the SuperScript III First-Strand Synthesis SuperMix (Cat.No.: 11752-050, Life Technologies), according to manufacturer's protocol. The same amount of template RNA between different samples (up to 1 µg, with the total amount determined by the lowest concentration sample) was loaded for each simultaneous reaction, and samples to be cross-compared were reverse-transcribed simultaneously. cDNA was diluted 1:50 in H_2O.

3.6.2 qPCR

qPCR reactions were performed in triplicates. To perform the qPCR reaction, either TaqMan Universal PCR Master Mix (Cat.No.:4352042, Life Technologies) and TaqMan gene expression assays (Life Technologies) or Fast SYBR Green Master Mix (Cat.No.:4385612, Life Technologies) and specific primers were used. The primers and TaqMan gene expression assays used are listed in the tables below.

Table 3.6.2.1 | Primers used for SYBR qPCR

Primers used for SYBR qPCR		
Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')
Cdh3	AGAAGGGTTTGGATTTTGAG	TTGACATCTTTCACATGGAC
Eomes	CCTGGTGGTGTTTTGTTGTG	TTTAATAGCACCGGGCACTC
Gapdh	CCCACTAACATCAAATGGGG	CCTTCCACAATGCCAAAGTT
Hand1	CCCCTCTTCCGTCCTCTTAC	CTGCGAGTGGTCACACTGAT
Tfap2a	CAAGATAGGATTGAATCTGCC	GGTTGAGAAATTCTGCTACTG

Table 3.2.6.2 | Applied Biosystems TaqMan standard assays used for qPCR

Applied Biosystems TaqMan standard assays used for qPCR		
Gene	Dye	Probe ID
Gapdh	VIC	4352339E
Fgf5	FAM	Mm00438919_m1
Gata4		Mm00484689_m1
Nanog		Mm02384862_g1
Rex1		Mm03053975_g1
T-brachyury		Mm01318252_m1

The reactions were performed on StepOnePlus Real Time PCR System (Cat.No.:4376600, Applied Biosystems). Delta Ct values to GAPDH were calculated and relative quantities were calculated as $2^{-\Delta Ct}$. The values were normalized as indicated.

3.7 Chromatin ImmunoPrecipitation (ChIP)

H3K9me3, H3K27ac and H3K27me3 ChIPs were performed as follows:

3.7.1 Sample preparation

Cells ($2x10^7$ for each condition) were dissociated with TrE express, washed in GMEM, resuspended in respective media (SL or 2iL), to which 0.5mL of fixing solution (0,1M NaCl, 1mM EDTA, 0.5mM EGTA, 50mM HEPES at pH 7.5, Protease inhibitors, 4.07% formaldehyde) was

added for 10 mins at room temperature and quenched with 0.785mL 1M cold glycine, then washed 3 times with ice-cold PBS and the respective pellets frozen at -80. Lawrence Bates provided assistance during the above described harvesting steps.
Parallel protein collection was performed at the same time, to validate Oct4 depletion.

3.7.2 ChIP

Pellets were incubated for 10 min at 4°C in lysis buffer 1 (50 mM HEPES (Cat. No.: 1002346455, Sigma-Aldrich) at pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol (Cat. No.: 1017342545, Sigma-Aldrich), 0.5% NP-40, 0.25% Triton X-100), pelleted, incubated for another 10 min at 4°C in lysis buffer 2 (10mM Tris pH 8.0, 200mM NaCl, 1mM EDTA, 0.5 mM EGTA (Cat. No.: E8145-106, Sigma-Aldrich)), pelleted, resuspended in shearing buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8.0) and sonicated to obtain an average DNA fragment size of 250-500 base pairs. Dynabeads Protein G magnetic beads (Cat. No.: 10004D, Life Technologies) were pre-incubated with isotype IgG antibody. Lysates were then diluted 1:10 in dilution buffer (50 mM Tris-HCl at pH 8.0, 167 mM NaCl, 1.1% Triton X-100 and 0.11% Na-deoxycholate (Cat. No.: 1000784206, Sigma-Aldrich)), precleared for 2h at 4°C with IgG pre-incubated Dynabeads, and incubated overnight at 4°C with 1 µg of the respective antibody or 1 µg of an isotype IgG control. The antibodies used are listed in the following table.

Table 3.7.2 | Antibodies used for ChIP

Antibodies used for Chromatin ImmunoPrecipitation			
Target	**Species**	**Type**	**Company and Cat. No.**
H3K9me3	Rabbit	Polyclonal	Abcam, ab8898
H3K27ac	Rabbit	Polyclonal	Abcam, ab4729
H3K27me3	Rabbit	Polyclonal	Millipore, 07-449
IgG	Rabbit	-	Santa Cruz, sc-2027

Before the overnight antibody incubation, 10% of the total chromatin volume was collected to serve as 10% input reference. Dynabeads were blocked with 1mg/mL BSA overnight at 4°C, and then washed 3 times for 5 mins at 4°C with wash buffer 1 (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 0.5 mM EGTA). The overnight antibody incubated chromatin was pelleted to remove protein aggregates and then incubated for 1h at 4°C with blocked Dynabeads. The beads were washed for 5mins at 4°C twice in wash buffer 1, once in wash buffer 2 (50 mM Tris-HCl at pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 0.5 mM EGTA), once in wash buffer 3 (50 mM Tris pH 8.0, 250 mM LiCl (Cat. No.: 101405405, Sigma-Aldrich), 0.5% Na-deoxycholate, 0.5% NP-40, 1 mM EDTA, 0.5 mM EGTA) and twice in wash buffer 4 (50 mM Tris pH 8.0, 10 mM EDTA, 5 mM EGTA). Chromatin was eluted twice for 15 min at 65°C in elution buffer (1% SDS, 0.1 M NaHCO3 (Cat. No.: 101454524, Sigma-Aldrich), and eluates pooled. Samples, including the 10% input collected, were incubated overnight at 65°C with 0.2M NaCl to reverse the cross-linking and purified using the QIAquick PCR Purification kit (Cat. No.: 28104, QIAGEN).

3.7.3 ChIP-qPCR

Pulled down chromatin was analyzed by SYBR green real-time PCR. Enrichment was calculated relative to the IgG ChIP. Primers used for ChIP analysis were designed to best fit known binding peaks of the immunoprecipitated protein to the specific locus, tested for efficiency and are listed in table below.

Table 3.7.3 | Primers used for ChIP-qPCR

Primers used for ChIP-qPCR			
Interest Histone Peak	Genomic Region	Forward primer (5' to 3')	Reverse primer (5' to 3')
K9me3	Airn enhancer	AGAGCGCATTTCGAGTGGAT	GCACTCCTTAGCACAATGGC
	Mndal enhancer	AGGAAATTGGTCAGCAGTGAAC	GACCCATCACTGAGGCTGTG
	Dcpp1 enhancer	TGGAGGTCCCTCCCAATTCT	CTGCCTCCATCCTTGGTGAA
	Cdh3 enchancer	CTGGCTTCCATCTTCAGCCA	GTGAACCTGTGCTGTACCCA
	Tfap2a enhancer	GTCCACTATCTCCAGCGTCTC	CAAACTTAACACCCGCGCA
K27ac	Klf4 enhancer	TGTCCTCTCCACTCCCACAA	AGGAGTGACTGCGTCAAACA
	Esrrb enhancer	CCGGTGCTGTGAAGAAACAC	GGTCGCCTCTTCCCCTATCT
	Nanog enhancer	CACCTCTTCGCTCGGATCTT	CTCCGGGTCAAAGGAGTCTG
	Rex1 promoter	CTCAATGTGCATCTACTGTGCTC	ACTGCATCGATCACCAGCTAC
K27me3	Cdh3 enhancer	GCGAGCGATCAAAGCACATT	CAGCTGATAGATGCCCGAGT
	Pax6 promoter	ACAGGTGCCCAGGAGTCTAA	CAAACTGCTCACGGCTCAAC
	Tfap2a enhancer	CAGTTGCGGCAGTCACTTTC	TTGGAGGCCGGTTCTTTGTT

The primers for Tfap2a-K9me3, Klf4enhancer-K27ac and Esrrb-K27ac were designed by Lawrence Bates.

CHAPTER 4

Results & Discussion

Characterization of the O4AID depletion system

The ZhbTO4AID is a new cell line which has not been described before. To answer the functional questions about Oct4 using the novel AID depletion system and these cells, first the system needs to be characterized and validated to provide biologically relevant results. For that, both the characteristics of the cell line itself and of the IAA induction for Oct4 depletion were examined.

Since these cells have transgenic Oct4, tagged with AID, it is important to assess whether ZhbTO4AID cells proliferate well in culture, which was confirmed (Figure 4.1A). They exhibit a morphology similar to WT ESCs in SL and to the parental ZhbTc4.1 cell line, both in Serum/LIF (SL) and 2i/LIF (2iL) conditions. It should be noted that the morphology in 2iL presented is not similar to WT ESCs in 2iL, suggesting suboptimal naïve culture. This is also observed for the parental cell line, suggesting it is not a problem of the O4AID but of the parental characteristics of the cell line. In SL conditions, a small percentage of the population can spontaneously differentiate, similar to what has been described in WT ESCs. If the AID tag interfered significantly with the function of Oct4, cells would not be expected to proliferate; therefore, this seems to be a viable system in this aspect and the cells appear to behave as WT ESCs.

To deplete Oct4, IAA is added to culture media. After 2h of IAA induction, Oct4 protein is no longer detectable by Western Blot, both in SL and 2iL conditions (Figure 4.1B). O4AID has a slightly higher molecular weight than WT Oct4 (63kDa and 45kDa, respectively). This shows that the system is very rapid and that the depletion occurs with a lot of efficiency. This rapid and efficient depletion was replicated several times, and depletions for further experiments were validated in this way.

In some cases, it is possible to see fainter bands between the O4AID and the WT Oct4 weight (Figure S1A), which might be products of Oct4 degradation. The appearance of these bands in uninduced samples as well reflects that there may be residual Tir1 transgene activity even without IAA being added, but, if that is the case, it is very minimal. Moreover, sometimes a faint band can appear in induced samples after very long exposures (10min) (Figure S1B); however, that should not confound experiments using this system because of the need of a long exposure to see it and because even if a small amount of Oct4 protein is still present, the proportion of Oct4 loss in this system makes it still relevant for the questions proposed to ask.

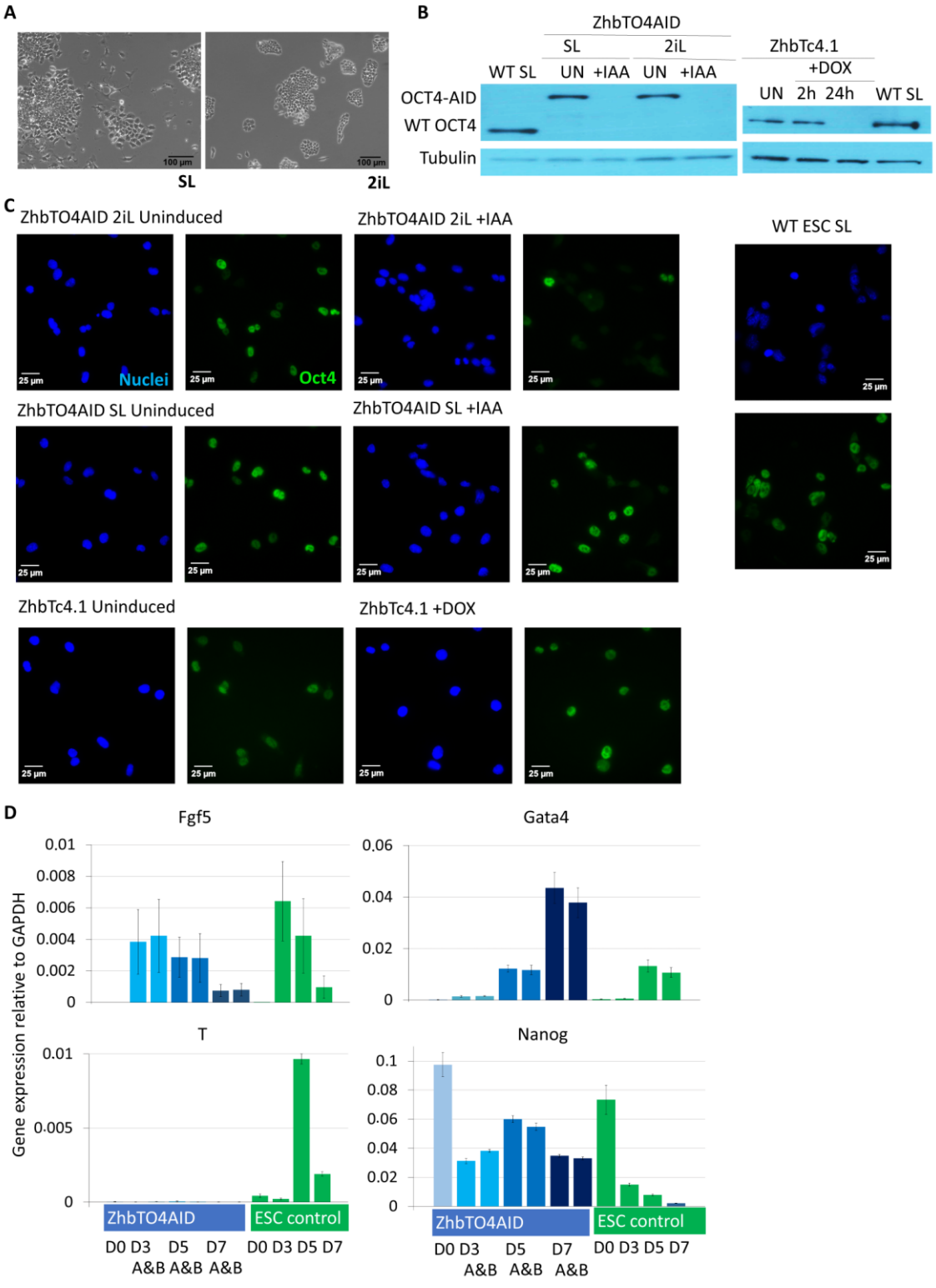

Figure 4.1 | O4AID depletion system is rapid and effective

Figure 4.1 | O4AID depletion system is rapid and effective

A) Phase images representative of ZhbTO4AID cells in SL and 2iL culture conditions.

B) Oct4 Western Blot of ZhbTO4AID cells in SL and 2iL, in uninduced (control) conditions (UN) or 2h after Oct4 depletion (+IAA); of ZhbTc4.1 cells in SL uninduced (control) conditions or after 2h or 24h of Oct4 transcriptional repression (+DOX); and of WT ESCs. Tubulin loading control is presented.

C) Oct4 immunostaining of ZhbTO4AID cells in SL and 2iL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA); of ZhbTc4.1 cells in SL in uninduced (control) conditions or after 2h of Oct4 transcriptional repression (+DOX); and of WT ESCs. Nuclei (blue) are stained using DAPI. Pictures shown are representative of 7 different taken in each condition and are a projection of 10 slices at 63X.Data shown are from 1 of 5 (SL), 3 (2iL, WT ESC) or 2 (ZhbTc4.1) representative experiments.

D) qPCR analysis of ectoderm (Fgf5), endoderm (Gata4), mesoderm (T-Brachyury) and pluripotency (Nanog) marker expression during EB differentiation of ZhbTO4AID cells (blue bars) maintained in 2iL and ESC control (green bars) cells. D0-D7 indicate number of days of differentiation. For ZhbTO4AID cells, biological replicates are shown for D3, D5 and D7 (bars in the same shade of blue, stated as A&B). Data shown are the mean of 3 replicates and error bars are ±s.d. This experiment was performed separately from Figure S1C. The cDNA of the positive ESC controls was kindly provided by Katie Tremble.

The kinetics of Oct4 protein depletion by the O4AID system are much faster compared to the DOX-inducible Oct4 transcript repression of the parental ZhbTc4.1 cell line, in which Oct4 protein can still be detected by Western Blot after 2h of induction (Figure 4.1B). This comparison highlights the utility of the O4AID system, as it is the only system currently capable of degrading Oct4 protein so rapidly and efficiently. This allows the study of the direct effects surrounding Oct4's loss for the first time. Effects of immediate depletion can be traced more closely to the function of Oct4 than late effects, from which is hard to distinguish direct from indirect effects.

To examine with rigor the efficiency of this system it is important to look at the single-cell level to see if cells are homogeneously depleting Oct4. While Western Blot protein detection can be very useful, it only provides information at the population-level. Immunostainings are at single-cell level, therefore allowing the study of the homogeneity of cell behavior. By immunostaining, it is possible to see that in 2iL the vast **majority** of the cells have deleted Oct4 efficiently by 2h of IAA induction (Figure 4.1C), and when looking at the parental ZhbTc4.1 cell line (Figure 4.1C), there is no Oct4 degradation as fast as 2h. The cells which efficiently deleted Oct4 by 2h show the same unspecific background staining as the Oct4-negative cells in the heterogeneous WT ESC population (Figure 4.1C). These findings come to validate the efficacy of this system.

After 2h of IAA induction in SL, the cells do not respond as efficiently, with a high percentage of cells still having Oct4 protein present (Figure 4.1C). This difference of efficacy between culture medium can possibly be explained by a more effective transgene selection in 2iL medium. On the other hand, the permissive and heterogeneous nature of the SL media allows that some cells may be refractory to the action of the Tir1 transgene and selection is not as well maintained as

in 2iL, leading to a high heterogeneity and high percentage of cells where there is no depletion of O4AID. However, the Western Blot suggests efficient Oct4 degradation in SL. This issue with the SL condition shall be bore in mind throughout this thesis, when interpreting results coming from cells cultured in this condition. This shall be optimized in the future so that one could ask questions that are best answered in SL conditions. 2iL medium pharmacologically maintains high expression of many pluripotency factors and could therefore buffer effects of acute Oct4 depletion, so depletion in SL conditions could produce more pronounced immediate effects on the pluripotency network. This could be done by assessing different depletion timepoints or by using a higher IAA concentration. Another feature to validate the Oct4 protein loss in this system that can be done in the future is to show the lack of Oct4 binding to the regulatory regions of genes by Oct4-ChIP.

Nevertheless, regarding this body of work, this difference in efficacy does not confound the interpretation of our results, the majority of which are presented under the very efficient 2iL conditions. It will be made clear in the text when this is not the case.

To further characterize this new cell line, its general ability to differentiate was evaluated by embryoid body (EB) differentiation assay. The gene expression patterns presented by the ZhbTO4AID cells (uninduced – expressing O4AID protein), compared to WT ESCs (Figure 4.1D), indicate that these cells can differentiate into the endodermal (Gata4) and ectodermal (FgF5) lineages, but present an impairment progressing to mesoderm (T). This finding is consistent with the morphology of this cell line on EB outgrowth (Figure S1C) and with the fact that they do not form beating cardiac cells, while both WT ESCs and the parental ZhbTc4.1 cells are capable to form those efficiently (Figure S1C – see also supplementary videos of beating cells). This lack of capacity to form beating cardiac cells has been replicated in three experiments. Moreover, ZhbTO4AID cells are not able to efficiently shut down the expression of pluripotency marker Nanog, as they should in such assay. The cDNA of the positive ESC controls was kindly provided by Katie Tremble.

Although it would be ideal that these cells would follow the usual differentiation pattern, for the purposes of this study and system, which is the evaluation of the effects of acute Oct4 protein depletion, the long-term identity progression is not entirely crucial and does not confound the results. Nevertheless, these impairments may indicate that the cells with O4AID do not behave as WT ESCs, something that is already hinted at by their 2iL morphology being not 2iL-like. It is also possible that these impairments are due to resistance of the transgenes to the natural effects of the differentiation process itself. If the O4AID transgene is not being repressed completely during differentiation, the cells may become trapped in an Oct4-low state which

promotes pluripotency and does not allow them to shut down pluripotency genes properly. This could be confirmed in the future by Oct4-qPCR.

It will be important in the future to compare the gene expression levels of EB differentiation with the pattern followed by the parental ZhbTc4.1 cell line. In the case that this cell line also presents inefficient Nanog shut down and no upregulation of T, it will suggest that this is not a feature of the O4AID transgene but a feature of this cell line itself. However, it is possible that this is not the case, as the parental ZhbTc4.1 cell line was able to form beating cardiac cells, indicative of positive progression to mesoderm. Because of this, in the future it will be of utmost importance to validate the main findings using the O4AID system in a different cell line.

ZhbTO4AID cells succeed in differentiating into TE – exhibiting a flat giant-cell morphology (Figure 4.2A) –, confirming that ZhbTO4AID cells behave according to the published long-term phenotypes of Oct4 depletion. This is an interesting observation, since this acute depletion of Oct4 could have led to a different cell identity as such a system has never been described before in the literature.

In fact, when PCA analysis of the top 2000 differentially expressed genes from Oct4 protein depletion via IAA during a 24h time course is plotted against published RNA sequencing (RNA-seq) data from different cell states (Figure 4.2B) – pre-implantation epiblast, post-implantation epiblast, primitive endoderm, ESCs in 2iL (Boroviak *et al*., 2015) ; and cultured TE stem cells (TSCs) (Calabrese *et al*., 2015) – it is possible to see a progression from an identity closer to the pre-implantation epiblast and 2iL ESCs (naïve pluripotency) into what seems the direction of TE stem cell identity, and away from the normal differentiation route (the post-implantation epiblast). It is important to state that it is not possible to draw immediate conclusions from such plot, because these comparisons are between 3 different studies, with major differences in their methodology. Nevertheless, the facts that the ZhbTO4AID cells follow an actual trajectory during the depletion time course and that there is a clear separation between the three embryonic states, seem indicative that the interpretation of this plot has biological relevance. This Principal Component Analysis (PCA) plot and the respective sequencing analysis to its creation were performed by Lawrence Bates.

When observing the morphological progression towards the TE state upon Oct4 protein depletion via the AID system, the flat giant-cell morphology starts to appear mostly around 3 days of Oct4 depletion (circled in Figure 4.2C) – a common feature between cells whether originally cultured in 2iL or SL –, and by the 4^{th} and 5^{th} days the cells exhibit this identity completely. In SL, it is possible to see some cells refractory to Tir1 activity – exhibiting morphology characteristic of non-depletion of Oct4 – still present at day 5 of Oct4 depletion

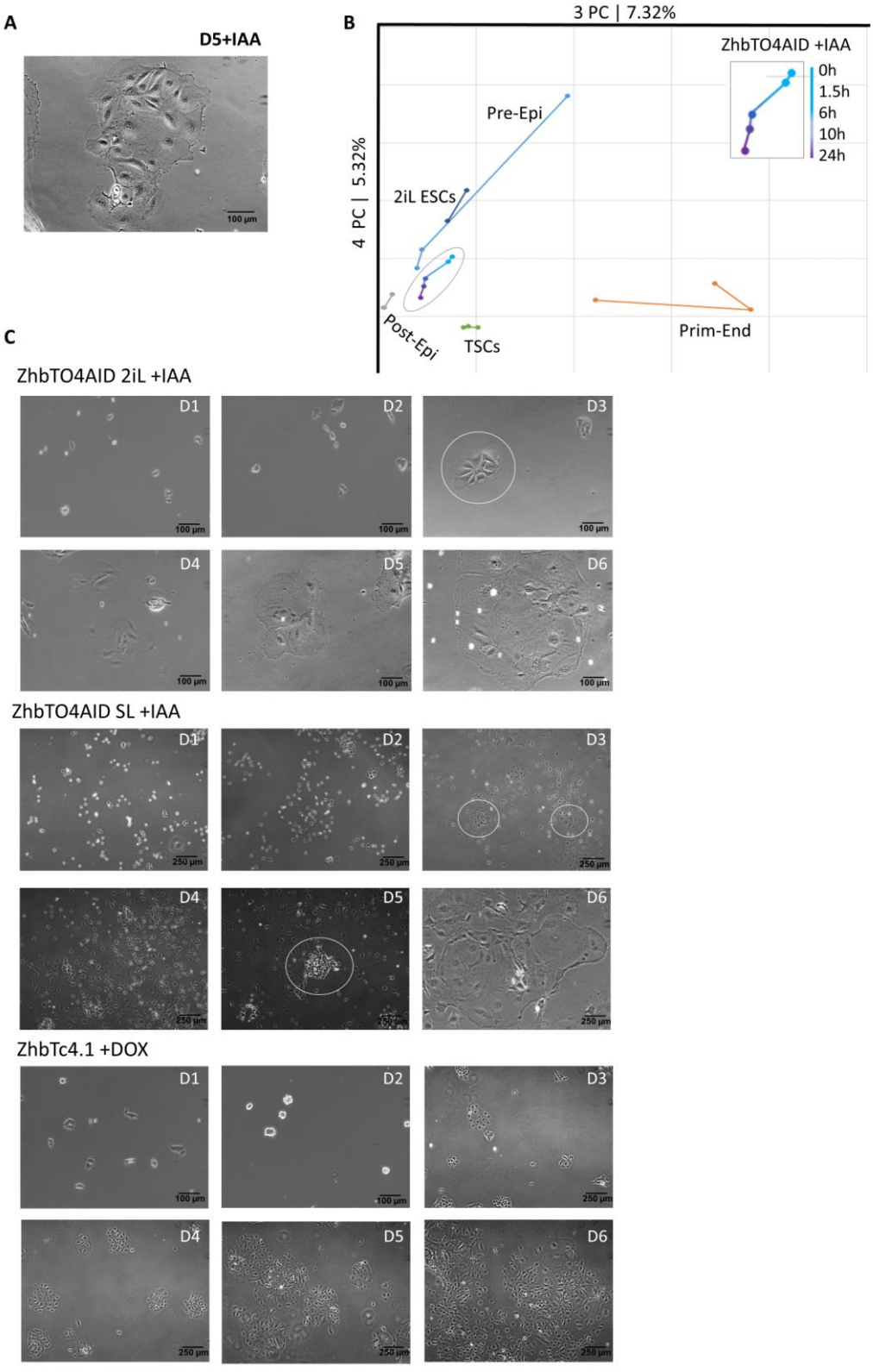

Figure 4.2 | Oct4 depletion via AID system leads to continuous progression into TE in the absence of an Oct4-low state

Figure 4.2 | Oct4 depletion via AID system leads to continuous progression into TE in the absence of an Oct4-low state

A) Phase image representative of ZhbTO4AID cells originally maintained in 2iL culture conditions, after 5 days of TE differentiation assay (SL culture conditions +IAA).

B) Principal component analysis of the 2000 genes most differentially expressed between the following samples: Post-Epi, Pre-Epi and TSCs. This analysis used RNA-seq samples from publicly available published data (Pre-Epi – pre-implantation epiblast; 2iL ESCs; Post-Epi – post-implantation epiblast; Prim-End – primitive endoderm; TSCs – TE stem cells) and RNA-seq data from an Oct4 depletion time course with the AID system. In the upper right corner, insert shows zoom of the trajectory followed by ZhbTO4AID cells (maintained in 2iL) and the respective legend of each timepoint of the Oct4 depletion time course. The percentage of variance of the third component is 7.32% and of the fourth component is 5.32%. This analysis was performed by Lawrence Bates.

C) Phase images representative of TE differentiation assays of ZhbTO4AID cells originally maintained in 2iL culture conditions (ZhbTO4AID 2iL+IAA), ZhbTO4AID cells originally maintained in SL culture conditions (ZhbTO4AID SL+IAA) and of ZhbTc4.1 cells originally maintained in SL culture conditions (ZhbTc4.1 + DOX). D0-D6 indicate number of days of differentiation. Representative features discussed in the result section are circled. In the images of D3 of both 2iL and SL conditions of ZhbTO4AID cells, TE-like cells are circled. The circled image at SL D5 shows a cluster of cells refractory to Tir1 expression, therefore still having Oct4 present and not differentiating into TE. Data shown are from 1 of 1 (ZhbTO4AID 2iL+IAA), 2 (ZhbTc4.1 + DOX) or 3 (ZhbTO4AID SL+IAA) representative experiments.

(circled in Figure 4.2C). This was observed in several independent experiments, and gene expression shows that pluripotency factors Nanog and Rex1 are not efficiently shut down in this medium condition (Figure S2A).

It is very clear that the ZhbTO4AID cells follow a different trajectory from the parental ZhbTc4.1 cell line. The parental ZhbTc4.1 cells, upon DOX induction for repression of Oct4 transcription, exhibit a transiently compacted morphology at day 2, and, by day 3, all cells have transformed to being TE-like (Figure 4.2C). This round morphology at day 2 is suggestive of a possible Oct4-low state, promoter of pluripotency, which could be explained by the slow kinetics of the Oct4 repression of this system. The ZhbTO4AID cells, on the contrary, present a continuous progression towards TE, with no presence of this Oct4-low state (Figure 4.2C). These results reinforce the power of this system, which provides a rapid and efficient way of Oct4 protein depletion, assuring that the effects observed will be direct consequences of the loss of Oct4 and assuring that these observations are not confounded by an Oct4-low state.

Moreover, LIF signaling is not required for the cells to progress towards TE upon depletion of Oct4, as can be seen in Figure S2A. This agrees with the literature which describes that TE formation does not require STAT3 (Vinh Do *et al*., 2015), although STAT3 overexpression, if exceeding certain levels, can induce the transformation of ESCs into TE, via Tfap2c and consequent Cdx2 induction (Tai *et al*., 2014).

Expectedly, if cells are maintained in 2iL and Oct4 protein is deleted, these cells eventually die, as they cannot survive as ESCs without Oct4 (Figure S2B).

These results taken altogether highlight and prove the efficacy of this AID system for Oct4 protein depletion, which will allow the biological answers resulting from this system to be direct consequences of Oct4 loss, and their interpretation will be clear from being compromised by an Oct4-low state. This is the first time that this novel system is characterized and the vast techniques used can assure the confidence in this characterization. With this system, now it is possible to ask relevant biological questions regarding the function of Oct4 that could never be answered before, and understand better how this important transcription factor works, something that has not been elucidated despite the many years this protein has been studied. For these reasons, the O4AID depletion is a very powerful tool that will help me address interesting questions in this body of work, but, most importantly, will open innumerable doors for pluripotency and cell-state transition studies in the present and in the future. Using this novel tool, I will try to study the transcriptional and epigenetic changes arising from the depletion of Oct4 protein, in order to shed light on its action in naïve pluripotency.

Transcriptional consequences of acute Oct4 depletion

The main function of Oct4 is to bind to regulatory regions of genes to affect their transcription. Thus, the first question to arise is: Are there immediate transcriptional consequences upon rapid depletion of Oct4 protein? To answer this, RNA from ZhbTO4AID cells in 2iL was sequenced in order to compare the whole transcriptome of uninduced cells with that of cells 1.5h after Oct4 depletion. Comparing this immediate 1.5h timepoint with later 6, 10 and 24h after Oct4 depletion allows the identification of the most direct Oct4 targets – the ones that show the earliest effects upon Oct4 loss. This identification is something that was not possible before the creation of this system and the ability to deplete Oct4 so rapidly. Moreover, this time course allows the investigation of the transcriptional kinetics upon the loss of Oct4 protein.

Given the crucial role of Oct4 in maintaining naïve pluripotency, the first genes to evaluate were pluripotency factors. Most of the pluripotency factors show only a slight change at 1.5h after Oct4 depletion but at 6h already show significant changes (Figure 4.3A). Since the cells have lost a factor indeed essential to maintain the naïve identity, one would expect these changes to occur in the form of decreasing of gene expression, as they do. It is interesting to observe that the major changes occur at 6h rather than 1.5h but this slightly later timepoint is still very soon after depletion and may reveal direct effects of Oct4 loss, while 1.5h may simply be too early to observe changes in bulk mRNA levels. By 24h, most of the factors have decreased substantially in their expression, as the cells are exiting naïve pluripotency towards the TE lineage. In fact, in the previously shown PCA plot (Figure 4.2B), it is possible to see that the 1.5h sample clusters

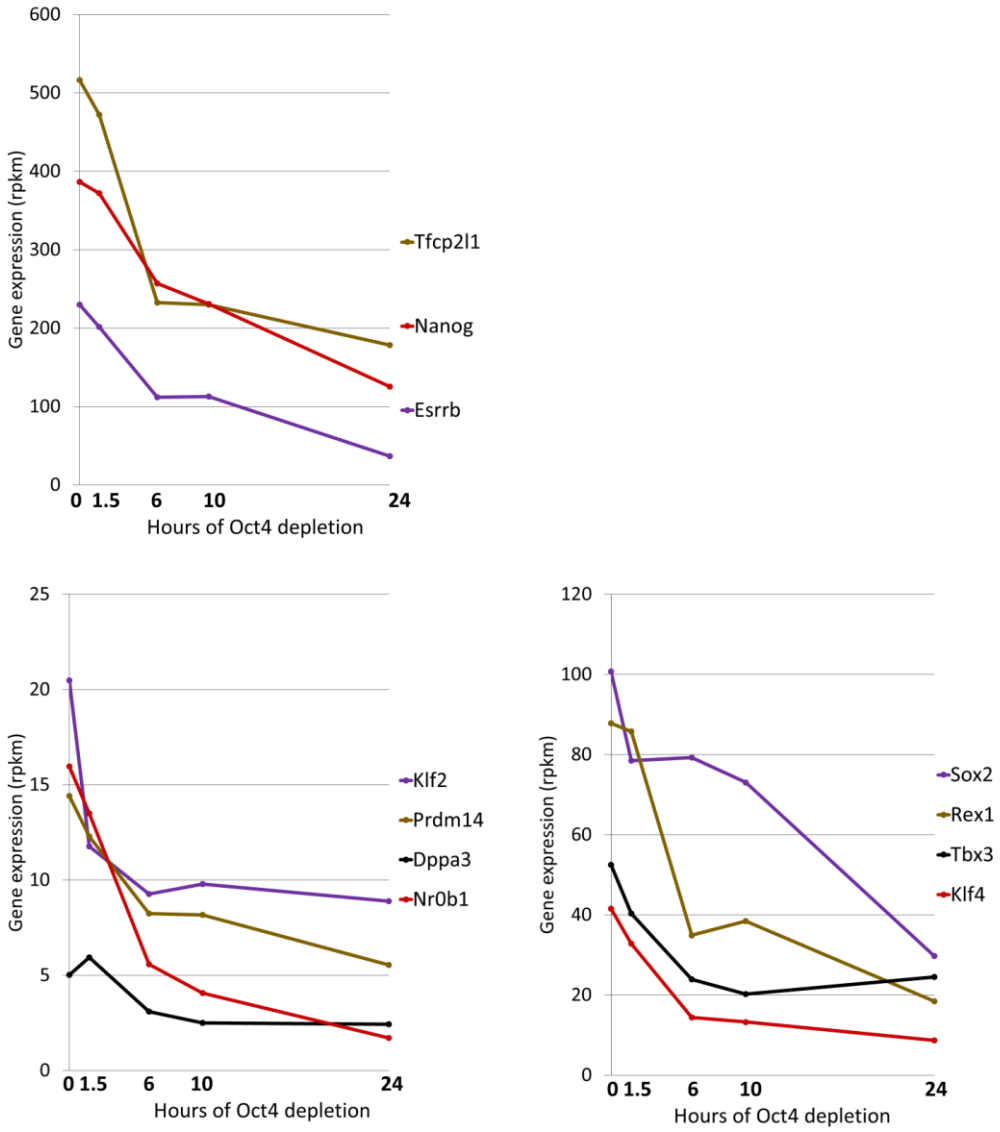

Figure 4.3 | Kinetics of pluripotency network transcription is affected early upon Oct4 depletion

A) Gene expression of pluripotency genes detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

closely to the 0h sample, and that the 6h sample starts to diverge more in the trajectory to TE identity. The word “direct” can have very different interpretations in terms of timing, which may depend on the system being used. In this case, the fact that Oct4 is lost by 1.5-2h justifies the criteria to think that 1.5h could already show direct effects.

By comparing the changes at 1.5h and 6h, it is possible to identify putative direct targets of the function of Oct4, which show the most immediate changes. That is the case for Klf2 (Figure 4A), expression of which halves in 1.5h after Oct4 depletion. This hints at the possibility that from all

the genes composing the pluripotent network, Klf2 might be the most direct target of Oct4, since it is possible to see these transcriptional changes so fast. Klf2 has been described in the literature to be positively regulated by Oct4 (Niwa *et al*., 2009), but this is the first time that is possible to compare the immediacy of effects between different pluripotency factors. Thus, it is the first time that such "hierarchy" in relationship is suggested and that Klf2 is hinted at as one of the most direct targets of Oct4. Interestingly, Klf2 is one of the most efficient single reprogramming drivers from primed to naïve cells, a feature related with the ability to maintain a conserved level of Oct4 (Stuart, unpublished). This is an interesting finding to be assessed further in the future.

The fact that most pluripotency factors only start undergoing major transcriptional changes at 6h could be a result of some other pluripotency factor initially buffering the function of Oct4, trying to maintain the network immediately upon Oct4 loss, but failing to maintain it in the longer term. In fact, long-term effects have been described in the literature as loss of pluripotency transcription factor binding to chromatin by 24h after Oct4 transcriptional repression (King *et al*., 2017). The data presented in this thesis suggests that the transcription of pluripotency genes can occur in the absence of Oct4 protein, meaning that Oct4 may not be essential for the expression of pluripotency genes in the short-term, given a pluripotency-promoting culture media as 2iL. This function could be fulfilled by another pluripotency factor in the short-term, such as Nanog. In fact, Nanog binding at most loci is not reduced by acute depletion of Oct4 (Bates, unpublished), further strengthening this hypothesis. Moreover, Oct4-low cells, which are locked in a pluripotency state, progress to TE if Nanog is knocked down (Silva, personal comms). This suggests that Nanog has some power to substitute Oct4 in the short term, but cannot fully replace it, as it is observed that as early as 6h the pluripotency network is already collapsing. In the future, it would be interesting to further examine this hypothesis.

It is important to note that different mRNAs possess very different half-lives. Thus, genes that would absolutely require the presence of Oct4 protein for their transcription might not appear to drop in their expression by 1.5h in the case that their half-lives are long. This is something to bear in mind in the interpretation of the results of this time course when comparing the expression of different genes over different timepoints. In the future, to assess such immediate effects, it would be interesting to do capture of nascent RNA. This would allow the measurement of changes in nascent RNAs, thus giving a more accurate snapshot of transcriptional dynamics upon the loss of Oct4 protein.

In Table S3B, the hits of genes with significant expression changes in each timepoint of Oct4 depletion are presented, compared to the uninduced control expression level. It is important to note that the RNA-seq experiment was performed without technical replicates. Nevertheless, the trends observed for key factors such as Rex1 and Nr0b1 have been replicated by qPCR (data not shown). In this list can be found several pluripotency factors, with Klf2 being the only present at the 1.5h hits. At 6h, pluripotency factors Tfcp2l1, Rex1, Tbx3, Klf4, Klf2 and Nr0b1 also present significant changes, and by 10h that is the case for Tbx3, Klf4 and Nr0b1.

Additionally, Tet2 is also significantly affected by Oct4 loss throughout the time course. Tet1/Tet2 proteins are 5-methylcytosine hydroxylases and have been described to be recruited by Nanog in reprogramming to promote the expression of certain genes (Costa *et al*., 2013). By 6h, when all the naïve factors start to be repressed, these proteins are also affected. It would be of future interest to look at all of these hits and try to identify any other coordinated changes, in particular the in hits at 1.5h. It would be interesting to validate the expression changes at 1.5h, most directly following Oct4 loss, and confirm that removal of Oct4 is the true cause of these changes, or if some might be indiscriminate. Cyp1a1, one of the hits present in all timepoints and which increases greatly throughout this time course, is a possible case of a false positive caused by the degradation system itself. This gene is expressed transiently during development, but is described to be highly responsive of foreign chemicals (Campbell *et al*., 2005), so it may be responding to the IAA used to remove Oct4, rather than to the loss of Oct4 itself. This could be tested by assessing if IAA addition in cells without the O4AID transgene (so, without removing Oct4) would cause a similar pattern. It is important to mention that the significance criteria chosen ($P<0.001$) allows a confident interpretation of the hits obtained, but it is possible that some interesting targets may be excluded by such tight criteria. Looser criteria suggest that Sox2, a known partner of Oct4, which already may be transcriptionally changing by 1.5h.

The suggestion that the transcription of pluripotency genes could occur in the absence of Oct4 protein for a short-time led the further examination of the consequences of Oct4 loss in the pluripotent network, at the protein level. Interestingly, by 2h after Oct4 depletion, of the pluripotency factors studied, only Sox2 shows a decrease at the protein level detected by Western Blot (Figure 4.4B) in 2iL media conditions. In the case of Esrrb and Nanog, the protein level is not affected until 24h after Oct4 depletion. Moreover, for Nanog and Sox2 in SL, there is still protein expressed at 24h, in agreement with the lower efficiency of this condition in deleting Oct4. Another interesting observation is that Nanog levels seem to be higher in uninduced conditions of the ZhbTO4AID cell line when compared to WT ESCs. This suggests that the

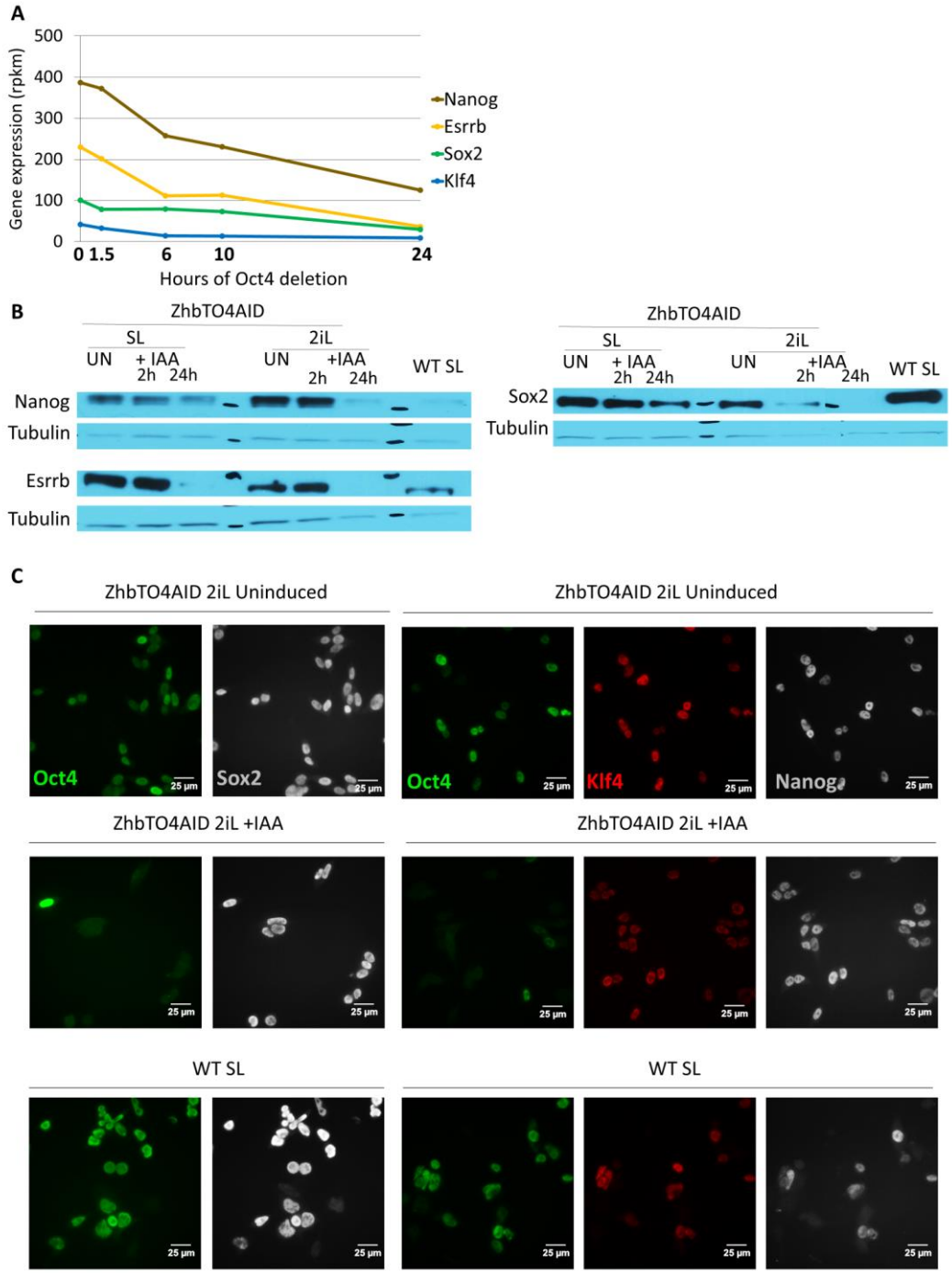

Figure 4.4 | Changes at the protein level of pluripotency factors upon Oct4 acute depletion are not immediate

Figure 4.4 | **Changes at the protein level of pluripotency factors upon Oct4 acute depletion are not immediate**

A) Gene expression of pluripotency genes detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

B) Western Blots of Pluripotency factors (Nanog, Esrrb and Sox2) of ZhbTO4AID cells in either SL or 2iL in uninduced (control) conditions (UN) or 2h or 24h after Oct4 depletion (+IAA), and of WT ESCs. Tubulin loading control is presented.

C) Pluripotency factors (Sox2, Klf4 and Nanog) immunostainings of ZhbTO4AID cells in 2iL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA); and of WT ESCs. Oct4 staining is represented in parallel. Pictures shown are representative of 7 different taken in each condition and are a projection of 10 slices at 63X. Data shown are from 1 of 2 representative experiments in the case of Sox2 immunostaining. Nuclei shown in Figure S4B.

ZhbTO4AID cell line resembles Oct4 low cells to some extent, which could introduce some caveats to the study. The idea of comparing the naïve state with its collapse would be compromised if the starting point were Oct4-low instead of naïve pluripotency. Regarding the assessment of protein level by immunostaining, no changes were observed by 2h after Oct4 depletion in Sox2, Klf4 or Nanog levels (Figure 4.4C). Once again, this suggests that the pluripotency network is able to buffer the effects of Oct4 loss in the short-term. Regardless of this initial ability of the naïve network to cope with Oct4 loss, Oct4 is essential for the maintenance of naïve identity, as the network starts collapsing by 6h.

Moving to another key aspect of Oct4 depletion, the kinetics of TE differentiation in this context were evaluated. It is interesting to assess whether acute Oct4 loss leads to immediate increase in TE markers.

Most TE-related factors are, as expected, very lowly expressed in naïve pluripotency (Figure 4.5A). The increase in expression of most of these genes does not occur immediately neither at 1.5h or 6. One could expect this increase to be much more striking and immediate, but most changes are only more significant by 24h after Oct4 depletion. It should be mentioned that the gene list displayed only includes genes known to be bound by Oct4, to assure their relevance in this context. This comes in agreement with the morphological observations reported before (Figure 4.2), as TE-morphology only appears by day 3. This is also consistent with what was seen in SL during a 6 day TE differentiation (Figure 4.5B), where the greatest increase in expression of these factors is later, around day 6. It would be interesting to compare the transcriptional kinetics of TE differentiation between this system and the parental ZhbTc4.1 cell line, which goes through an Oct4-low state, and assess whether the O4AID system produces faster upregulation of TE markers.

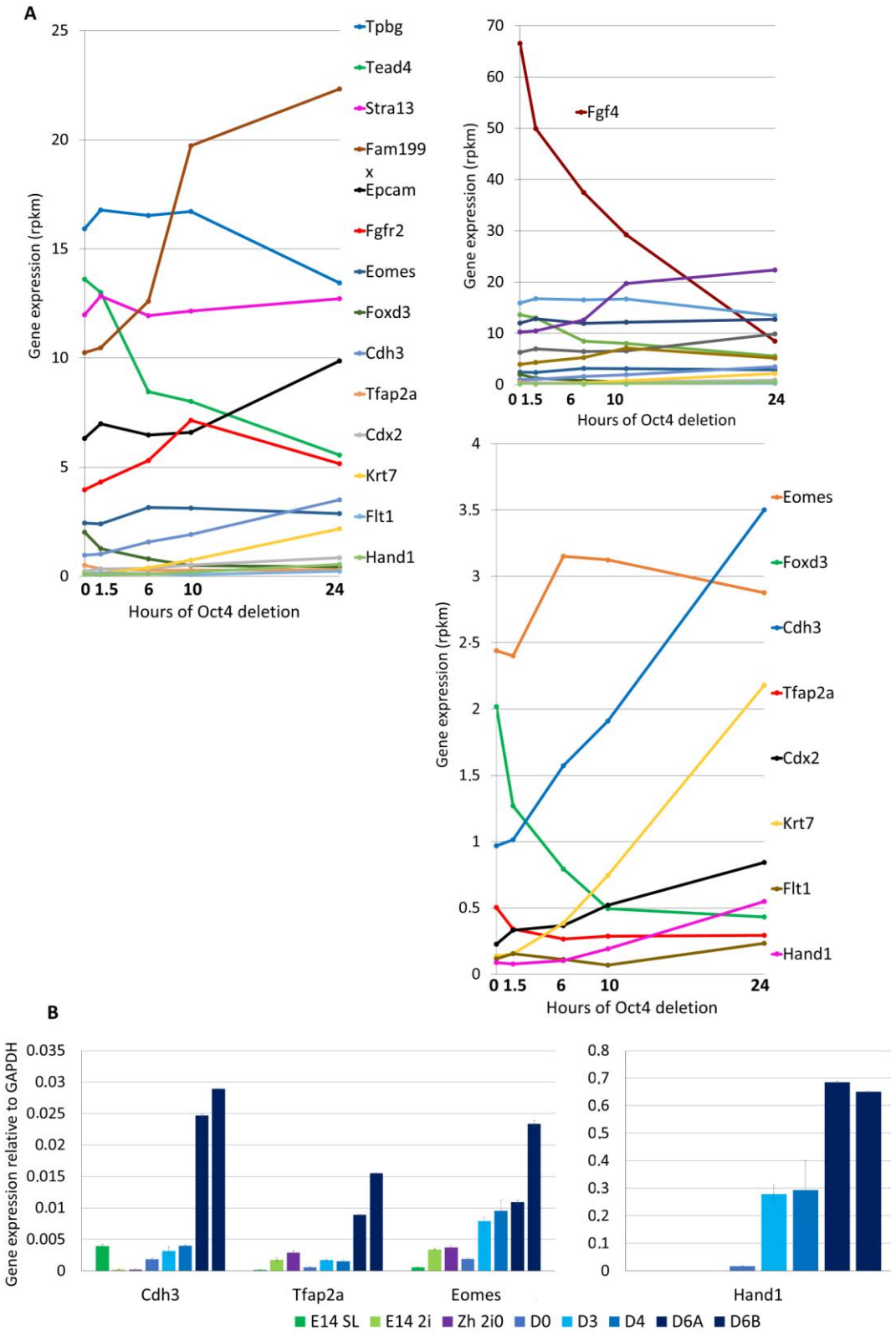

Figure 4.5 | Increase in TE-related gene expression upon Oct4 loss is not immediate

Figure 4.5 | Increase in TE-related gene expression upon Oct4 loss is not immediate
A) Gene expression of TE-related genes which are bound by Oct4 detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.
B) qPCR analysis of TE markers (Cdh3, Tfap2a, Eomes and Hand1) expression during TE differentiation. Legend from left to right: WT ESCs in SL (E14 SL), WT ESCs in 2iL (E14 2i), Uninduced ZhbTO4AIDs in 2iL, Uninduced ZhbTO4AIDs in SL, D0-D6 indicate number of days of differentiation from ZhbTO4AIDs in SL.D6A and D6B are replicates. Data shown are the mean of 3 replicates and error bars are ±s.d.

Taking advantage of this immediate system, it is interesting to observe which are the earliest TE genes to become expressed. Cdx2, Hand1 and Krt7 appear to be the first such genes to be expressed (Figure 4.5A). This hints at a more direct interaction between Oct4 and the expression of these TE genes. It would be interesting to follow up this finding in the future.
Once again, the temporal kinetics of the expression of TE genes comes to agree with the trajectory followed by the PCA plot shown before (Figure 4.2B) and also supports the hypothesis that upon depletion of Oct4, initially naïve pluripotency is being sustained by another factor in place of Oct4.

The most relevant observations for this system are, indeed, the evaluation of the early kinetics of the collapse of the naïve network, and the kinetics of TE differentiation. But given the breadth of this data set, it is possible to ask other questions.

Regarding lineage markers, most genes are lowly expressed to begin with and fluctuate slightly throughout the 24h after Oct4 depletion (Figure 4.6A). These fluctuations occur while maintaining a low level of expression. Only the neural-related gene Nes appears to be more highly expressed and increases expression by 24h, but when compared with the levels of expression shown by more abundant genes, Nes expression is still relatively low.

As has been mentioned before, the Oct4 interactome has been studied several times, producing several diverse and non-overlapping candidate Oct4 partners or interactors. The transcriptional consequences upon Oct4 loss are very variable among the genes of these candidates (Figure 4.6B). It is also important to remember that many of these genes are involved in chromatin remodelling, therefore their expression is expected to be quite ubiquitous. This is an aspect that could be looked into more deeply in the future.

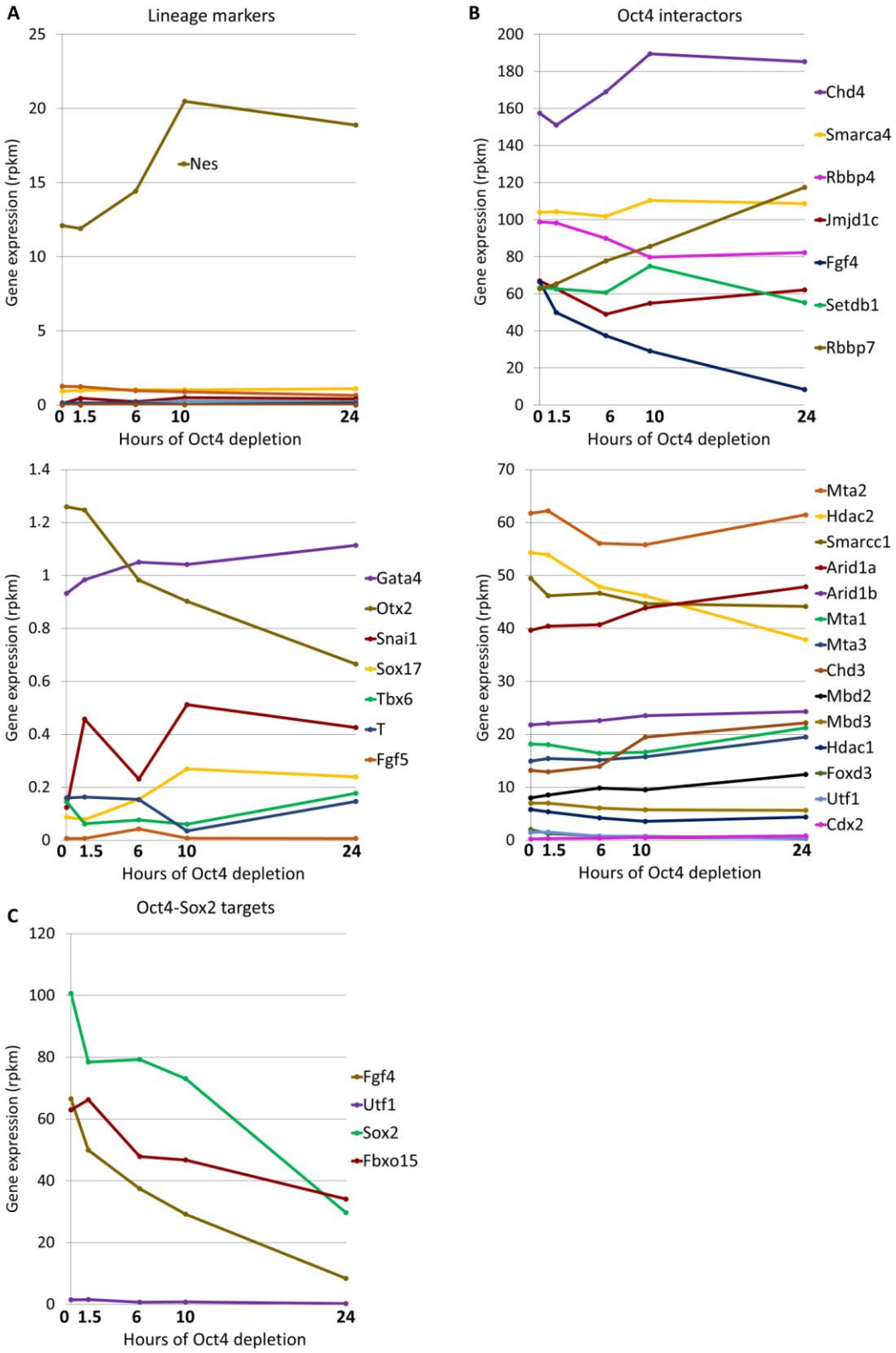

Figure 4.6 | Impact of immediate Oct4 depletion on gene expression of lineage markers, Oct4 interactors and targets of Oct4-Sox2

Figure 4.6 | Impact of immediate Oct4 depletion on gene expression of lineage markers, Oct4 interactors and targets of Oct4-Sox2

A) Gene expression of lineage markers detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

B) Gene expression of described Oct4 interactors detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

C) Gene expression of genes known to be bound by both Oct4 and Sox2 detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

Another interesting aspect to investigate using this system, is the relationship between Oct4 and Sox2. It would be interesting to compare what happens transcriptionally at genes which are co-bound by Oct4 and Sox2 against genes bound solely by Oct4. It could be expected that Sox2 would initially be able to buffer the consequences of Oct4 loss; therefore, there would be more immediate changes to the transcription of factors only bound by Oct4. Another possibility is that Oct4 is required for Sox2 to bind correctly at regulatory regions; therefore, this expression of these genes would be as affected as of genes only bound by Oct4. Regarding factors that are described to be bound by both Oct4 and Sox2 (Rodda *et al*., 2005), Fgf4 presents the more immediate and striking reduction in transcription. It is difficult to ascertain which targets are bound only by Oct4, because this factor very often binds with its partner Sox2. But that would be an interesting future follow-up. Moreover, before I have hypothesized that Nanog could be buffering the loss of naïve pluripotency upon Oct4 degradation, but it is possible that Sox2 is also acting to maintain the network. In the future, it would be of much interest to ChIP Sox2 and understand, for example, if Sox2 is compensating for the loss of Oct4 by increasing its binding at certain loci.

These are only a subset of examples from endless aspects that could be assessed in more detail. Examples of other aspects to consider in the future would be the transcriptional effect of Oct4 loss on the genes belonging to the POU-family, and other Sox-family genes, many of which are known to partner with Oct4.

It will be important to replicate the key findings of this experiment in the future, for example, by qPCR, to confirm these results. If the efficiency of the system is optimized in SL it will also be interesting to assess the transcriptional consequences of Oct4 loss throughout 24h, given the

more permissive nature of this condition, which would allow cells to progress unimpeded towards a TE-like identity, potentially resulting in more pronounced effects.
Overall, the presence of transcriptional consequences as immediate as 1.5h upon the depletion of Oct4 protein attests to the cruciality of this factor in naïve pluripotency maintenance. Most importantly, this system allows the dissection of how this factor affects transcription, allowing the identification of candidates for its most immediate targets.

Epigenetic consequences of acute Oct4 depletion

Now that the transcriptional consequences of Oct4 loss have been evaluated and that the function of Oct4 as a transcriptional regulator has been discussed, I will move to another layer of naïve identity regulation – the epigenetic landscape. It has been noted before that naïve ESCs possess defining epigenetic features that distinguish them from differentiated cells and are thought to play a role in naïve identity maintenance. Moreover, even though a feature of the Oct4 interactome studies is the small rate of overlap between the candidate interactors obtained, it has been mentioned before that a common feature arising from those studies is the interaction of Oct4 with chromatin remodellers. In fact, Oct4 is thought to be a pioneer factor, on which other factors rely to be able to bind to chromatin. It is thus very interesting to ask whether the function of Oct4 in naïve pluripotency is linked to chromatin remodeling. With the AID degradation tool, the existence of an effect upon Oct4 loss on the presence of certain histone marks to the regulatory regions of genes can be examined. If this is the case, then the role of Oct4 in supporting histone marking can be evaluated in more detail, further paving the way to an understanding of the function of Oct4.

Starting with the hallmarks of naïve identity, the expression of pluripotency genes is promoted by, among other events, the presence of active histone mark H3K27ac to their regulatory regions. Moreover, Oct4 is known to bind to these regions as well. Thus, it is relevant to observe the events following the loss of Oct4 protein with regards to H3K27ac presence at this naïve pluripotency loci.

By ChIP-qPCR it was possible to observe that the presence of this active mark at pluripotency genes Esrrb, Nanog, Rex1 and Klf4 decreases upon Oct4 loss (Figure 4.7A). This is reflected transcriptionally by the reduction of expression of these genes by 6h (Figure 4.7B), which can be considered quite early in the 24h Oct4 depletion time course. This suggests that when Oct4 is binding to the regulatory regions of these genes it is acting to maintain the active H3K27ac mark. Thus, when Oct4 is degraded, this mark fails to keep at these loci and these genes subsequently lose their expression, quite dramatically in the case Rex1. This effect, being so immediate,

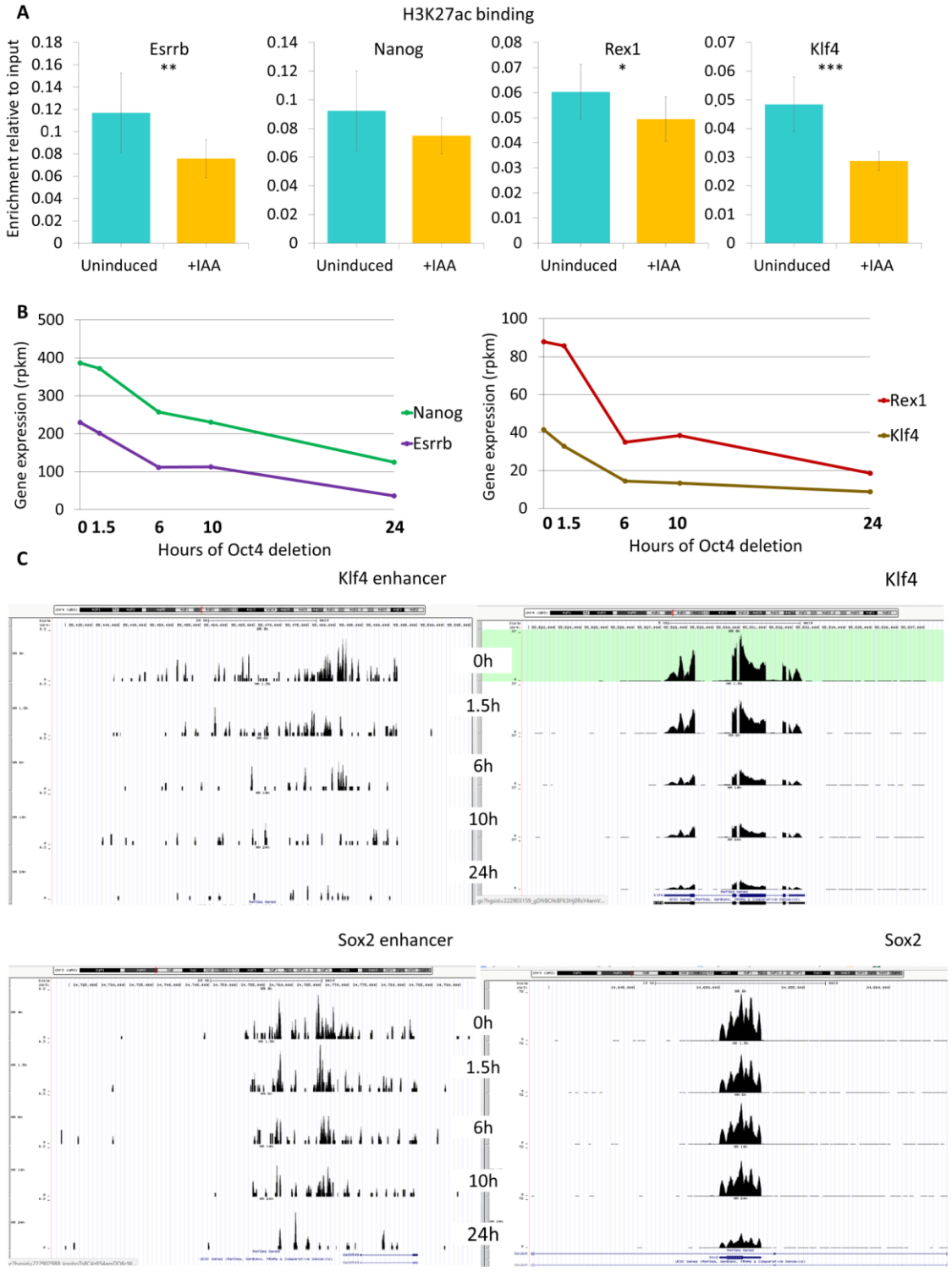

Figure 4.7 | Consequences in regulatory regions of pluripotency genes upon acute depletion of Oct4 precede transcriptional consequences

Figure 4.7 | Consequences in regulatory regions of pluripotency genes upon acute depletion of Oct4 precede transcriptional consequences

A) ChIP-qPCR analysis of H3K27ac binding at the regulatory regions of pluripotency genes (Esrrb (enhancer), Nanog (enhancer), Rex1 (promoter) and Klf4 (enhancer)) also bound by Oct4 in ZhbTO4AID cells maintained in 2iL, in uninduced (control) conditions or 1.5h after Oct4 depletion (+IAA). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 3 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d. Output of Student's T-test analysis of the significance of changes is shown by * = P ≤ 0.05, ** = P ≤ 0.01, *** = P ≤ 0.001. p-values: Esrrb = 0.00917, Nanog = 0.115, Rex1 = 0.0345, = Klf4 0.000171.

B) Gene expression of pluripotency genes detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

C) Genome browser (UCSC) visualization of RNA-seq reads at pluripotency genes Klf4 and Sox2 and the non-coding RNA of their enhancers, over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

suggests that Oct4 may protect the locus from deacetylases, which would act quickly upon its loss. It would be interesting to assess the effect on H3K27ac marking at the regulatory regions of Klf2, as this factor is the more immediate to change upon Oct4 loss (Figure 4.3A). In the future, it would be also of relevance to perform ChIP against other active histone marks at pluripotency genes, such as H3K4me3.

Given that the transcriptional consequences of Oct4 depletion by 1.5h are quite slight, it is fair to say that the effects in the presence of this active histone mark at the genome precede the interference in gene expression. Moreover, there are transcriptional changes at the non-coding level that precede gene expression changes. This is the case for Klf4 enhancer and Sox2 enhancer, both bound by Oct4, where enhancer-RNA quantity starts to decrease earlier than the expression of Klf4 or Sox2 genes, in a 24h Oct4 depletion time course (Figure 4.7C). Thus, upon loss of Oct4, transcriptional changes are preceded not only by a decrease in active histone marks but also by the downregulation of these regulatory non-coding RNAs.

This is the first time that direct consequences of loss-of-function of Oct4 have been studied. Taken together, these observations suggest that Oct4 is acting primarily in chromosomal organization, which will then affect gene expression. Such suggestion hints at a more regulatory role of Oct4 rather than direct locus oriented gene expression promotion/repression. This new work opens doors to discover more about this relationship of Oct4 with the organization of chromatin in the future.

One of the possible future avenues of worth is to assess whether these consequences upon loss of Oct4 are due to changes in the 3D chromosome looping. DNA looping has been described to determine the pluripotency of stem cells (Kagey *et al*., 2010), and components of the Mediator and cohesin complexes MED12 and SMC1 were shown to bind to regulatory regions of Oct4, suggesting that they contribute to the that activation of endogenous Oct4, among other pluripotency genes, in reprogramming to induced pluripotency (Zhang *et al*., 2013).

Oct4 could be interacting directly with the cohesion-mediator complex to act as a maintainer of the correct looping of chromatin for RNA polymerase to be recruited properly. If that is the case, then the decrease in expression of the non-coding regulatory regions of pluripotency factors upon the loss of Oct4 could be mediated by this complex. It could be that, these complexes that support chromosomal organization start to undergo structural changes upon acute Oct4 depletion, which would lead to impairment in RNA polymerase recruitment, which would then lead to a decrease in the transcription of such loci where this would occur. To further confirm this hypothesis and examine if there are changes to chromatin structure upon Oct4 loss, chromosome conformation capture (3C) could be used (Dekker *et al*., 2002).

With the loss of Oct4, the identity of the cell shifts from the naïve state to a TE-like state. This transition is blocked during naïve identity maintenance by the presence of repressive histone marks such as H3K9me3 at TE-related loci, which prevent their expression. These regions are also bound by Oct4. It is, then, of much interest to interrogate whether H3K9me3 presence would undergo any changes as rapid as 2h after Oct4 protein loss, remembering that the transcriptional effects observed are not as immediate.

Using ChIP-qPCR it was possible to observe that, upon Oct4 loss, there is a striking decrease in the repressive marking of regions by H3K9me3 - regulatory regions of genes which are related with TE and where Oct4 is usually bound, such as Cdh3 and Tfap2a (Figure 4.8A & S8A). It is surprising to see such a dramatic effect in such a short period, which implies a direct relationship between the binding of Oct4 at these loci and the presence of H3K9me3. These changes are reflected transcriptionally later in the case of Cdh3, of which expression increases over the 24h course of Oct4 depletion (Figure 4.8B). Tfap2a maintains a low level of expression throughout the entire time course.

In normal conditions, in the naïve state, Oct4 could be promoting this repressive marking by shaping the conformation of the complexes bound to DNA in agreement to protect demethylases from acting at this locus. The structure of these complexes could be disrupted if Oct4 protein is not present, leading to the impairment of H3K9me3 maintenance. Oct4 could also be recruiting chromatin modifiers to promote the trimethylation of H3K9 at these loci. It

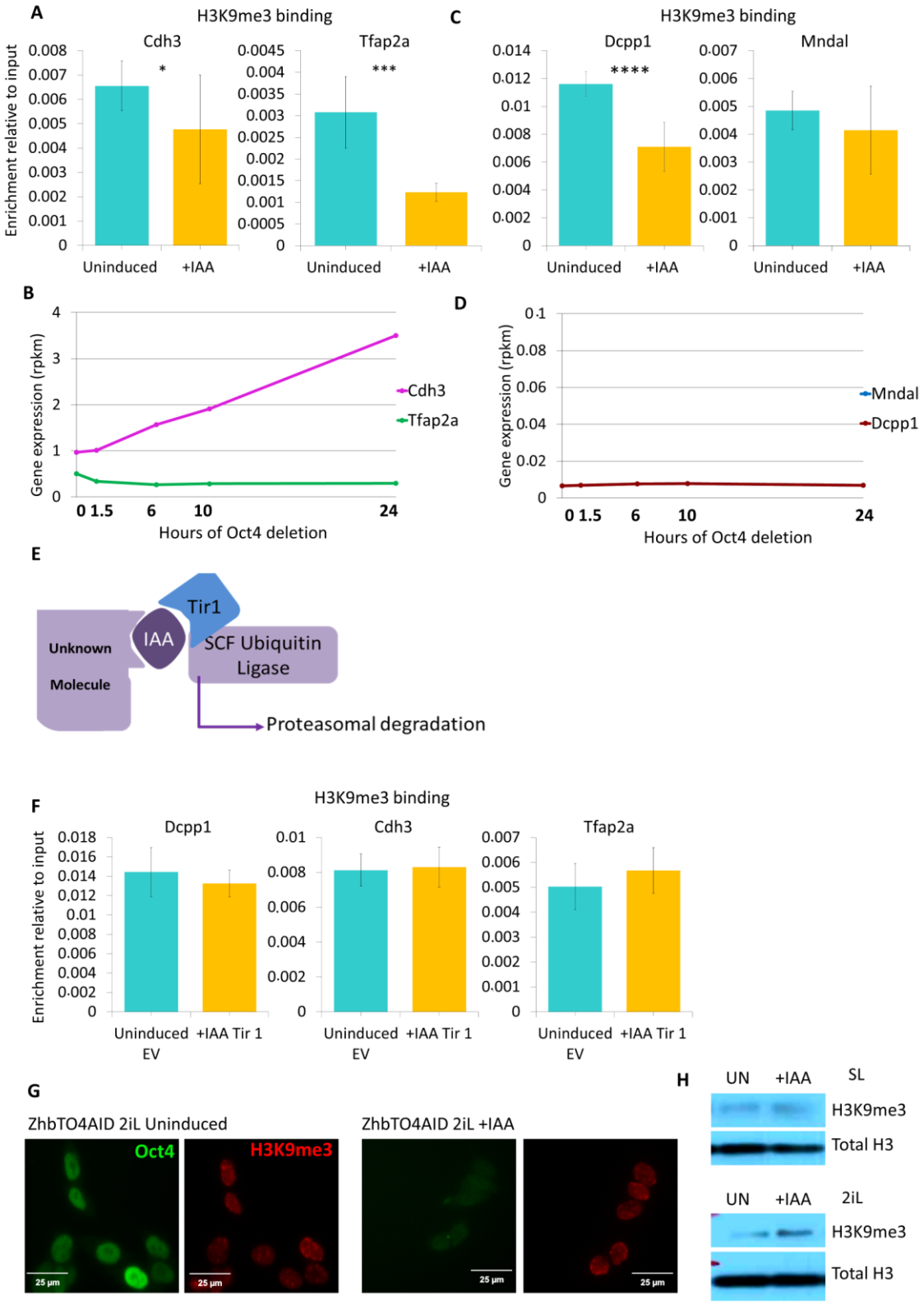

Figure 4.8 | Oct4 protein depletion results in striking decrease of H3K9me3 binding regardless of Oct4 occupation binding to respective loci

Figure 4.8 | Oct4 protein depletion results in striking decrease of H3K9me3 binding regardless of Oct4 occupation binding to respective loci

A) ChIP-qPCR analysis of H3K9me3 binding at the regulatory regions of TE genes (Cdh3 (enhancer) and Tfap2a (enhancer)) also bound by Oct4 in ZhbTO4AID cells maintained in 2iL, in uninduced (control) conditions or 1.5h after Oct4 depletion (+IAA). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 3 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d. Output of Student's T-test analysis of the significance of changes is shown by * = P ≤ 0.05, *** = P ≤ 0.001. p-values: Cdh3 =0.0508, Tfap2a = 0.000112.

B) ChIP-qPCR analysis of H3K9me3 binding at the regulatory regions of genes not bound by Oct4 (Dcpp1 (enhancer) and Mndal (enhancer)) in ZhbTO4AID cells maintained in 2iL, in uninduced (control) conditions or 1.5h after Oct4 depletion (+IAA). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 3 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d. Output of Student's T-test analysis of the significance of changes is shown by **** = P ≤ 0.0001. p-values: Mndal = 0.243, Dcpp1 = 1.842E-05.

C) Gene expression of TE genes Cdh3 and Tfap2a detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

D) Gene expression of MndaI and Dcpp1 detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

E) Schematic representation of possible (but refuted in Figure 4.7F) indirect interactions of the Tir1 transgene in the AID system. On the left, there is an unknown molecule that would bind to IAA, and on the right, there is transgenic F-box protein Tir1, which targets IAA to proteasomal degradation.

F) ChIP-qPCR analysis of H3K9me3 binding at the regulatory regions of Dcpp1 (enhancer), Cdh3 (enhancer) and Tfap2a (enhancer) in ZhbTc4.1 cells transfected with an empty vector or ZhbTc4.1 cells transfected with the Tir1 transgene after 2h of IAA induction, maintained in SL. Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 3 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d.

G) Oct4 and H3K9me3 immunostainings of ZhbTO4AID cells in 2iL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA). Pictures shown are representative of 7 different taken in each condition and are a projection of 10 slices at 63X. Data shown are from 1 of 3 replicate experiments.

H) H3K9me3 Western Blot of ZhbTO4AID cells in SL (upper panel) and 2iL (lower panel), in uninduced (control) conditions (UN) or 2h after Oct4 depletion (+IAA). Total H3 loading control is presented. The membrane was stripped before assessing Total H3 immunoreactivity.

has been noted before that Oct4 interacts with H3K9 demethylase Jmjd2c, H3K9 methyltransferases SetDB1 and chromatin-modifying repressor complexes NuRD an SWI/SNF. However, for Oct4 to change these enzymes' activity it would need to affect either their gene expression, or their localization, or the gene expression of some molecule responsible for either effect. Since 2h was shown to be too early to observe major transcriptional consequences, this hypothesis is less likely. In fact, our RNA-SEQ data shows that neither Jmjd2c or SetDB1 suffer significant transcriptional changes 1.5h after Oct4 depletion nor later on (Figure S8D). There is another hypothesis, where if Oct4 is usually sequestering another transcription factor or molecule, upon its depletion this molecule can start acting structurally and maybe result in these immediate effects.

The reduction in H3K9me3 is stronger in S/L conditions (Figure S8A) than in 2iL conditions (Figure 4.8A), which could be explained by the more permissive nature of S/L compared to the tight and strong naïve maintenance of 2iL. It is possible and probable that the loss of Oct4 can be initially buffered by the optimal naïve environment that 2iL medium heavily promotes. Moreover, SL signaling required for formation of TE is absent in 2iL, resulting in a slighter trend. Thus, assessing effects as rapid as 2h, these will be less pronounced. Nevertheless, one shall bear in mind the lower efficiency of the depletion in SL, as apparent by the immunostaining results described earlier, even though the Western Blot validation suggests efficient Oct4 degradation (Figure S7B).

Next generation high throughput sequencing analysis would be useful to expand these results and assess their relevance. One could evaluate, for example, whether this striking H3K9me3 presence decrease is occurring at the majority of the TE-related loci, reinforcing the participation of Oct4 in this process. One could also look into other genes of interest repressed in the naïve state and see if this repression is also lost in such a rapid manner. For example, it would be interesting to look at the epigenetic repression of other lineages. It could be the case that Oct4 acts to block differentiation into many lineages, but that there is less redundancy in the repression of the trophectoderm programme, explaining why the identity of the cell follows this route and not another upon Oct4 loss. This could be confirmed if the rapid effect of the loss of Oct4 in the repressive marking of genes is revealed to be more striking at TE-related targets rather than other lineage-related loci upon Oct4 depletion.

It is, then, of future interest to perform next generation high throughput sequencing of these ChIP experiments. After obtaining this information, it will be possible to further dissect how Oct4 acts, by, performing ChIP for factors that are good candidates to be players in this phenomenon (such as chromatin remodellers), for example, and see if their binding to the genome change after Oct4 protein depletion. Additionally, sequential ChIP could be performed to assess possible interactions between Oct4 and other factors, binding at the same region to act in synergy.

Very surprisingly, the same striking reduction in H3K9me3 marking 2h after Oct4 protein loss is observed at loci where Oct4 does not usually bind. This was the case for Dcpp1, Mndal and Airn (Figure 4.8C& S8B). All of these genes are lowly expressed both in uninduced conditions and over the 24h Oct4 depletion time course (Figure 4.8D & S8C). This is not in disagreement with the ChIP data, as gene expression is regulated by various mechanisms, such as DNA methylation, so it is possible that one particular change won't affect the overall regulation of transcription. These results were surprising due to the fast nature of effects in places which, not having Oct4

bound, are thought to be more distantly and indirectly related. Performing ChIP for loci where Oct4 does not bind is a useful experiment that could be done in the future for the H3K27ac and H3K27me3 ChIPs, in order to better dissect the impacts of Oct4 depletion.

The interference from some other IAA interactor in this experiment was most likely, so I wanted an appropriate control experiment to assess if this was the case. Thus, upon these surprising results, firstly, it was confirmed that this was not a result of indirect effects, coming from unspecific binding of IAA to any other molecule rather than the auxin-binding domain fused in Oct4, resulting in the degradation of such molecule (Figure 4.8E). This was done by comparing H3K9me3 presence at the loci of interest between the parental ZhbTc4.1 cell line transfected with an empty vector or transfected with just the Tir1 transgene and not the O4AID transgene. In this way, if IAA were causing the dysregulation of another factor – whether by Tir1-induced degradation or any other mechanism – we would expect to see a similar decrease in H3K9me3 presence. When comparing the presence of H3K9me3 at the loci tested above, no differences can be seen. This confirms that the system is not having indirect effects that confound the analysis of repressive marking by H3K9me3 (Figure 4.8F).

At the protein level, no changes can be seen in the foci of H3K9me3 by immunostaining (Figure 4.8G), and by Western Blot the hypothesis of total protein level decrease is excluded (Figure 4.8H). This agrees with the literature, which shows that there is no difference in total amount of H3K9me3 present between the ICM and TE (Huang *et al*., 2007). So, it does not appear to be the case that this effect is a result of a decrease in the global level of H3K9me3 upon acute Oct4 loss.

Having confirmed that these results are reliable, it would be valuable to interrogate how Oct4 protein loss can be affecting the histone marking of regulatory regions of genes that it does not usually bind to. This suggests a long-distance effect, and for this to be seen in such a rapid fashion is very intriguing. Once again, Oct4 can be acting directly through chromatin modifiers, however this may not be highly probable due to the fast nature of the effect. Maybe in this long-distance case, the hypothesis of a sequestered factor, described above, seems a more probable cause of Oct4's loss affecting sites where Oct4 usually is not bound. This effect would be not locus-oriented but due to the desequestration of an enzyme that would subsequently lead to a decrease in trimethylation of H3K9 independently of direct Oct4 binding. There is also the possibility that Oct4 is binding in the vicinity and having more direct effects on H3K9me3 at these loci, because Oct4 is known to bind at a multitude of places in the genome and Oct4 ChIP-seq has not been performed in this cell line yet. Being possible that in this cell line the binding places

of Oct4 are slightly different that the ones published before, another possibility is that these sites are actually strongly bound by it. Nevertheless, these hypotheses are no more than speculations that need to be assessed and answered with scientific experiments, such as Oct4 ChIP-seq or sequential Oct4 ChIP screening for a putative candidate to be the sequestered molecule.

Even though these control regions were chosen exclusively based on the presence of H3K9me3 and absence of Oct4, it is worth noting that they do relate to early embryogenesis. Dcpp1 is present in the mouse oviductal lumen and has been described to promote the growth of preimplantation embryos (Lee *et al*., 2006). Moreover, Airn is known to silence insulin-like growth factor 2 receptor (Igf2r) – vital for the regulation of embryonic growth (Wang *et al*., 1994) – expression during differentiation of ESCs (Santoro *et al.*, 2013). Igf2r is expressed fairly evenly throughout the 24h depletion time course (Figure S8C). Nevertheless, these sites are not described to be related with TE.

Given that this is the case, next generation high throughput sequencing in the future will be particularly useful to understand how widespread this effect of loss of Oct4 is at sites where it doesn't bind, and if the loss of Oct4 is causing global H3K9me3 presence reduction, or a global shift in the places this histone modification is located. Notably, a small shift in the location of H3K9me3 of as little as 300bp could lead to complete loss of detection due to the site-specificity of qPCR. There is the possibility that it is a very small number of loci where Oct4 usually is not bound that suffer this dramatic effect. It will be important to look at loci not involved in gene regulation that are marked by H3K9me3, such as pericentromeric repeats, to assess whether this reduction in H3K9me3 is an indiscriminate effect of Oct4 loss. Though the immunostaining and Western Blot data suggest that this will not be the case, in the event of the confirmation of such direct effect of Oct4 in H3K9me3 binding being global, there is a lot to unveil and discover about the interaction of this transcription factor with the epigenetic machinery, proving itself once again as a powerful jack of all trades and master of naïve cell identity regulation.

The striking direct effect of Oct4 loss in the marking of H3K9me3 shows that this transcription factor is needed for the maintenance of the correct level of this repressive histone mark in ESCs, thereby maintaining the right pattern of gene repression. These results once again hint at the link between the interaction of Oct4 with chromosome organization, since effects at regulatory regions precede the transcriptional consequences. Even though these results need further follow-up for a detailed and mechanistic view of the influence of Oct4 on cell identity, these studies can now be done thanks to the existence of the O4AID depletion system, which proves

itself as a very important tool that can be used for a multitude of studies. Altogether, these will surely be highly informative about the nature and behavior of this transcription factor. And we might learn that the primary function of Oct4 in fact relates to chromosome organization.

So far, the results described concern the epigenetic marking of pluripotency genes and TE-related genes, but it is also interesting to see what happens to lineage markers of the three embryonic germ layers. One of the defining features of naïve pluripotency is the presence of lineage specifiers which are bivalently marked by H3K27me3 and H3K4me3, which have been described before. Upon loss of pluripotency, an important event is the resolution of bivalent genes into on or off states, depending on which mark remains. One of the histone modifications characteristic of bivalent marking is H3K27me3, which represses gene expression. It would be expected that with the loss of Oct4 some genes would lose this mark, especially genes which are related to TE and need to be activated.

Using ChIP-qPCR in SL cells, it was possible to observe that, upon Oct4 loss, there is a significant decrease in the presence of the repressive mark H3K27me3 at bivalent genes which are also bound by Oct4. This is the case for TE-related Tfap2a and Cdh3 and neural lineage-related Pax6 (Figure 4.9A). Once again, to see such a rapid effect comes to strengthen the hypothesis that Oct4 is a strong definer of histone marking to regulatory regions of genes which are crucial for the instruction of cell identity. As described before, the gene expression of Tfap2a is very low throughout the 24h depletion time course, but both Pax6 and Cdh3 show transcriptional effects corresponding to the loss of repressive marks (Figure 4.9B). These transcriptional effects occur later than the immediate epigenetic effects, suggesting once more that Oct4 can be more than a transcriptional regulator and maybe acts primarily on chromosome structure, thus when it is lost, histone marking is lost.

In the future, it will be interesting to assess what are the consequences of acute Oct4 depletion on the marking of active H3K4me3 at these bivalent loci. It is expected to stay the same in genes which lose H3K27me3, promoting their expression

To further evaluate the role of Oct4 in bivalent marking, it is natural to wonder which of a list of 257 described bivalent genes which are bound by Oct4 (Marks *et al*., 2012, Buecker *et al*., 2014) would undergo significant changes from the RNA-seq 24h time course of Oct4 depletion. Of the eight genes that showed significant changes, four increase their expression, two decrease and others present relatively minor variation (Figure 4.9C). It would be very interesting in the future

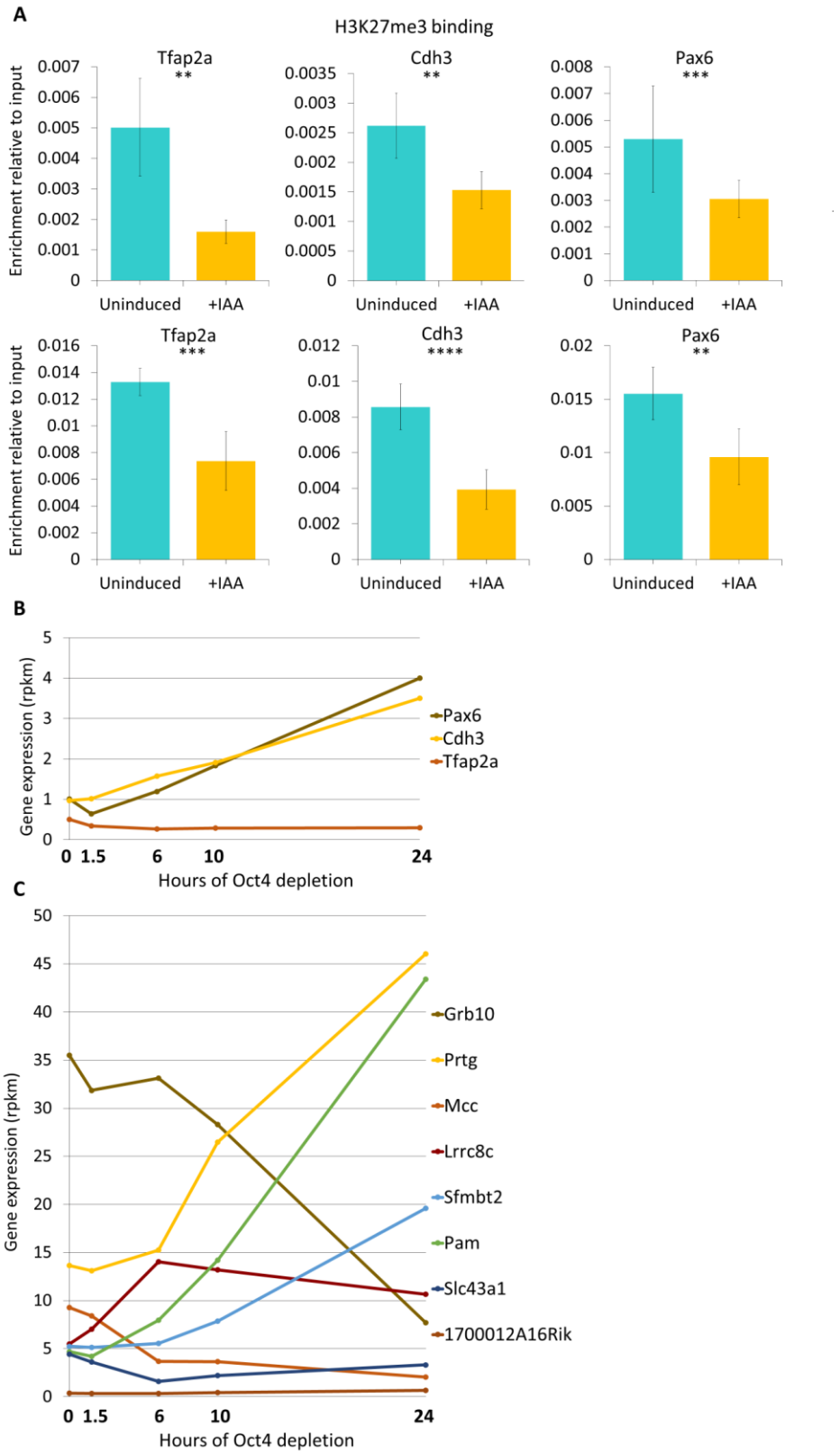

Figure 4.9 | Bivalently-marked genes undergo changes in gene expression and H3K27me3 binding profiles upon rapid Oct4 loss

Figure 4.9 | Bivalently-marked genes undergo changes in gene expression and H3K27me3 binding profiles upon rapid Oct4 loss

A) ChIP-qPCR analysis of H3K27me3 binding at the regulatory regions of bivalent genes (Tfap2a (enhancer), Cdh3 (enhancer) and Pax6 (promoter)) also bound by Oct4 in ZhbTO4AID cells maintained in SL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA). Two replicate experiments are shown (top and bottom panels). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 2 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d. Data shown for Pax6 (upper panel) are the mean of two qPCRs replicates. Output of Student's T-test analysis of the significance of changes is shown by * = P ≤ 0.05, ** = P ≤ 0.01, *** = P ≤ 0.001, **** = P ≤ 0.0001. Data shown are 2 technical (ChIP) replicates. p-values: upper row – Tfap2a = 0.00283, Cdh3 = 0.00299, Pax6 = 0.00250; lower row – Tfap2a = 0.000542, Cdh3 = 5.771E-05, Pax6 = 0.00235.

B) Gene expression of Pax6, Cdh3 and Tfap2a detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data shown is of one single experiment.

C) Gene expression (detected by RNA sequencing) of bivalent genes known to be bound by Oct4 which show significant (P <0.001) changes over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL, shown by reads per kilo base per million mapped reads (rpkm). The RNA-seq data shown is of one single experiment.

to look at these loci to assess the histone marking profiles upon immediate Oct4 depletion. It **would** be expected that those that show increase of expression over time would lose the repressive marking and the ones that go down would lose the active mark. Next generation high throughput sequencing of such ChIPs would be of future interest to be able to scan the wide number of bivalent genes.

This is a very interesting dynamic to interpret. These immediate changes suggest that Oct4 plays a role in the resolution of bivalent marks into either an on or an off state. If one considers that at the regulatory region of one gene there is binding of, among other molecules, Oct4 and two histone marks (one active and one repressive), and that, upon loss of Oct4, there is, as a consequence (either by structural changes, or the unsequestering of a molecule such as an enzyme, or any other mechanism), the loss of one of this mark, how is this loss specific? How does Oct4 loss in one specific locus lead to loss of the repressive mark and in another locus of lead to loss of the active mark? Maybe this can be explained by different loci having a different complex of molecules as a whole binding to them, with which Oct4 may interact differently, leading to different resolution upon its loss. It could also be that this resolution does not depend of Oct4 at all, but that the acute loss of Oct4 acts through some molecule which is the real regulator of bivalency at these loci. Once again, the idea that Oct4 is unsequestering some factor is a possibility.

Overall, this system proves itself as an extremely powerful tool to understand not only the dynamics of the action of Oct4 but also the regulation and instruction of cell identity and the immediate responses of all these flexible networks respond to interference.

These results, taken altogether, inform us of how Oct4 functions to maintain the naïve pluripotent identity, as it is necessary for active marks to be present at pluripotency genes, repressive marking at TE loci, and the maintenance of a correct bivalent pattern at lineage markers. The observation of such direct effects in the described mechanisms upon Oct4 loss suggests that one of Oct4 primary functions is to maintain the correct chromosomal organization which will then contribute to the correct gene expression pattern that instructs the cell to retain naïve pluripotency.

Brief discussion of the O4AID system in SL conditions

Throughout this work it has been described that the SL culture conditions did now work as efficiently as the 2iL conditions, based on immunostaining findings (Figure 4.1C), even though Western Blots showed clearly degradation of Oct4. That premise led to a more cautious interpretation of results obtained in SL, and structured this body of work to privilege the findings in 2iL conditions. However, very close to the submission of this MSc thesis I have obtained some preliminary data that increases the confidence in the SL condition.

When doing immunostainings for the AID tag that is present in the transgenic version of Oct4 of ZhbTO4AID cells, the percentage of cells still AID positive after 2h of IAA induction (Figure 4.10) is similar to the percentage of cells still Oct4 positive when this induction is made in 2i (Figure 4.1C). This percentage is very low. This finding suggests that the Oct4 immunostaining could be detecting another protein in SL conditions, such as another protein from the POU family. Nonetheless, one should investigate further why would such false positive only occur in SL and bearing in mind that the specificity of the antibodies used have been thoroughly checked before by members of the Silva lab in their diverse projects.

If it is confirmed that, in fact, the efficacy of the system in SL is trustworthy, then it would be interesting to interpret the effects of Oct4 loss in transcription in the SL context. This is an interesting point to be addressed in the future.

Very importantly, I shall underline the striking effect observed in H3K9me3 presence upon Oct4 loss, which under SL conditions is very much pronounced. If it is confirmed that the SL system is

efficient, it will be very important to look more deeply into this and possibly sequence the H3K9me3 ChIP performed in cells coming from SL.

Even if the system is as efficient, there are notable and interesting changes of behavior between the two conditions, which are expected and have been discussed before, one example being Figure 4.4B. It should be noted that the immunostaining results of Figure 4.4.C and Figure 4.8.G are similar for SL conditions (Figure S10).

Nevertheless, this preliminary finding needs to be replicated and the situation looked into more detail to assess what is giving a false positive in SL immunostainings and considering the difference in transcriptional kinetics observed over the course of 6 days of TE differentiation (Figure S2A), and the presence of cells that don't delete Oct4 in replicated TE assays when ZhbTO4AIDs would come from a SL background (Figure 4.2C). In the future, immunostaining studies should be done in parallel with staining for the AID tag instead of Oct4 staining.

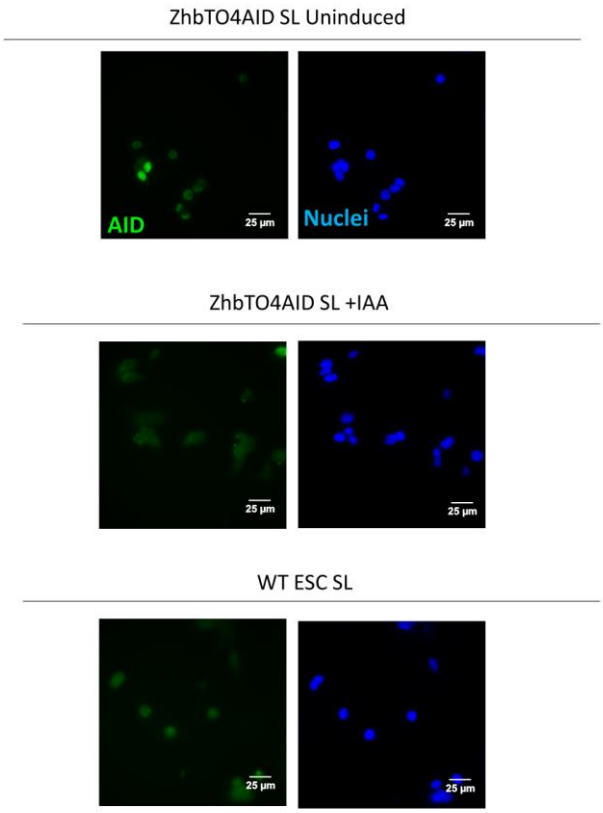

Figure 4.10 | AID immunostaining in SL
AID immunostaining of ZhbTO4AID cells in SL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA) and of WT ESCs. Nuclei (blue) are stained using DAPI. Pictures shown are representative of 4 different taken in each condition and are a projection of 10 slices at 63X.

CHAPTER 5

Conclusions & Perspective

"Although much remains obscure, and will long remain obscure, I can entertain no doubt, after the most deliberate study and dispassionate judgement of which I am capable… "

On the origin of the species, Charles Darwin

“The times they are a-changing” (Bob Dylan).

For years Oct4 has been known to be crucial to naïve pluripotent stem cell identity. However, the field has struggled to perform functional studies, in particular loss-of-function ones, that would allow the mechanistic dissection of this pluripotency factor. But the times have changed and this body of work is the first to describe a novel system of Oct4 depletion, established by Lawrence Bates. Using this system, the work presented above describes a direct molecular mechanism by which Oct4 works in the maintenance of naïve pluripotency for the first time.

In this study, I characterize this system of acute Oct4 depletion through an extensive number of techniques for the first time. This characterization was and is crucial to validate the use of this system to answer questions regarding the function of Oct4 in mouse ESCs and highlights the novelty and power of the system when compared to previous systems. The rapid nature of this Oct4 depletion has been convincingly demonstrated. I have also demonstrated that these cells don’t exhibit an Oct4 low state, unlike the previous depletion models. Due to this characterization, the optimal methods to maintain these cells and ensure the efficiency of the system were established. Being such a useful system, it will certainly open doors to answer many diverse scientific questions in the future. For that to occur, and for other researches being able to use this system to interrogate the function of Oct4, naïve pluripotency, and cell identity itself, the system needed to have this thorough characterization.

Regarding transcription, I found that Klf2 is probably the most direct target of Oct4 between all pluripotency factors. This “hierarchical” comparison between the interaction of Oct4 and other members of the pluripotency network was only possible due to the novel acute characteristic of this Oct4 depletion system. This observation accompanies the observations that the transcription of pluripotency genes can occur in the absence of Oct4 protein and that TE-related genes gradually increase in expression throughout the 24h following Oct4 depletion, suggesting that some members of the naïve network may be capable of substituting the function of Oct4 for a short period, with cell identity collapsing as early as 6h without Oct4.

Regarding the epigenetic landscape, it was shown that there are very immediate effects upon Oct4 loss on the active marking of pluripotency genes by H3K27ac, the repressive marking of TE genes by H3K9me3 and the repressive marking by H3K27me3 of genes marked bivalently. This shows a direct relationship of Oct4 with the maintenance of the correct histone modification patterns of the regulatory regions of genes which compose naïve identity. In the case of the activation of pluripotency genes, it was also described in this work that the regulatory regions of these genes undergo transcriptional changes prior to the coding regions of the genes

themselves, hinting at a primary function of Oct4 in maintaining the right regulatory behavior of the chromatin of these cells. In the case of H3K9me3, it was shown that loss of Oct4 also has consequences at loci where Oct4 does not bind, suggesting a major role of Oct4 in the maintenance of the global H3K9me3 pattern in naïve pluripotent stem cells. These results, being as immediate as 1.5h-2h upon Oct4 depletion, are very striking.

Thus, taken altogether, the results of this MSc thesis suggest that the molecular mechanism behind Oct4's maintenance of naïve pluripotency consists primarily of Oct4 sustaining the right chromosomal organization that supports the appropriate modification of histones, and the correct expression of non-coding regulatory regions, thus leading to the modulation of the gene expression patterns that define the naïve identity. It is the first time that evidence supports this hypothesis, that if confirmed, could change our understanding of Oct4. These observations do not only inform us about Oct4 function but also tell us a lot about cell identity, particularly that of naïve pluripotency stem cells. It is even possible to abstract from stem cell identity and learn lessons from these observations regarding the interplay of histone marking and transcriptional changes, and the interaction of transcription factor with the epigenome.

However, a lot more needs to be examined to lead us towards a greater comprehensive understanding of the real mechanism of Oct4. Various questions arise from the observations of this MSc thesis and hopefully will be answered in the future. How does Oct4 sustain the right structure of chromatin complexes? Is Oct4 sequestering a specific molecule to prevent its action during naïve pluripotency? Which are the exact factors that need Oct4 irreplaceably for their expression? This system really is a paradigm changer in the pluripotency field and not only will more be known about Oct4 in the future but also about this cell identity itself. It suggests that chromatin organization is the first layer of control of cell identity, and that the correct organization is maintained by proteins like Oct4 rather than being a result of transcriptional cues.

In the immediate follow-up of this work, as I have mentioned before, it will be interesting to perform next generation high throughput sequencing of the histone mark ChIPs in order to understand the effect of the loss of Oct4 more widely. It will also be very important to validate the key findings in a new cell line, which does not present some impairments to standard ESC function such as incapability of forming mesoderm upon differentiation or the exhibition of a morphology indicative of suboptimal naïve culture. To truly dissect whether the action of Oct4 is primarily upon chromatin organization prior to transcriptional effects, it will be important to perform nascent RNA capture and dissect the transcriptional kinetics of Oct4 loss, in order to

formally demonstrate the order of events. Chromosome conformation capture studies should also be performed in order to understand if Oct4 really has a role in the maintenance of correct DNA looping.

Many other questions could be answered using this system, such as evaluating the binding profiles of pluripotency factors to the genome in order to understand whether Oct4 contributes to their correct position in the genome. For this, ChIP for pluripotency factors could be performed. Other aspects of naïve pluripotency could be explored, such as the DNA methylation patterns, which are very distinct in this identity. For this, bisulfite sequencing could be performed, as well as the evaluation of levels/localization of DNA methylases upon Oct4 loss. It would be also interesting to validate these loss of function results with gain of function experiments; it should be possible to wash off IAA from the cells (Zhang *et al*., 2015) so that O4AID protein would be expressed again (since this system only degrades Oct4 protein, transcription would continue). If this system could be incorporated into *in* vivo studies it would be very interesting to assess the immediate consequences in vivo of Oct4 loss during epiblast formation and interrogate the effect on the levels of naïve pluripotent genes, for example. If a reliable *in vitro* human naïve ESC model is established, this O4AID system could be leveraged to examine how far these functions of Oct4 are conserved in human pluripotency, a state that is remarkably difficult to study *in vivo*. Furthermore, if this AID system could be combined with a reporter system, it would be possible to live image the disappearance of Oct4, and consequently track in real time the effects it would have at the single-cell level. It would be also interesting to transpose this system to look not only to naïve pluripotency but also to cell state transitions and the role of Oct4 in these. For example, one could try to assess the consequences of acute Oct4 depletion at certain stages of reprogramming of differentiated cells to induced pluripotency and examine the role of Oct4 in this process. By disrupting the reprogramming protocol with Oct4 depletion at various stages it could be detailed in which timings of reprogramming is Oct4 truly essential. The system could also be used to disrupt the differentiation into certain lineages by deleting Oct4 and seeing the consequences. This system could also be used to better understand the interactome of Oct4, by, for example, analyzing the quantities of different pull downs between the control and throughout a time course of Oct4 depletion.

Naïve pluripotent stem cells are powerful biological entities that can become any type of cell of the adult organism. The maintenance of this identity, in mechanistic detail, is still much of a black box. But with this work, we are one step towards a better understanding of how naïve pluripotent stem cells maintain their characteristics, in particularly through the action of Oct4.

And with this system, many steps can be taken and many more interesting observations will occur.

A wide variety of researchers could take advantage of this system and its study to refine their studies. Molecularly, this study provides interesting insights that, followed up in the future, could lead to a better knowledge on how transcription factors interact with histone marking and DNA looping. The better understanding of Oct4, due to its consequential information on naïve pluripotency, reprogramming to induced pluripotency and differentiation, will allow the refinement of conditions to maintain cells in culture as close as possible to their correct identity (such as iPSCs), that way providing more accurate results that could be transposed to areas such as disease modeling, drug screenings, or even stem cell therapy. It could also improve fundamental research on various topics from disease to development, given that they could fine tune better the cells they are using if those identities are understood better. Moreover, it provides insights that could be transposed to *in vivo* development.

It is impossible not to be amazed by the characteristics of stem cells and wonder how they work. We all were naïve pluripotent cells in the first place. And with our increasing understanding of these phenomena, we will be able to understand better where we came from.

CHAPTER 6

Supplementary Figures

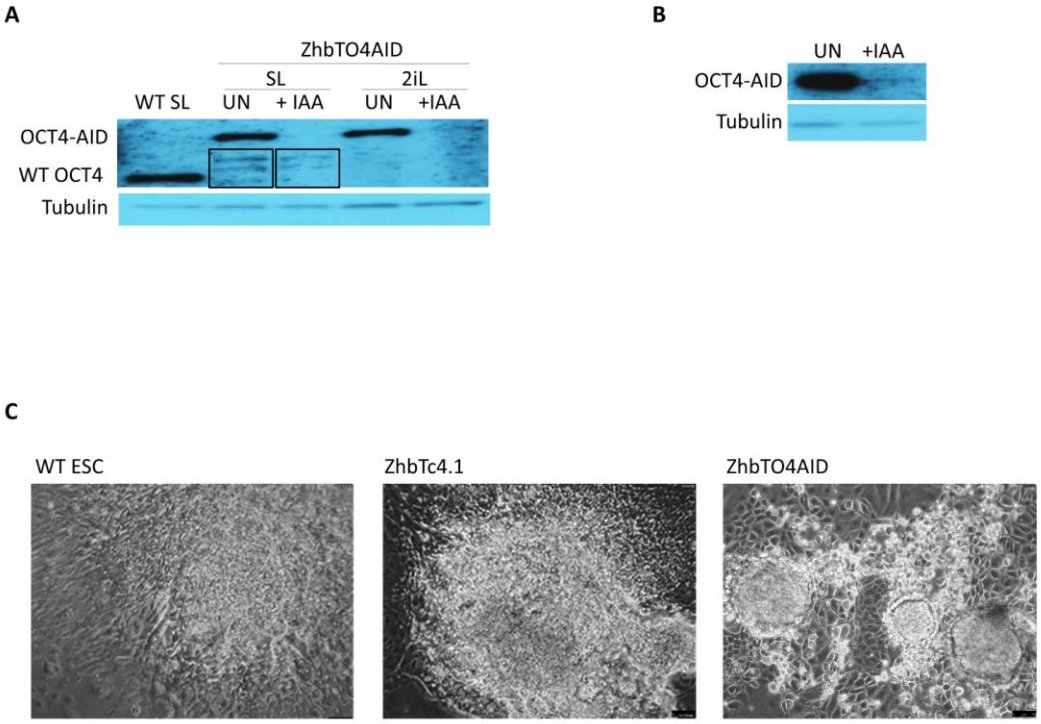

Figure S1 | O4AID depletion system

A) Oct4 Western Blot of ZhbTO4AID cells in SL and 2iL, in uninduced (control) conditions (UN) or 2h after Oct4 depletion (+IAA); and of WT ESCs. Tubulin loading control is presented.

B) Oct4 Western Blot of ZhbTO4AID cells in SL, in uninduced (control) conditions (UN) or 2h after Oct4 depletion (+IAA). Tubulin loading control is presented.

C) Phase images at 10X of representative of EB outgrowth of WT ESCs, ZhbTc4.1s, and of ZhbTO4AID cells originally maintained in 2iL culture conditions, after 10 days of EB differentiation assay. The WT ESC and ZhbTc4.1 are frames of the recordings showing representative beating cardiac cells formed in this assay. These can be found at https://figshare.com/s/17fa3167ecf4eb3d5ca9 (ZhbTc4.1s from 2iL_EB assay: 2 beating bodies' videos) and https://figshare.com/s/5c1a2d0a16228bfd9b39 (WT ESCs from 2iL_EB assay: 2 beating bodies' videos). This experiment was performed separately from Figure 4.1D.

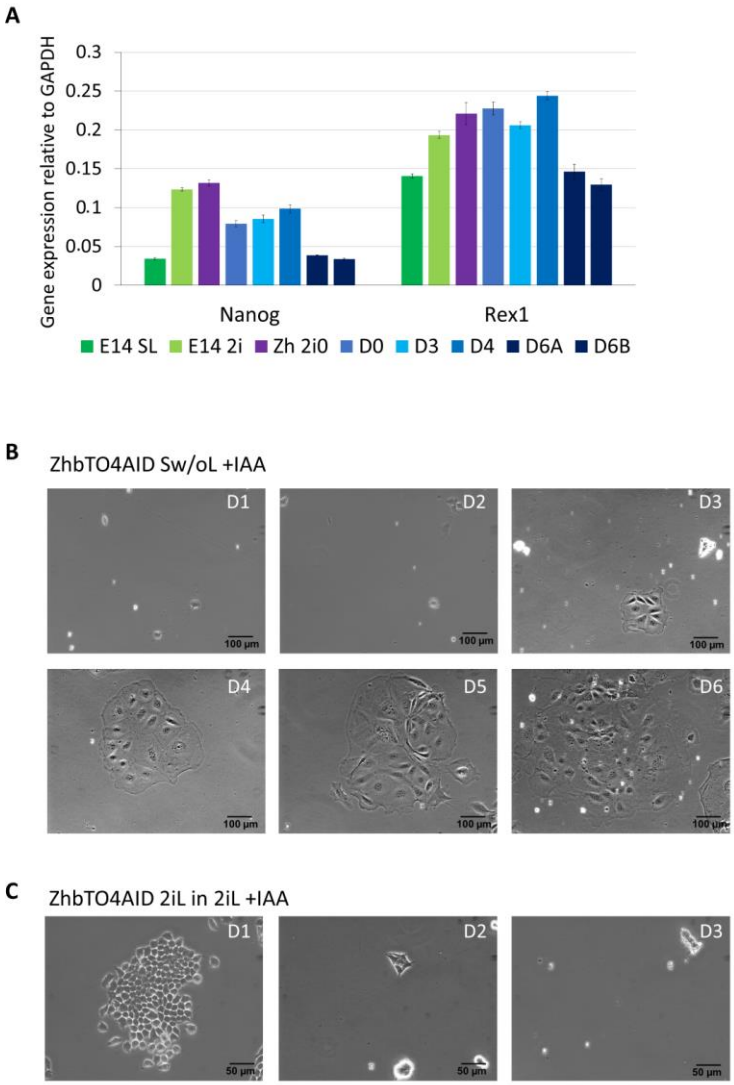

Figure S2 | TE differentiation upon Oct4 depletion via AID system

A) qPCR analysis of pluripotency markers expression during TE differentiation. Legend from left to right: WT ESCs in SL (E14 SL), WT ESCs in 2iL (E14 2i), Uninduced ZhbTO4AIDs in 2iL, Uninduced ZhbTO4AIDs in SL, D0-D6 indicate number of days of differentiation from ZhbTO4AIDs in SL. D6A and D6B are replicates. Data shown are the mean of 3 replicates and error bars are ±s.d.

B) Phase images representative of TE differentiation assays of ZhbTO4AID cells originally maintained in SL culture conditions but where the TE assay is performed without the presence of LIF (ZhbTO4AID Sw/oL +IAA). D0-D6 indicate number of days of differentiation. Representative features discussed in the result section are circled. The circle at D3 shows a cell exhibiting TE-like morphology. Data shown are from 1 of 2 representative experiments.

C) Phase images representative of TE differentiation assays of ZhbTO4AID cells originally maintained in 2iL culture conditions and kept in 2iL culture conditions throughout Oct4 depletion by adding IAA (ZhbTO4AID 2iL in 2iL+IAA). D0-D6 indicate number of days of differentiation.

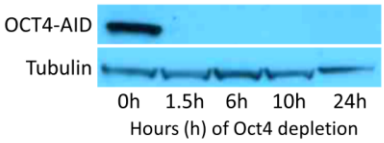

Figure S3 | RNA-seq induction validation

Oct4 Western Blot of ZhbTO4AID cells in 2iL over a time course of Oct4 depletion. Tubulin loading control is presented. This experiment was done in parallel to the RNA-seq time course and to the ChIP experiments shown in Figures 7 and 8, to validate the depletion of Oct4.

1.5h	6h	10h	24h
Fzd5	Fzd5	Fzd5	Pam
Amotl2	AC139347.1	AC139347.1	Anxa2
Chac1	Ahrr	Accsl	Arrb1
Cyp1a1	Aldh1a3	Ahrr	Cidea
Ffar3	C1ql2	C1ql2	Cyp1a1
Gadd45b	Cidea	Cdx1	Cyp1b1
Klf2	Cobl	Cidea	Esrrb
Mat2a	Cyp1a1	Cobl	Fgf4
Ppp1r10	Cyp1b1	Cyp1a1	Frem2
Prr15	Gadd45g	Cyp1b1	Gm14136
Sgk1	Gm13128	Fgfrl1	Gm1564
Slc38a2	Gm14136	Gm14136	H19
Spry2	Gss	Gprc5c	Jam2
Zfp296	H2-M5	H2-M5	Klf4
	Hpse	Hpse	L1td1
	Il10	Il10	Lin28a
	Jam2	Jam2	Mmp15
	Klf2	Kcnd3	Nfatc2ip
	Klf4	Klf4	Nr0b1
	L1td1	L1td1	Pecam1
	Lrrc8c	Lrpap1	Peg10
	Myh7b	Myh7b	Rec8
	Myo10	Nfatc2ip	Rhox6
	Nfatc2ip	Nkx1-2	Rragd
	Nkx1-2	Nmnat2	Sprr2a1
	Nr0b1	Notum	Sprr2a2
	Pdgfc	Nr0b1	Sprr2a3
	Peg10	Parm1	Stra8
	Phc1	Peg10	Tmprss2
	Pigl	Phc1	Trpm1
	Plekho2	Pigl	Txnip
	Plk3	Plk3	Upp1
	Pou2f3	Pou2f3	Wls
	Ptch2	Ptch2	Zfp42
	Slc24a5	Rragd	
	Slc2a1	Slc7a11	
	Slc7a11	Sprr2a3	
	Spry2	Spry2	
	Tbx3	Tbx3	
	Tdgf1	Tdgf1	
	Tet2	Tet2	
	Tfcp2l1	Tinagl1	
	Tiparp	Tiparp	
	Usp29	Tmprss2	
	Zfp296	Usp29	
	Zfp42	Vcan	
	Zfp703	Wnt6	
		Zfp296	
		Zfp36l3	
		Zfp703	

Table S3 | RNA-seq P <0.001 hits

Table of gene hits from whole-transcriptome RNA-seq analysis which shown significant (P <0.001) changes over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. Significance is determined relative to 0h for each timepoint using an intensity dependent t-test. The RNA-seq data is of one single experiment. This list was obtained by Lawrence Bates.

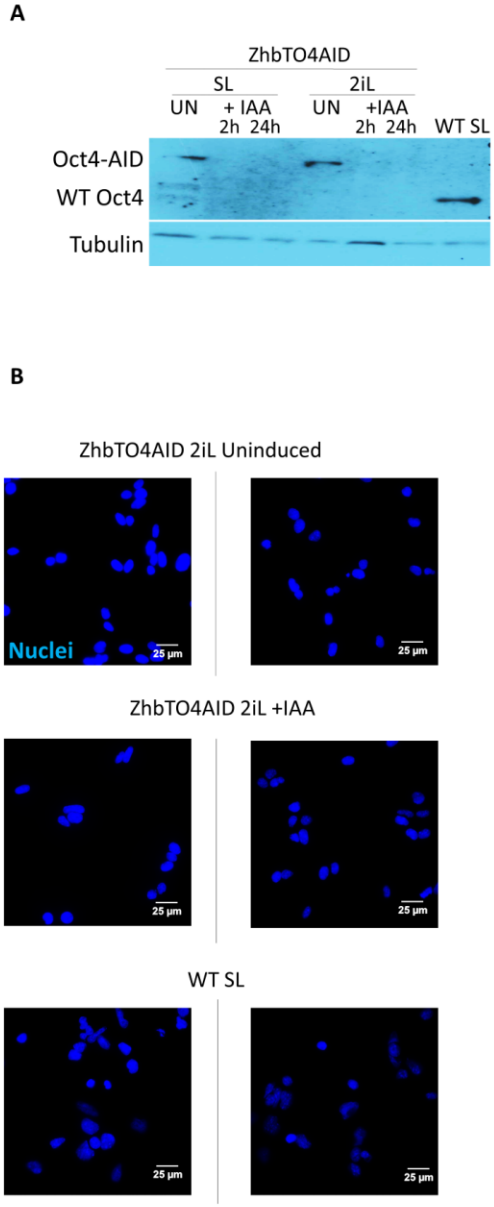

Figure S4 | Induction validation for pluripotency network Western Blots

A) Oct4 Western Blot of ZhbTO4AID cells in either SL or 2iL in uninduced (control) conditions (UN) or 2h or 24h after Oct4 depletion (+IAA), and of WT ESCs. Tubulin loading control is presented. These are the same samples used in Figure 4.4B, to validate the depletion of Oct4.

B) Nuclei (blue) immunostaining using DAPI, respective to Figure 4.4C, in the same order as shown in Figure 4.4C. Immunostainings of ZhbTO4AID cells in 2iL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA); and of WT ESCs. Pictures shown are a projection of 10 slices at 63X.

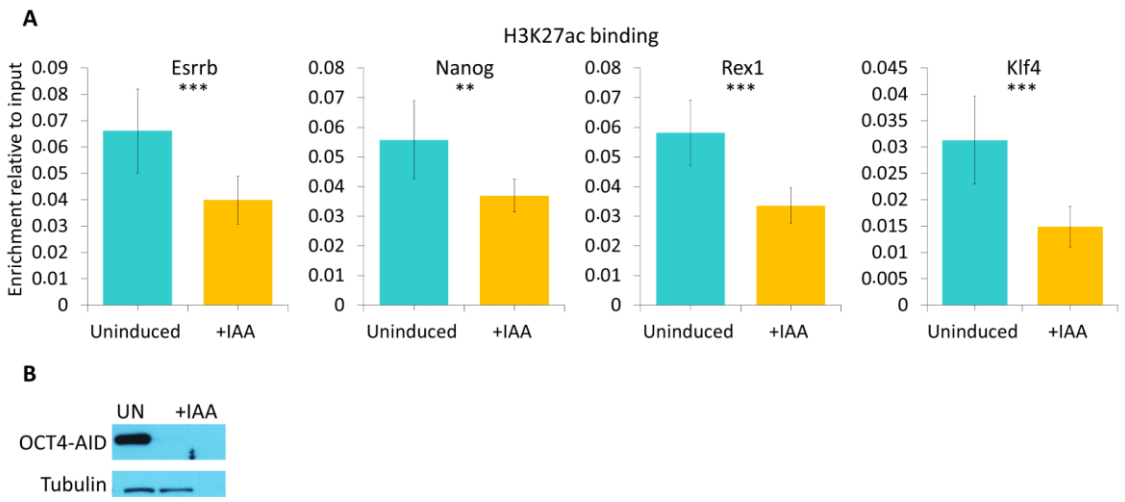

Figure S7 | Binding of H3K27ac in SL and induction validation

A) ChIP-qPCR analysis of H3K27ac binding at the regulatory regions of pluripotency genes (Esrrb (enhancer), Nanog (enhancer), Rex1 (promoter) and Klf4 (enhancer)) also bound by Oct4 in ZhbTO4AID cells maintained in SL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 3 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d. Output of Student's T-test analysis of the significance of changes is shown by * = P ≤ 0.05, ** = P ≤ 0.01, *** = P ≤ 0.001. p-values: Esrrb = 0.000891, Nanog = 0.00229, Rex1 = 0.000170, Klf4 = 0.000210. Data shown are from 1 of 2 replicate experiments.

B) Oct4 Western Blot of ZhbTO4AID cells in SL, in uninduced (control) conditions (UN) or 2h after Oct4 depletion (+IAA). Tubulin loading control is presented. This experiment was done in parallel to the SL ChIP experiments shown in Figures S7, S8 and 3.9, to validate the depletion of Oct4.

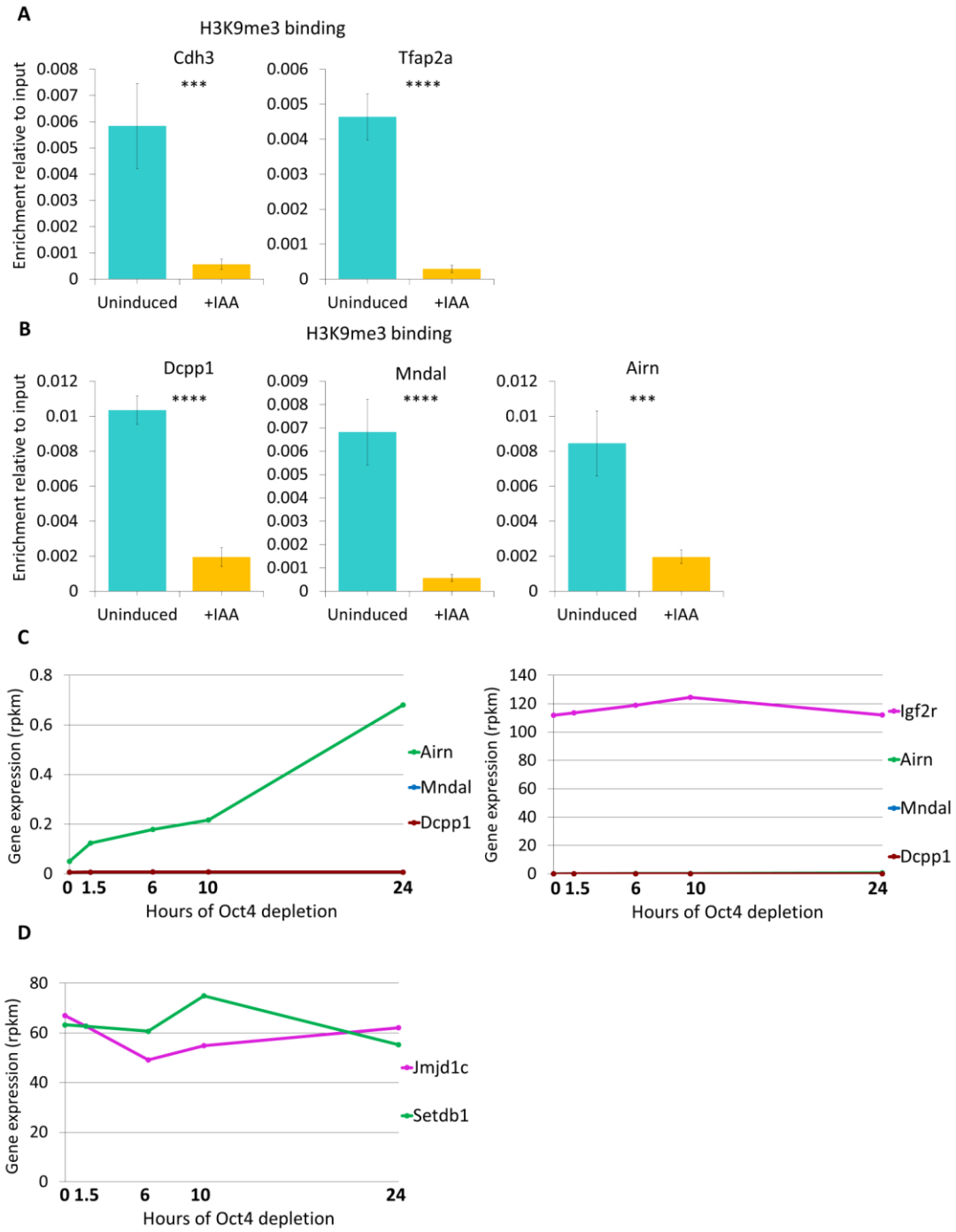

Figure S8 | H3K9me3 binding in SL and RNA-seq of respective loci

A) ChIP-qPCR analysis of H3K9me3 binding at the regulatory regions of TE genes (Cdh3 (enhancer) and Tfap2a (enhancer)) also bound by Oct4 in ZhbTO4AID cells maintained in SL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 2 replicate IPs for each condition (each with 3 qPCR replicates), error bars are ±s.d. Tfap2a result is the mean of two replicate qPCRs. Output of Student's T-test analysis of the significance of changes is shown by *** = P ≤ 0.001, **** = P ≤ 0.0001. p-values: Cdh3 = 0.000450, Tfap2a = 7.124E-11.

B) ChIP-qPCR analysis of H3K9me3 binding at the regulatory regions of genes not bound by Oct4 (Dcpp1 (enhancer), Mndal (enhancer) and Airn (enhancer)) in ZhbTO4AID cells maintained in SL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA). Data is shown as enrichment relative to input with the IgG background subtracted. Data shown are the mean of 2 replicate IPs for each condition (each with

3 qPCR replicates), error bars are ±s.d. Output of Student's T-test analysis of the significance of changes is shown by *** = P ≤ 0.001, **** = P ≤ 0.0001. p-values: Mndal = 0.000101, Dcpp1 = 7.696E-09, Airn = 0.000268.

C) Gene expression of Mndal, Dcpp1, Airn and Igf2r detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data is of one single experiment.

D) Gene expression of Jmjd1c and Setdb1 detected by RNA sequencing, shown by reads per kilo base per million mapped reads (rpkm), over a time course of Oct4 depletion in ZhbTO4AID cells maintained in 2iL. The RNA-seq data is of one single experiment.

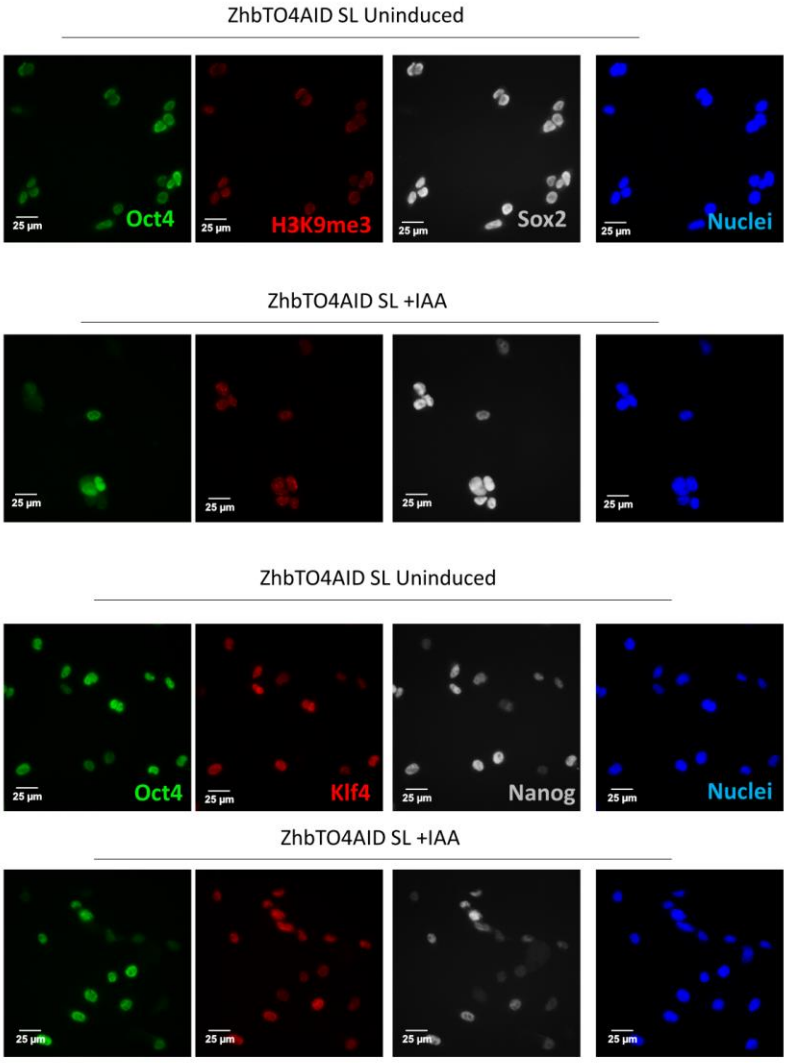

Figure S10 | SL Immunostainings

Oct4, Pluripotency factors (Sox2, Klf4 and Nanog) and H3K9me3 immunostainings of ZhbTO4AID cells in SL, in uninduced (control) conditions or 2h after Oct4 depletion (+IAA); and of WT ESCs. Pictures shown are representative of 7 different taken in each condition and are a projection of 10 slices at 63X. Data shown are from 1 of 2 representative experiments in the case of Sox2 immunostaining. Nuclei (blue) were stained with DAPI.

CHAPTER 7

References

Ambrosetti, D.C. *et al*. (2000) Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J Biol Chem 275, 23387–23397

Ang, Y.S. *et al*. (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell 145, 183-197

Askoy, I. *et al*. (2013) Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J. 32, 938-953

Avilion, A.A. *et al*. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17, 126–140

Azurara , V. *et al*. (2006) Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8, 532-538

Ben-Shushan, E. *et al*. (1993) Extinction of Oct-3/4 gene expression in embryonal carcinoma x fibroblast somatic cell hybrids is accompanied by changes in the methylation status, chromatin structure, and transcriptional activity of the Oct-3/4 upstream region. Mol. Cell. Biol. 13, 891-901

Ben-Shushan, E. *et al*. (1998) Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. Mol. Cell. Biol. 18, 1866–1878

Bernstein, B.E. *et al*. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315-326

Bilodeau, S. *et al*. (2009) SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. 23, 2484-2489

Bourillot, P.Y. *et al*. (2009) Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. Stem cells 27, 1760-1771

Boroviak, T. *et al*. (2015) Lineage-Specific Profiling Delineates the Emergence and Progression of Naïve Pluripotency in Mammalian Embryogenesis. Dev Cell 35, 366-382

Botquin, V. *et al*. (1998) Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. Development 130, 1673–1680

Boyer, L.A. *et al*. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-956

Boyer, L.A. *et al*. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349-353

Brehm, A. *et al*. (1997) The carboxy-terminal transactivation domain of Oct4 acquires cell specificity through the POU domain. Mol. Cell. Biol. 17, 154–162

Buecker, C. *et al*. (2014) Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. Cell Stem Cell 14, 838–853

Calabrese, J. M. *et al*. (2015) A Survey of Imprinted Gene Expression in Mouse Trophoblast Stem Cells. G3 5, 751-759

Campbell, S.J. *et al*. (2005) The Murine Cyp1a1 Gene Is Expressed in a Restricted Spatial and Temporal Pattern during Embryonic Development. J Biol Chem 280, 5828-5835

Carey, B.W. *et al*. (2011) Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell Stem Cell 9, 588–598

Chamberlain, S.J. *et al*. (2008) Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem cells 26, 1496-1505

Chambers, I. *et al*. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643–655

Chambers, I. *et al*. (2007) Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-1234

Chen, T. & Dent, S.Y. (2014) Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat Rev Genet. 15, 93–106

Chen, X. *et al*. (2008) Integration of external signalling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106–1117

Chowdhury, F. *et al*. (2010) Soft Substrates Promote Homogeneous Self-Renewal of Embryonic Stem Cells via Downregulating Cell-Matrix Tractions. PLoS ONE 5(12): e15655

Costa, Y. *et al*. (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 495, 370-374

Creyghton, M.P. *et al*. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107, 21931-21936

Dekker, J. *et al*. (2002) Capturing chromosome conformation. Science. 295, 1306–11

Deveale, B. *et al*. (2013) Oct4 is required e7.5 for proliferation in the primitive streak. PLoS Genet. 9, e1003957

Ding, J. *et al*. (2012) Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res. 22, 155–167

Dodge, J.E. *et al*. (2004) Histone H3-K9 methyltransferase ESET is essential for early development. Molecular and cellular biology 24, 2478-2486

Downs, K.M. (2008) Systematic localization of Oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. Dev. Dyn. 237, 464–475

Efe, J. *et al*. (2011) Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. Nat. Cell Biol. 12, 215-222

Efroni, S. *et al*. (2008) Global transcription in pluripotent embryonic stem cells. Cell stem cell 2, 437-447

Esch, D. *et al*. (2013) A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat. Cell Biol. 15, 295–301

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

Festuccia, N. *et al*. (2012) Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell 11, 477–490

Frum, T. *et al*. (2013) Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. Dev. Cell 25, 610–622

Guo, Y. *et al*. (2002) The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. PNAS 99, 3663-3667

Guo, G. *et al*. (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev Cell 18, 675–685

Hall, J. *et al*. (2009) Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell 5, 597–609

Hammachi, F. *et al*. (2012) Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. Cell Rep. 1, 99–109

Hay, D.C *et al*. (2004) Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells 22, 225-235

Hayashi, K. *et al*., (2008) Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem. Cell 3, 391-401

Holland, P. *et al*. (2007) Classification and nomenclature of all human homeobox genes. BMC Biol. 5, 47

Huang, J.C. *et al*. (2007) Comparison of histone modifications in in vivo and in vitro fertilization mouse embryos. Biochemical and Biophysical Research Communications 354, 77-83

Imagawa, M. *et al*. (1991) Stringent integrity requirements for both trans-activation and DNA-binding in a trans-activator, Oct-3. Nucl. Acids Res. 19, 4503–4508

Jaenisch, R. & Young, R. (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567–582

Jiang, J. *et al*. (2008) A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol. 10, 353-360

Kagey, M.H. *et al*. (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430-435

Karwacki-Neisius, V. *et al*. (2013) Reduced Oct4 expression directs a robust pluripotent state with distinct signalling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell 12, 531–545

Kehler, J. *et al*. (2004) Oct4 is required for primordial germ cell survival. EMBO Rep. 5, 1078-1083

Kim, J. *et al*. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132, 1049-1061

Kim, J. *et al*. (2011) Direct reprogramming of mouse fibroblasts to neural progenitors. Proc. Natl. Acad. Sci. U.S.A. 108, 7838–7843

Kim, J.B. *et al*. (2009a) Oct4-induced pluripotency in adult neural stem cells. Cell 136, 411–419

Kim, J.B. *et al*. (2009b) Direct reprogramming of human neural stem cells by Oct4. Nature 461, 649–653

King & Klose (2017) The pioneer factor Oct4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife 6:e22631

Koche, R. P. *et al*. (2011), Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell 8, 96–105

Kopp, J.L. *et al*. (2008) Small increases in the level of Sox2 trigger the differentiation of mouse embryonic

stem cells. Stem Cells 26, 903–911

Kumar, S.M. *et al*. (2012) Acquired cancer stem cell phenotypes through Oct4-mediated dedifferentiation. Oncogene 47, 4898-4911

Kunath, T. *et al*. (2007) FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development 134, 2895-2902

Kuo, C.T. *et al*. (1997) LKLF: A transcriptional regulator of single positive T cell quiescence and survival. Science 277, 1986-1990
Kurimoto, K. *et al.,* (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res 34, e42

Lenardo, M.J. *et al*. (1989) Repression of the IgH 1030 enhancer in teratocarcinoma cells associated with a novel octamer factor. Science 243, 544–546

Lee, K-F. *et al*. (2006) Demilune Cell and Parotid Protein from Murine Oviductal Epithelium Stimulates Preimplantation Embryo Development. Endocrinology 147, 79-87

Leeb, M. *et al*. (2010) Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev 24, 265-276

Leitch, H.G. *et al*. (2013) Naïve pluripotency is associated with global DNA hypomethylation. Nature structural & molecular biology 20, 311-316

Lessard, J.A. & Crabtree, G.R. (2010) Chromatin regulatory mechanisms in pluripotency.
Annual review of cell and developmental biology 26, 503-532

Li, P. *et al*. (2008) Germline competent embryonic stem cells derived from rat blastocysts. Cell 135, 1299-1310

Li, R. *et al*. (2010) A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell 7, 51–63

Li, Y. *et al*. (2011) Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. Cell Res. 21, 196–204

Liang, J. *et al*. (2008) Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell. Biol. 10, 731–739

Lim, H.Y. *et al*. (2009) Implication of human Oct4 transactivation domains for self-regulatory transcription. Biochem Biophys Res Commun. 385, 148-53

Livigni, A. *et al*. (2013) A conserved Oct4/POUV-dependent network links adhesion and migration to progenitor maintenance. Curr. Biol. 23, 2233–2244 57

Locke, J. (1689), An Essay Concerning Human Understanding

Loh, K.M. & Lim, B. (2011) A precarious balance: pluripotency factors as lineage specifiers. Cell Stem Cell 8, 363–369

Loh, Y.H. *et al*. (2007) Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes Dev 21, 2545-2557

Loh, Y-H. *et al*. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. 38, 431–440

Lohmann, F. *et al*. (2010) KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. Stem cells 28, 201-212

Lu, Y. *et al*. (2012) Avian-induced pluripotent stem cells derived using human reprogramming factors. Stem Cells Dev. 21, 394–403

Luo, J. *et al*. (1997) Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. Nature 388, 778-782

Marks, H. *et al*. (2012) The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell 149, 590-604

Martello, G. & Smith, A. (2014) The nature of embryonic stem cells. Annu. Rev. Cell Dev. 20, 647-675

Martello, G. *et al*. (2013) Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. 32, 2561–2574

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

Masui, S. *et al*. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat. Cell Biol. 9, 625–635

Masui, S. *et al*. (2008) Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. BMC developmental biology 8, 45

Matoba, R. *et al*. (2006) Dissecting oct3/4-regulated gene networks in embryonic stem cells by expression profiling. PLoS ONE. 1, e26

Matsuda, T. *et al*. (1999) STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. EMBO J 18, 4261-4269

Matsui, Y. *et al*. (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 70, 841–847

Mermod, N. *et al*. (1989) The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell 58, 741–753

Meshorer, E. *et al*. (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell 10, 105–116

Mikkelsen, T.S. *et al*. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-560

Mikkelsen, T. S. *et al*. (2008) Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49–55

Minucci, S. *et al*. (1996) Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy *in vivo*. EMBO J. 15, 888–899

Morrison, G.M. and Brickman, J.M. (2006) Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. Development 133, 2011–2022

Nakagawa, M. *et al*. (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. 26, 101–106

Niakan, K. K. *et al*. (2006) Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. Mol Gen and Metabolism 88, 281-271

Nichols, J. & Smith, A. (2009) Naïve and primed pluripotent states. Cell Stem Cell 4, 487–492

Nichols, J. *et al*. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379–391

Nichols, J. & Smith, A. (2012) Pluripotency in the embryo and in culture. Cold Spring Harbor perspectives in biology 4, a008128

Nishimoto, M. *et al*. (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol. Cell. Biol. 19, 5453–5465

Nishimoto, M. *et al*. (2005) Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Mol. Cell Biol. 25, 5084-5094

Nishimura, K. *et al*. (2009) An auxin-based degron system for the rapid deletion of proteins in nonplant cells. Nature Methods 12, 917-923

Nishioka, N. *et al*. (2009) The Hippo signalling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Developmental cell 16, 398-410

Niwa, H. *et al*. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev 12, 2048–2060

Niwa, H. *et al*. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. 24, 372–376

Niwa, H. *et al*. (2002) Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. Mol. Cell. Biol. 22, 1526–1536

Niwa, H. *et al*. (2005) Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. Cell 123, 917-929

Niwa, H. (2007) Open conformation chromatin and pluripotency. Genes Dev. 21:2671-6

Niwa, H. *et al*. (2008) Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. Evol. Dev. 10, 671–682

Niwa, H. *et al*. (2009) A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature 460, 118–122

Ogle, B.M. *et al*. (2004) The Biology of Cell Fusion: Cells of different types and from different species can fuse, potentially transferring disease, repairing tissues and taking part in development. American Scientist 92, 420-427

Okamoto, H. *et al*. (1990) A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 60, 461–472

Okazawa, H. *et al*. (1991) The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. EMBO J. 10, 2997–3005

Osorno, R. *et al.* (2012) The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. Development 139, 2288–2298

Ovitt, C. & Schöler, H.R. (1998) The molecular biology of Oct-4 in the early mouse embryo. Mol. Human Rep, 4, 1021–1031

Palmieri, S.L. *et al*. (1994) Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Dev. Biol. 166, 259–267

Pan, G. *et al*. (2002) A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. FASEB J. 20, 1730-1732

Pardo, M. *et al*. (2010) An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell 6, 382–395

Pasini, D. *et al*. (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Molecular and cellular biology 27, 3769-3779

Pasini, D. *et al*. (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306–310

Pereira *et al.* (2008) Heterokaryon-Based Reprogramming of Human B Lymphocytes for Pluripotency Requires Oct4 but Not Sox2. PLoS Genet 4(9): e1000170

Pesce, M. & Schöler, H.R. (2001) Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells 19, 271–278

Pesce, M. *et al*. (1998) Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. Mech. Dev. 71, 89–98

Petryniak, B. *et al*. (1990) Characterization of chicken octamer-binding proteins demonstrates that POU domain-containing homeobox transcription factors have been highly conserved during vertebrate evolution. PNAS 87, 1099-1103

Polo *et al*., J.M. (2012) A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617–163
2

Radzisheuskaya, A. & Silva, J. (2014) Do all roads lead to Oct4? The emerging concepts of induced pluripotency. Trends in Cell Biology 24, 275-284

Radzisheuskaya, A. *et al*. (2013) A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* 15, 579–590

Ralston, A. *et al*. (2010) Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development 137, 395-403

Resnick, J.L. *et al*. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature 359, 550551

Reynolds, N. *et al*. (2012) NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell stem cell 10, 583-594

Rizzino, A. & Wuebben, E. (2016). Sox2/Oct4: A delicately balanced partnership in pluripotent stem cells and embryogenesis. BBA 1859, 780-
791

Rodda, D.J. *et al*. (2005) Transcriptional regulation of nanog by Oct4 and SOX2. J. of Biol. Chem. 280,

24731-24737

Rosfjord, E. *et al*. (1995) Phosphorylation and DNA binding of the octamer binding transcription factor Oct-3. Biochem. Biophys. Res. Comm. 203, 1795–1802

Rosner, M.H. *et al*. (1990) A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature 345, 686-692

Rossello, R.A. *et al*. (2013) Mammalian genes induce partially reprogrammed pluripotent stem cells in non-mammalian vertebrate and invertebrate species. Elife 2, e00036

Ruvkun, G. & Finney, M. (1991) Regulation of transcription and cell identity by POU domain proteins. Cell 64, 475–
478

Ryan, A. K. & Rosenfeld, M. G. (1997) POU domain family values: flexibility, partnerships, and developmental codes. Genes Dev. 11, 1207-
1225

Samavarchi-Tehrani, P. *et al*. (2010) Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 7, 64–77

Santoro *et al.* (2013) Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. Development 140, 1184-95

Schindelin, J. (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 28, 676-82

Schöler H.R. *et al*. (1989) A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. EMBO J. 8, 2543–2550

Schöler H.R. *et al*. (1990) New type of POU domain in germ line-specific protein Oct-4, Nature 344, 435–
43
9

Schöler H.R. *et al*. (1990) Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J. 9, 2185–
2195

Schöler H.R. *et al*. (1991) A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. *Cell* 66, 291–
304

Schöler H.R. (1991) Octamania: the POU factors in murine development. Trends Genet. 7, 323–329
Shakya, A. *et al*. (2015). Oct4 Mediates Stepwise Nucleosome Demethylation and Deletion. Mol. Cell. Biol. 35, 1014-
1025

Shimazaki, T. *et al*. (1993) Hybrid cell extinction and re-expression of Oct-3 function correlates with differentiation potential. EMBO J. 12, 4489–
4498

Shimozaki, K. *et al*. (2003) Involvement of Oct3/4 in the enhancement of neuronal differentiation of ES cells in neurogenesis-inducing cultures. Development 130, 2505–2512

Shu, J. *et al*. (2013) Lineage Specifiers: New Players in the Induction of Pluripotency. Genomics, Proteomics & Bioinformatics 11, 259–263

Smith, A.G. *et al*. (1998) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336, 688-690.

Smith, Z.D. *et al*. (2010) Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nature Biotech. 28, 521–
526

Soufi, A. *et al*. (2012) Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994–1004

Sridharan, R. *et al*. (2009) Role of the murine reprogramming factors in the induction of pluripotency. Cell 136, 364–
377

Sutherland, A.E. *et al*. (1990) Inner cell allocation in the mouse morula: the role of oriented division during fourth cleavage. Dev Biol 137, 13-25

Szabo, E. *et al*. (2010) Direct conversion of human fibroblasts to multilineage blood progenitors. Nature 468, 521–
526

Tai, C-I *et al*. (2014) Stat3 signaling regulates embryonic stem cell fate in a dose-dependent manner. Biology Open 3, 958-965

Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676

Takeda, J. *et al*. (1992) Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. Nucleic Acids Res 20, 4613-4620

Tapia, N. *et al*. (2012) Reprogramming to pluripotency is an ancient trait of vertebrate Oct4 and Pou2 proteins. Nat. Commun. 3, 1279

Tesar, P.J. *et al*. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196-199

Thier, M. *et al*. (2012) Direct conversion of fibroblasts into stably expandable neural stem cells. Cell Stem Cell 10, 473–479

Thomson, M. *et al*. (2011) Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145, 875–889

Thomson, J.A. *et al*. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282, 1145-1147

Thomson, J.A. *et al*. (1995) Isolation of a primate embryonic stem cell line. Proc Natl Acad Sci U S A 92, 7844-7848

Toyooka, Y. *et al*. (2008) Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909-918

Tropepe, V. *et al*. (2010) Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. Neuron 30, 65-78

Tsai, S.Y. *et al*. (2011) Single transcription factor reprogramming of hair follicle dermal papilla cells to induced pluripotent stem cells. Stem Cells 29, 964–971

van den Berg, D.L. *et al*. (2010) An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 6, 369–381

Verrijzer, C.P. *et al*. (1991) POU proteins bend DNA via the POU-specific domain. EMBO J. 10, 3007-3014

Vigano, M.A. & Staudt, L.M. (1996) Transcriptional activation by Oct-3: evidence for a specific role of the POU-specific domain in mediating functional interaction with Oct-1. Nucleic Acids Res 24, 2112–2118

Vinh Do, D. *et al*. (2015) A genetic and developmental pathway from STAT3 to the Oct4–NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes Dev 27, 1378-1390

Wang, Z-Q. *et al*. (1994) Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. Nature 372, 464-467

Wang, G. *et al*. (2013) Critical regulation of miR-200/ZEB2 pathway in Oct4/Sox2-induced mesenchymal-to-epithelial transition and induced pluripotent stem cell generation. Proc. Natl. Acad. Sci. U.S.A. 110, 2858–2863

Wang, J. *et al*. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature 444, 364–368

Wang, Z. *et al*. (2012) Distinct lineage specification roles for NANOG, Oct4, and SOX2 in human embryonic stem cells. Cell Stem Cell 10, 440-454

Watanabe, S. *et al*. (2006) Activation of Akt signalling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. Oncogene 25, 2697-2707

Webster, D.M. *et al*. (2009) O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. BMC Dev. Biol. 9, 28

Wei, F. *et al*. (2007) Sumoylation of Oct4 Enhances Its Stability, DNA Binding, and Transactivation. J. of Biol. Chem. 282, 21551-21560

Whyte, W.A. *et al*. (2013) Master transcription factors and mediator establish superenhancers at key cell identity genes. Cell 153, 307-319

Williams, R.L. *et al*. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336, 684-687

Wray, J. *et al*. (2011) Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol 13, 838-845

Wray, J. *et al*. (2010) The ground state of pluripotency. Biochemical Society transactions *38*, 1027-1032

Wu, T. *et al*. (2011) Reprogramming of trophoblast stem cells into pluripotent stem cells by Oct4. Stem Cells 29, 755–763

Wu, J. *et al*. (2016) The landscape of accessible chromatin in mammalian preimplantation embryos. Nature 534, 652–657

Yagi, R. *et al*. (2007) Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development 134, 3827-3836

Ye, S. *et al*. (2013) Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. EMBO J

Yeap, L.S. *et al*. (2009) ERG-associated protein with SETdomain (ESET)-Oct4 interaction regulates pluripotency and represses thetrophectoderm lineage. Epigenet Chromatin 2:12

Yeom, Y.I. *et al*. (1991) Structure, expression and chromosomal location of the Oct-4 gene. Mech. Dev. 35, 171–179

Yeom, Y.I. *et al*. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122, 881–894

Ying, Q-L. *et al.* (2003a) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115, 281-292

Ying, Q-L. *et al*. (2003b) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol 21, 183-186

Ying, Q-L. *et al.* (2008) The ground state of embryonic stem cell self-renewal. Nature 453, 519-523
Young, R.A. (2011) Control of the embryonic stem cell state. Cell 144, 940-954

Yuan, H. *et al*. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev 9, 2635–2645

Yuan, P. *et al.* (2009) Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev 23, 2507–2520

Yuan, X. *et al*. (2011) Brief report: combined chemical treatment enables Oct4-induced reprogramming from mouse embryonic fibroblasts. Stem Cells 29, 549–553

Zaret, K. & Carroll, J. (2011) Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25, 2227-2241

Zhang, H. *et al*. (2013) Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. Cell Stem Cell 13(1):30-5

Zhang, L. et al. (2015) The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development 142, 4374–4384

Zhu, J. *et al.* (2013) Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell 152, 642–654