THE UNIVERSITY *of* EDINBURGH

The Evolution and Functional Plasticity of Vertebrate Class V POU Proteins in Pluripotency

WORANOP SUKPARANGSI

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2015

To My Dear Sukparangsi and Maitreejit Families

"*Natural Selection almost inevitably causes much Extinction of the less improved forms of life and induces what I have called Divergence of Character.*"

Charles Darwin

ACKNOWLEDGEMENTS

I would like to express great attitude towards several people who make my dream come true. It is not all about getting the new title in front of my name or certificate but it is about a chance to explore science. Many people have given me a trust and allow me to follow my own path. I have received valuable professional experiences during five years of my PhD study. The first person I would like to thank is my supervisor Professor Joshua M Brickman who has given me such a great opportunity to start my research in stem cell field. He has already given me a lot of supports and trust in my science, especially my new area of study: evolutionary developmental biology (EVO-DEVO). He allowed me to explore this EVO-DEVO field with freedom and peruse my own lines of investigation. Importantly, thank you so much for the best mojito ever since my first year of Phd..Josh!

During this journey to the final story of this thesis, I am indebted to several people in DanStem (Copenhagen) and SCRM (Edinburgh) who provided me an invaluable source of both technical and conceptual advice and help to establish this exciting story with essence of science. Here are friends and collaborators who kindly helped me to this final point and I would like to give a lot of thanks to them: Alessandra Livigni, Hanna Peradziyri, Keisuke Kaji, Kumiko A Iwabuchi, Fabian Roske, Kathryn Anderson, Jurriaan J Hölzenspies, William Hamilton, Dora Papp, Michaela Rothova, Sophie Morgani, Javier Martin Gonzalez, Yung Hae Kim, Lydie Carole Flasse, Evan Manuel Figueiredo-Larsen, Laurence Anne Emmanuelle Lemaire, Gelo Victoriano Dela Cruz, Stephen Frankenberg (University of Melbourne), and Fumiaki Sugahara (RIKEN Institute, Japan).

I would like to thank following people who always inspire me to work on biology field. Firstly, my parents Nopparat Sukparangsi and Passara Maitreejit, my brother Akara Sukparangsi, my sister Thanthong Sukparangsi and Porjai Sukparangsi have always given unlimited supports and love and encouraged me to continue working in what I would like to do. Secondly, my grandmother Nittaya Maitreejit and my aunt Wanida Maitreejit had

supported me in education, in particular during my highschool at Assumption College Sriracha (Chonburi, Thailand) and undergraduate study at Chulalongkorn University (Bangkok, Thailand). She inspired me to get into this Thailand top university and supported me all the ways until I got scholarship to study aboard. Thirdly, lecturers at the Department of Biology, Chulalongkorn University in particular Assistant Professor Pongchai Harnyuttanakorn who gave me great opportunities to start working in wetlab since the first year of my undergraduate study. Following four years of malaria research under supervision of Pongchai, I have learned how to work in science and make me love this field even more, which finally encouraged me to consider PhD program aboard. Next, teachers at Assumption College Sriracha, in particular Desha Kunthajit who firstly taught me and let me to new exciting world of biology arena.

Without doubt, this new life of my science research can be possible because of two funding bodies: Royal Thai Government Scholarship DPST (Development and Promotion of Science and Technology Talents Project) and Danish Stem Cell Centre (DanStem), University of Copenhagen.

Lastly, I would like to thank following friends who always help to maintain my positive attitude towards life: Kittichate Visuttijai (Kitti), Kulabutr Komenkul (Tam), Phunpiti Bhovichitra (Pat), Neil Disney, Chinnavuth Vatanashevanapakorn (Jo), Pannarai Premwan (Pui), Ployphan Nipayaporn (Ploy), Orasa Achayapunwanich (Lek), Chatiya (Kwan), Wasin Sakulkoo (Gough), and Teeranun Rodsiri (Por).

DECLARATION

I declare that:

1. I composed this thesis.

2. The work presented in this thesis is of my own, unless otherwise stated.

3. This work has not been submitted for any other degree of professional qualification.

Woranop Sukparangsi

ABSTRACT

Oct4, a transcription factor belonging to the fifth class of POU proteins (POUV), plays essential roles in the maintenance of pluripotency, differentiation and the generation of induced pluripotent stem cells (iPSCs). Oct4 regulates two levels of pluripotency, which are distinguished by their gene expression profiles and epigenetic status, namely the naïve and primed state of pluripotency. Embryonic stem cells (ESCs) and embryonic germ cells (EGCs), which are isolated from inner cell mass and primordial germ cells in the embryo, respectively, are *in vitro* models in which the naïve state is propagated through self-renewal. Epiblast stem cells (EpiSCs) and traditional human ESCs have gene expression profiles that are closest to the post-implantation epiblast, which is closer to embryonic differentiation, and exhibit a primed state of pluripotency. As Oct4 is important for pluripotency in all these cell types, where it regulates different targets, it appears to have two distinct sets of functions, namely germ cell/naïve ESC-like activity and epiblast/primed pluripotency-like activity. Based on protein sequences and syntenic gene analysis, Oct4/POUV homologs of jawed vertebrates can be classified into two subfamilies: POU5F1 and POU5F3, which are thought to originate from a genome duplication event that occurred in a common ancestor. Most extant vertebrates have lost one of these paralogs, while a small fraction, including coelacanths, axolotls, turtles, and marsupials, retains both POUV forms.

In my thesis, I investigated the gene duplication event that underlies divergence of POU5F1 and POU5F3 in both expression pattern and specialised function. In particular, I focused on species that have retained both genes and asked whether POUV functional divergence correlates with ancestral origin. To test the function of POU5F1 and POU5F3, I substituted endogenous mouse Oct4/Pou5f1 with different POUV proteins using a cell line in which endogenous Oct4 expression can be silenced with tetracycline (ZHBTc4). Results showed that POU5F1 proteins had a greater capacity to support naïve ESC pluripotency and

self-renewal than POU5F3 proteins. Global transcriptome analysis of the POUV-rescued ESC lines revealed that coelacanth POU5F1 protein regulates gene expression in a similar manner to mouse Oct4, in that genes involved in stem cell maintenance, reproduction and development are upregulated in ESCs rescued by POU5F1, but not POU5F3. Coelacanth POU5F3 rescued lines, however, expressed genes involved in various cell differentiation programs, including cell adhesion (e.g. E-cadherin and N-cadherin). This suggests that POU5F3 plays a role in primed pluripotency, while POU5F1 regulates naïve pluripotency.

However, there is one POU5F3 factor that rescues ESCs like Oct4, the *Xenopus* gene Xlpou91 (Pou5f3.1). In *Xenopus*, a further duplication of POU5F3 gene enabled specialization, and Xlpou91 is expressed specifically in the primordial germ cells. Xlpou25 (Pou5f3.2) exhibits epiblast-specific activities and lacks the capacity to maintain naïve ESC pluripotency, similar to other POU5F3 proteins. This functional distinction between the different *Xenopus* POUV paralogs enabled us to address how specific Oct4 functions (germ cell-like versus epiblast-like activity) are related to the induction of pluripotency. To address this question, mouse Oct4 was replaced by either Xlpou91 or Xlpou25 in murine cellular reprogramming using a Nanog-GFP reporter line to monitor iPSC generation. Results showed that Xlpou91 and mouse Oct4 were required at similar levels to reprogram somatic cells toward iPSCs and reprogrammed cells emerged with similar kinetics. Conversely, Xlpou25 was required at higher expression levels and the resulting iPSCs appeared at a later timepoint, while the pluripotent population in these cultures appeared to be less stable and more prone to differentiate. I found that this phenotype of enhanced differentiation in Xlpou25 reprogrammed cultures may be a product of a different set of immediate early genes induced at the first stages of differentiation. Global transcriptome analysis of the naïve ESC-like pluripotent subpopulation of these iPSC lines confirmed the capacity of all *Xenopus* POUVs to drive reprogramming towards the pluripotent state. However, the gene sets induced by both Xlpou91 and mouse Oct4, but not Xlpou25, were somewhat enriched

for genes involved in reproduction, emphasizing the segregated role of Xlpou91 as a germ cell specific POUV protein.

Lastly, I explored the evolutionary origin of these two *POUV* paralogs and attempted to identify a *POUV*-related gene in jawless vertebrate (cyclostomes). Based on *in silico* analysis of genomic and transcriptome databases, my collaborators and I were able to identify a single *POUV* gene in the Japanese/arctic lamprey, thus providing the first insight into the origin of gnathosome *POUV* genes.

LAY SUMMARY

Embryonic stem cells (ESCs) are pluripotent stem cells, which are capable of giving rise to all somatic cell types (e.g. neurons, blood cells, intestinal cells). This capacity is called pluripotency. Somatic cells can also be converted back to an ESC-like state through a technology called cellular reprogramming. These converted cells are called induced pluripotent stem cells (iPSCs). A particular protein called Oct4 is a transcription factor essential for the maintenance of ESC pluripotency and inducing cellular reprogramming. In mouse embryos, there are two tissues expressing Oct4, which are the epiblast, the origin of all cell types of the adult body, and the germ cells, the origin of sperm and oocytes. Interestingly, other vertebrates' embryos, show a similar expression pattern of proteins with an amino acid composition similar to that of mouse Oct4. These proteins are called Oct4 homologs. Here I explore whether these proteins from different vertebrates have the same functions as mouse Oct4 and ask how these proteins have become specialised in vertebrates carrying more than one Oct4 homolog.

In my thesis, I employed two approaches to address this question. First, I used genetically modified mouse ESCs, whose level of Oct4 protein can be artificially regulated. Thus, I can address the functional similarity/distinction of Oct4 homologs from other animals by just replacing Oct4 in ESC and observing whether the cells retain stem cell character. With this approach, I can address the question of how functionally similar or distinct Oct4 homologs from other animals are, compared to mouse Oct4. Secondly, I replaced mouse Oct4 with Oct4 homologs in mouse cellular reprogramming, which is generally induced by mouse Oct4 and three other transcription factors: Sox2, Klf4 and c-Myc. By exploring these two functional tests for Oct4, I found that Oct4 homologs known to be expressed in germ cells were effective in both of these assays, suggesting that some

aspects of ESC pluriptoency has evolved from the gene expression program that evolved to control our reproduction. This study provides evidence that the mammalian stem cell network might be an ancient toolkit and originated as early as 400 million years ago before the emergence of mammalian lineages, and I present some computer based analysis that suggests that a single ancesteral Oct4 gene exist at the beginning of vertebrate evolution.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1
INTRODUCTION

CHAPTER 1 GENERAL INTRODUCTION

Oct4 or Oct43/4, a transcription factor encoded by the gene *Pou5f1*, is a central regulator of early embryonic development, differentiation potency and cellular reprogramming. Similar to other POU proteins, Oct4 contains a bipartite DNA binding domain including a POU-specific domain (POU_S) and POU homeodomain (POU_{HD}) that together classify Oct4 as a member of class V POU proteins (POUV proteins). In addition to its role in stem cells, Oct4 has important roles in at least three developmental stages: early pre-implantation, post-implantation epiblast and germ cells. Based on conserved POU domain sequences, several members of POUV proteins have been identified and their expression in early embryonic development is conserved among vertebrate species. In my thesis, I focus on understanding how the remarkable activity of Oct4 evolved, how different roles in development may explain its capacity to sustain and induce stem cell potency.

To better understand the key findings of my thesis, this chapter will introduce you to early mouse embryonic development and transcriptional regulation of pre- and post-implantation of mouse embryos (Section 1.1-1.2). Pluripotent stem cells from these early stages of mouse embryos can be isolated and maintained *in vitro*. In section 1.3, I describe the nature of pluripotent stem cells derived from different stages of development, and discuss current views on the extrinsic and intrinsic factors regulating pluripotency and differentiation. In section 1.4, I describe how cells can be reprogrammed from adult somatic cells to reactivate the pluripotency network by exogenous expression of the Yamanaka factors Oct4, Sox2, Klf4 and c-Myc. While recent work has suggested that other factors or combinations of Yamanaka factors can be used in reprograming, Oct4 is the most consistent requirement in most reprogramming protocols. In section 1.5-1.6, I provide an in-depth discussion of Oct4's roles in mouse embryonic development and pluripotent stem cells. In section 1.7-1.8, I review the evolution of vertebrates and the evolution of Oct4-related

proteins among different vertebrate taxa, Oct4/POUV expression profiles during early vertebrate embryonic development and conserved Oct4/POUV activity from the literature.

Section 1.1 Introduction to murine embryonic development

1.1.1 Morphological changes during pre-implantation development

Figure 1.1A summarises early murine pre-implantation embryonic development. After fertilization, the zygote initiates mitotic cell division, giving rise to smaller individual cells called blastomeres. Once the blastomeres proliferate and reach the 8-cell stage, a phenomenon called compaction takes place. During compaction, the blastomeres become polarised, acquire a strong adhesion between each other and the embryo becomes a solid ball of cells. Compaction results in a 16-cell stage morula, which contains the first two distinct cell types: apolar cells, which are located inside the embryo (inside cells) and polarised cells, which are located on the outside facing the extracellular environment (outside cells). The inside cells of the 16-cell stage morula develop into the inner cell mass (ICM), which will eventually contribute to the embryo proper. The descendants of polarised outside cells become trophoblast cells, which will eventually give rise to the extra-embryonic lineages of the placenta (Boroviak and Nichols, 2014; Hirate et al., 2013). During transition from the morula to the blastula stage, a process called cavitation takes place, during which trophoblast cells secrete fluid into the intercellular cavity, pushing the ICM to one side of a monolayer trophoblast cells and generating a blastocoel. This mammalian-specific structure of cavitated blastula is called the blastocyst. Trophoblast cells that are directly overlying the ICM are called polar trophectoderm, while trophoblast cells at the bottom of the blastocoel are called mural trophectoderm, which later endoreplicates to form mononuclear polyploid trophoblast giant cells (Rossant and Cross, 2001).

The cells in the ICM at the early blastocyst are composed of a heterogeneous mix of epiblast and primitive endoderm (PrEN) precursors (Chazaud et al., 2006). At the late blastocyst stage (E4.5), presumptive epiblast and PrEN physically segregate. PrEN forms a monolayer of cells on the surface of the epiblast (close to the blastocoel) and subsequently proliferates along the wall of the blastocoel. The PrEN adjacent to the mural trophectoderm will form parietal endoderm, while PrEN adjacent to the epiblast and polar trophectoderm will form visceral endoderm after implantation. The epiblast layer contains cells with the potential to become all somatic cell types and germ cells. The capability of ICM and epiblast to give rise to all embryonic germ layers including the germ line is called pluripotency, while the early cells of cleavage stages can give rise to both the embryo proper and the trophectoderm, this ability is called totipotency. Remarkably, ICM and epiblast cells can be isolated and maintained *in vitro* as pluripotent stem cells.

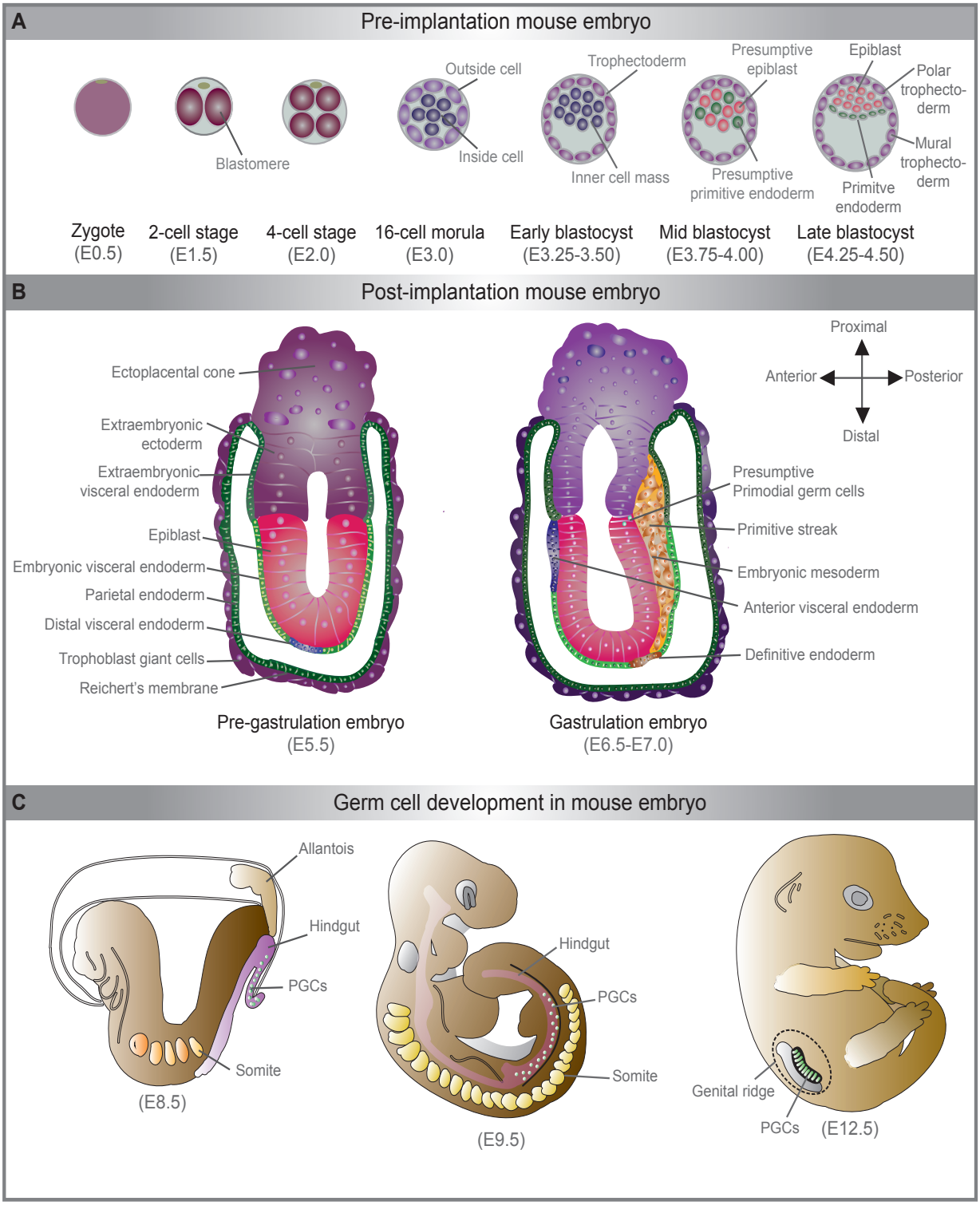


Figure 1.1 Schematic overview of early mouse embryonic development and germ cell development

A) Pre-implantation mouse embryonic development from embryonic day (E) 0.5 - 4 is illustrated. After fertilization, zygote proliferates and gives rise to smaller cells called blastomeres. The first lineage segregation takes place from compaction (8-cell stage) until early blastocyst when inner cell mass and trophectoderm are formed. The second lineage segregation takes place during blastocyst stage when epiblast and primitive endoderm are formed.

B) Early post-implantation mouse embryonic development before and during gastrulation is illustrated. Trophectoderm gives rise to extraembryonic ectoderm and ectoplacental cone while primitive endoderm gives rise to visceral endoderm and parietal endoderm. The epiblast undergoes rodent-specific morphogenesis and forms egg cyclinder-like structure. Complex transcriptional network happening during gastrulation determines lineage specification. Distal visceral endoderm and later anterior visceral endoderm help to specify embryonic axes: promixal-distal and anterior-posterior, respectively. During gastrulation, cells at the posterior side of the embryo undergoes changes in cell morphology and form a structure called primitive streak, where cells ingress and migrate underneath the epiblast layer to form mesoderm and intercalate with visceral endoderm to form endoderm.

C) Germ cell development during primodial germ cells (PGC)' migration towards genital ridge is illustrated.The PGCs are firstly found as a small cluster of approximately 40 cells at the base of the incipient allantois at E7.25. The PGCs then migrate to the developing hindgut endoderm, mesentery, and finally colonize at the genital ridges.

1.1.2 Morphological changes during early post-implantation development

Figure 1.1B summarises early post-implantation embryonic development in mouse. During implantation, the epiblast cells proliferate and undergo epithelial morphogenesis. The polar trophectodermal cells form the extraembryonic ectoderm (ExE) and ectoplacental cone, which will eventually form the placenta, an organ connecting the embryo to the mother's uterus (Rossant and Cross, 2001). At the same time, the PrEN gives rise to the parietal endoderm that extends around the outside of the embryo forming Reichert's membrane (Salamat et al., 1995), and the visceral endoderm (VE) that remains adjacent to the epiblast and extra-embryonic ectoderm. The epiblast layer is composed of pluripotent cells, which can give rise to all embryonic lineages including germ cells, and some extraembryonic tissues. From E6.5, there are a series of elaborate and co-ordinated cellular processes driving extensive morphological movements of the epiblast to form embryonic germ layers. This process is called gastrulation. However, before the onset of gastrulation, a number of important patterning events occur in the visceral endoderm layer that results in the establishment of the embryonic axes. At the distal end of the embryo, distal visceral endoderm (DVE) start to express a large number of Wnt and Nodal antagonists, shielding the distal epiblast from signals produced on the proximal side of the embryo. This creates the first embryonic axis: promixal-distal (P-D) axis at E5.5. The DVE then migrates to the anterior site of the embryo to form anterior visceral endoderm (AVE), which then ensures that the forebrain is formed, providing the first indication of the anterior-posterior axis formed at E6.0 (Arnold and Robertson, 2009).

Gastrulation begins with high levels of Wnt and Nodal signaling in the proximal posterior epiblast leading to the formation of a structure known as the primitive streak (PS), through which epiblast cells will migrate to form mesoderm (future bone, muscle, blood and cartilage), endoderm (furure gut and visceral organs) and the posterior neural tube. At the anterior most point of the PS is the node, the mammalian equivalent of the Spemann-Mangold organiser, an important embryonic signaling centre that is responsible for patterning of the emerging central nervous system (Arnold and Robertson, 2009; Grubb, 2006).

To commence migration through the PS, epiblast cells undergo a morphological change called epithelial-mesenchymal transition (EMT). EMT involves a number of molecular and cellular changes including loosening of epithelial adherens junctions, loss of association with the basement membrane and rearrangement of the cytoskeleton (Nakaya and Sheng, 2008; Nakaya et al., 2013; Williams et al., 2012). EMT allows gastrulating epiblast cells to delaminate and migrate out through the streak, enabling them to form mesoderm or to intercalate with the underlying visceral endoderm to form the future definitive endoderm (Viotti et al., 2014). Different parts of the primitive streak receive different intensities of signals, which result in different allocations of future cell fates along the anterior-posterior axis. The cells ingressing at the early PS stage (posterior PS) will form associated with extra-embryonic structures like the chorion, visceral yolk sac mesoderm and blood islands. Cells from the intermediate and anterior part of the PS form lateral plate mesoderm, paraxial mesoderm and cardiac mesoderm. The cells from the most anterior tip of the PS (late PS) form midline axial mesendoderm including prechordal plate, notochord, node and definitive endoderm (Arnold and Robertson, 2009).

1.1.3 Germ cell development in mice

Figure 1.1C summarises key events of germ cell development in mouse embryo. Germ cells are the source of gametes, both oocytes and sperms. In the animal kingdom, germ cell lineages can be specified either by maternally inherited determinants (preformation) or by inductive signals (epigenesis) (Extavour, 2003). The evolution of modes of germ cell specification in different vertebrates is described in section 1.4.3. In murine development, the most proximal posterior epiblast at gastrulation is induced to switch on the primodial germ cell (PGC)-specific program in the response to localised signaling (described in section 1.2.3). The PGCs are firstly found as a small cluster of approximately 40 cells at the base of the incipient allantois at E7.25. The PGCs then migrate to the developing hindgut endoderm, mesentery, and finally colonise at the genital ridges (figure 1.3a) (Hayashi et al., 2011; Saitou and Yamaji, 2012). During the early stages of PGC development, prior to their colonization at the genital ridges, PGCs undergo epigenetic reprogramming, including a genome-wide DNA de-methylation and re-acquisition of pluripotency (Saitou and Yamaji, 2012). Later, in the gonads, the PGCs begin sex-specific differentiation programs: either spermatogenesis or oogenesis.

Section 1.2 Transcriptional network regulating early mouse embryonic patterning

1.2.1 Transcriptional regulation in the pre-implantation embryo

Transcriptional regulation of early murine pre-implantation embryo is summarised in figure 1.2. Position-dependent Hippo signaling is an essential component of the first lineage segregation event, in which the inner ICM cells become distinct from the outside TE cells. In the inside cells, Nf2/Merlin-Angiomotin (Amot) can form a complex with basolateral adherens junctions, and recruit Lat1/2 that results in the activation of Hippo signaling and priming the cell towards an ICM fate allowing the expression of core pluripotency genes, Oct4, Sox2, Klf4 and Nanog (Hirate et al., 2013). On the other hand, the outside cells become polarised and cell polarity/Par-aPKC signaling sequesters Amot away from basolateral adherens junctions to the apical membrane, that leads to the inactivation of Lat1/2 inhibiting Hippo signaling. Thus, Yap1 can be localised to the nucleus to act as a co-activator of Tead4 to induce the expression of the trophectoderm program (e.g. *Cdx2* and *Gata3*).

ICM cells rapidly show mosaic gene and protein expression of two distinct populations. The lineage determining transcritpion factors Nanog (epiblast) and Gata6 (PrEN) become mutually exclusive as the blastocyst develops, with the presumptive epiblast becoming $Nanog^{+}Gata6^{-}$ and presumptive PrEN becoming $Gata6^{+}$ $Nanog^{-}$ (Chazaud et al., 2006). This progressive process of lineage segregation appears to be mediated by Fgf signaling (Cheng et al., 1998). Based on a number of experiments with culturing embryos in the presence and absence of Fgf signaling, it has become apparent that Fgf4 promotes a PrEN cell fate and the Fgf4 mutant embryos lack PrEN layer and comprise only epiblast

cells at the time of implantation (Kang, 2013). Presumptive PrEN cells initally express slightly higher levels of Fgfr2, making them more receptive to Fgf signaling, while presumptive epiblast cells express higher levels of secreted Fgf4 to stimulate PrEN identity in the neighbouring FgfR2 expressing cells (Arman et al., 1998). The transcription factors Nanog and Gata6 reinforce this loop by cross-regulating Fgf components and themselves leading to an amplification of an initially small distinction in their expressions. The final result of the antagonistic effect of these transcription factors is to create a salt and pepper distribution of epiblast (Nanog and Sox2 positive cells) and PrEN (Gata6, Gata4, Sox17 and Pdgfrα positive cells) in the ICM (Chazaud et al., 2006; Frum et al., 2013; Plusa et al., 2008). During blastocyst maturation, these cells segregate based on differential adhesion, resulting in the localization of PrEN cells at the surface of the blastocoel. The rearrangement of PrEN involves actin-dependent cell movements, retention of positional information, epithelialization and apoptosis (Gerbe et al., 2008; Meilhac et al., 2009; Plusa et al., 2008). PDGF signaling is required for PrEN survival during segregation (Artus et al., 2013).

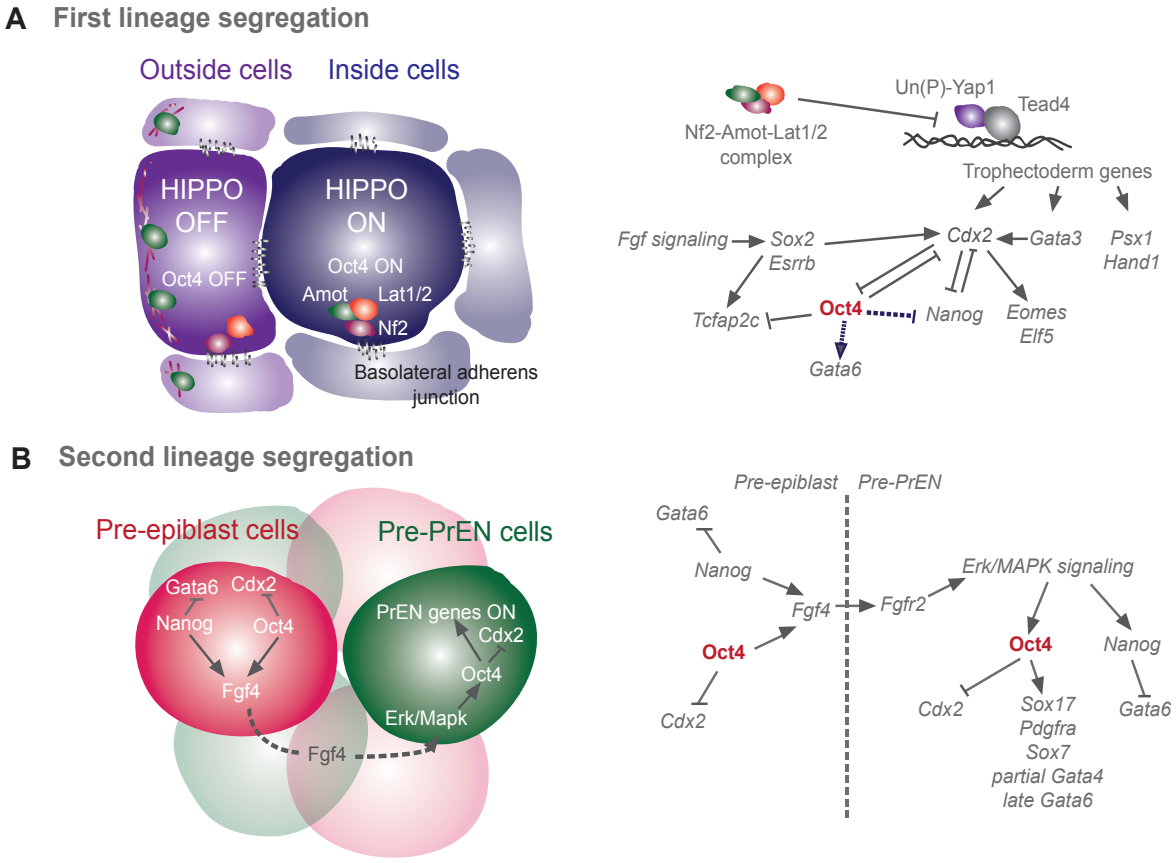


Figure 1.2 Transcriptional regulation of pre-implantation mouse embryo

A) The first lineage segregation of inner cell mass (ICM) and trophectoderm (TE) is specified by position-dependent Hippo signaling, The Nf2/Merlin-Angiomotin (Amot)-Lat1/2 complex links Hippo signaling to basolateral adherens junctions (E-cadherin-β-catenin-α-catenin). The Nf2/Amot/Lat1/2 complex phosphorylates Yap1 and blocks its nuclear localization. Yap is an essential co-activator of TEAD4 to turn on trophectodermal (TE) gene expression. In outside cells, Amot is sequestered to the apical membrane, thus Yap1 is unphosphorylated and can localize to the nucleus and activate TE programs (expression of Cdx2, Gata3 etc). Whereas, in the inside cells Yap1 is phosphorylated and is unable to translocate to the nucleus, thus TE circuitry is blocked and inner cell mass programs (expression of Oct4, Nanog etc) is instead switched on.

B) The second lineage segregation of epiblast and primitve endoderm (PrEN) is specified by FGF signaling. Nanog and Oct4 upregulates the expression of secretory Fgf4, which in turn induces PrEN gene expression (e.g. Sox17, Sox7, Gata4, Gata6) of the neighboring cells mediated by Erk/MAPK signaling and Oct4. This creates mosaic/salt and pepper pattern of presumptive epiblast (Nanog^{+}) and presumptive PrEN (Gata6^{+}) in the inner cell mass. Nanog^{+}cells and Gata6^{+} cells are later segregated to form epiblast and PrEN at the late blastocyst.

1.2.2 Transcriptional regulation in the early post-implantation embryo

Transcriptional regulation of early murine post-implantation embryo is summarised in figure 1.3. Once the blastocyst implants into the uterus, Nodal-Smad2 signaling begins to pattern the embryo leading to gastrulation. Epiblast cells release Nodal ligands which are maturated by the ExE-secreted enzymes Spc1 and Spc4. The VE receives Nodal signal from the epiblast and in turn produces more Nodal at the distal tip, where the initial levels of Nodal signaling are the highest and Nodal induces inhibitors of its own signaling including Cer1, Lefty1 and Dkk1. The production of these inhibitors in the so called distal VE (DVE) shields the distal end of the expanding epiblast from the effects of Nodal signaling distinct domains. Following its induction, the DVE migrates toward what will be the future anterior side of the embryo to become the anterior VE (AVE), where it will shield the prospective forebrain from Nodal and Wnt signaling, establishing the beginning of the A-P axis. Relocation of DVE to the anterior side of the embryo therefore generates a gradient of Nodal/Wnt signaling along the anterior-posterior axis. This combination of localised high levels of Nodal-Smad, canonical Wnt, and BMP (from ExE) signaling at the posterior site is required for primitive streak (PS) formation and the beginning of further lineage specification (Arnold and Robertson, 2009).

In the promixal posterior epiblast, closest to the extraembryonic ectoderm, where cells in the pre-gastrula epiblast receive the highest level of Nodal/Wnt signaling, PS formation is initiated. The PS extends dorsally, and cell lineages are specified based on the proximity of cells to the PS and where they enter it. Graded Nodal signaling, which is highest around the PS, induces differentiation in a dose-dependent manner. High levels of Nodal/Smad2/3/Wnt signaling initially specify axial derivatives at the very anterior of the PS (Vincent et al., 2003). These cells will migrate forward along the midline and differentiate

into the anterior definitive endoderm (ADE) and prechordal plate. The ADE will eventually displace the AVE and form the ventral foregut. Initially it produces the same antagonists as the AVE, but also has a role in patterning the anterior neural plate. The more posterior definitive endoderm is also produced by relatively high levels of Nodal/Wnt signaling and these cells will undergo an EMT and intercalate with the VE, to make the remainder of the embryonic gut. Slightly lower levels of Nodal signaling induce the formation of node and the axial cells migrating forward from the node will form the notochord and the floor plate of the neural tube. Low levels of Nodal/Smad2/Smad3 specify paraxial and laterial plate mesoderm (Chu, 2004; Dunn, 2004). Epiblast cells on the future anterior side of the embryo do not ingress through the PS and differentiate to make neuroectoderm (Lawson, 1999).

Before epiblast cells begin ingressing through the PS, these cells are linked via adherens junctions mediated by E-cadherin. E-cadherin is essential for the epithelial integrity throughout early mammalian development, but its downregulation is essential for the majority of movements associated with gastrulation. The downregulation of E-cadherin is mediated by a number of gastrulation stage signaling pathways including Fgf8, Nodal and Wnt signaling. In particular, Fgf8 is released from the visceral endoderm (VE) and stimulates Fgfr1 in the region of the PS resulting in the induction of *Snai1*, which in turn downregulates *Cdh1* (E-cadherin) transcription (Arnold and Robertson, 2009). Gastrulating cells downregulate E-cadherin and upregulate a N-cadherin. In the mesoderm, gastrulating cells remain N-cadherin postive, but in the endoderm this is not the case. Gastrulating endoderm is identified by the expression of the transcription factor Sox17, and after these cells intercalated into the VE, they form the definitive endoderm and re-express E-cadherin (Viotti et al., 2014), see figure 1.3C.

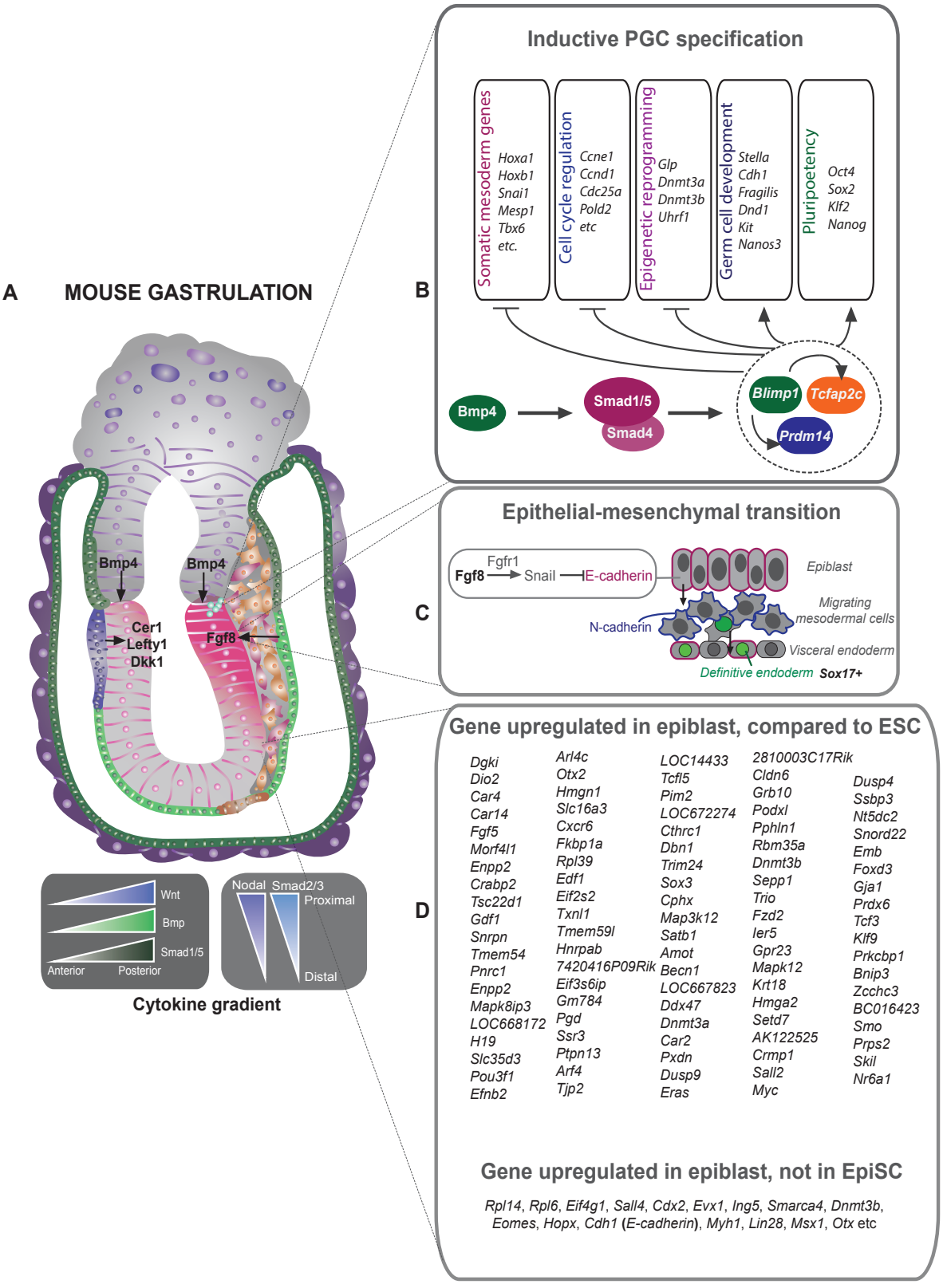


Figure 1.3 Transcriptional regulation of gastrulation and inductive germ cell specification of mouse embryo
(described in section 1.2.2) In gray boxes under the cartoon of mouse gastrulation indicate the gradient of some important cytokines required for lineage specification. Gene lists in D are obtained from Hayashi et al., 2011 and Kojima et al., 2014.

1.2.3 Transcriptional network of murine germ cell specification

Transcriptional regulation of murine germ cell specification is summarised in figure 1.3. Following the initiation of gastrulation, the positioning of the AVE and its production of signaling antagonists ensures that the region of the embryo experiencing the highest levels of BMP signaling is the proximal posterior epiblast. BMP cytokines are released from extraembryonic ectoderm (Bmp4 and Bmp8B) and visceral endoderm (Bmp2) adjacent to this region and high level of BMP signaling leads to the induction of the PGC program. BMPs, like Nodals are members of the TGF-β super family, except they signal through different cell signaling transducers (Smad1 and 5 instead of 2 and 3, see section 1.3.5). In this context, BMP signaling induces the expression of Prdm1 or Blimp1, PR domain-containing protein that is a master regulator of the germ cell lineage. Initially Prdm1 positive cells also express a number of mesoderm-related genes, including the *Hox cluster*, *Snai1*, *Tbx3*, *Tbx6*, *Mesp1*, *Sp5*, *Mixl1*, *Sall3*, *Ccnd1*, *Cdx1*, *Ets1*, *Foxf1*, *Plxna2*, and *Smad7* (Magnúsdóttir et al., 2013; Saitou and Yamaji, 2010), leading to speculation about a common origin of hematopoietic stem cells and PGCs. During this early stage, pluripotency markers including Sox2, Nanog and Zic are not expressed, supporting the notion that the epiblast must be reprogrammed to a pluripotent state during PGC induction. Later, the pluripotency genes are upregulated again alongside with primodial germ cell (PGC)-related markers, while mesoderm specific genes are downregulated.

Prdm14 is another PR domain-containing protein that plays essential roles in the birth of the germ cell lineage (Yamaji et al., 2008). At early stages, Prdm14 is induced independently of Prdm1 although later the maintenance of Prmd14 depends on Prdm1 expression. Interestingly, Prdm14 is also expressed in ICM, from which ESCs can be isolated. It is also highly expressed in pluripotent cells and appears specific to naïve pluripotency (Assou et al., 2007; Saitou and Yamaji, 2010). From the perspective of this thesis, Prdm14 appears specifically expressed in naïve pluripotent cells and in particular in

the PGCs or ICM derivatives. This makes it a very good marker for a PGC-like pluripotent state, whereas Prdm1 is expressed broadly in other tissues.

In addition, Tcfap2c (also known as AP2y) plays key roles together with Prdm1 and Prdm14 in germ cell specification (Weber et al., 2009). These three regulators form a tripartite early germ cell network that is essential for downstream expression of the following PGC specific transcripts: *Nanos3*, *Dnd1*, *Rhox9*, *Kit*, *Dppa3*, *Dazl*, *Ifitm3*, *E-cadherin (Cdh1)*, *Nanog*, *Sox2*, *Oct4 (Pou5f1)*, *Sox3*, *Elf3*, *Elk1*, *Isl2*, *Mycn*, *Klf2*, *Ddx4*, *Fiat*, *Sp8*, *Smad3*, *Zic3*, *Tcfe3*, *Epc1*, *Six4*, *Eya3*, *Ccne1*, *Ccnd1*, *Cdc25a* (Magnúsdóttir et al., 2013; Saitou and Yamaji, 2010). As mentioned above, the reinduction of pluripotent gene expression suggests that PGC induction involves some form of reprogramming. This also includes a number of other global chromatin and epigenetic changes including genome-wide DNA de-methylation, erasure of parental imprints, and re-activation of the inactive X-chromosome. These changes may be regulated by a set of chromatin modifiers and DNA demethylases that are upregulated during PGC specification, e.g. *de novo* DNA methyltransferases 3a and b (Dnmt3a, Dnmt3b) (Saitou and Yamaji, 2010).

While reprogramming event appears to occur during PGC specification, several of these pluripotent/PGC factors are maintained in spermatogonia in testis, for example, c-kit, Vasa (MvH), Dazl, Stra8, Epcam, Oct4 (Pou5f1), Nanos2, Nanos3, Ngn3, Sox-3, Taf4b, Bcl6b, Numb, Lrp4, Ret, Cdh1, UTF1 and Lin28 (Phillips et al., 2010). Spermatogenesis involves a classical adult stem cell-dependent process. The stem cells in the adult testis are called spermatogonia stem cells (SSCs). Several regulators e.g. Gdnf, Bcl6b, Etv5, Ihx1, Plzf and Taf4b have been shown to play a role in SSC maintenance (Oatley and Brinster, 2008). Interestingly, it has also been shown that the $Oct4^{+}c\text{-}Kit^{+}$ subpopulation shows less capacity in repopulation than the $Oct4^{+}c\text{-}Kit^{-}$ subpopulation. This suggests that Oct4 has roles in maintenance of the stem cell population in spermatogenesis process (Phillips et al., 2010).

Section 1.3 Pluripotent stem cells

Pluripotency refers to the potential of cells to differentiate toward cell lineages of all three embryonic germ layers (ectoderm, mesoderm and endoderm) as well as germ cells. However, it does not include the ability of cells to differentiate to the extra-embryonic lineages TE and PrEN. Cells harboring pluripotency are called pluripotent cells. In the embryo, pluripotent cells exist only transiently, but cell lines such as ESCs can be established *in vitro*. These cells have the ability to indefinitely propagate *in vitro* and exhibit self-renewal, meaning that their pluripotent character is passed through successive rounds of cell division. Pluripotency can be assessed by *in vitro* and *in vivo* differentiation. *In vivo* differentiation assay includes both teratoma formation and chimera generation. During teratoma formation, pluripotent cells are able to form teratomas, tumors containing all the lineages of an embryo, at ectopic sites in the adult. Pluripotent cells can also be introduced into the blastocyst stage embryo and undergo normal embryonic differentiation and contribute to all lineages (chimera formation). A brief introduction to the *in vitro* derivation and differentiation of pluripotent stem cells is illustrated in figure 1.4.

1.3.1 Mouse embryonic stem cells

Pluripotent cells exist transiently in the ICM of the pre-implantation embryo. Under certain conditions, these cells can be expanded *in vitro* to generate immortal, karyotypically normal cell lines known as ESCs. Based on the early stage of development from which these cells are derived, they are referred to as naïve, pluripotent cells. While traditionally these lines have only been generated in mouse, naïve ESCs have recently been derived in human (Ware et al., 2014). Originally, mouse ESCs were cultured in fetal bovine serum (FBS) and on feeder cell lines, but they can now also be cultured in the presence of LIF (the

cytokine produced by the feeders) and Bmp4 (component in FBS). For convenience, ESCs are frequently grown in the serum and LIF. More recently, completely defined conditions employing LIF and two small molecules inhibitors of MEK (mitogen-activated protein or extracellular signal-regulated kinase) and GSK3β (glycogen synthase kinase 3β) (conditions referred to as 2i/LIF) have been used extensively (Ying et al., 2008). Naïve ESCs exhibit the capacity to give rise to all cell lineages including germ cells and, importantly, they demonstrate high chimera contribution. In addition, naïve ESCs display active X chromosomes and differentiate under the activation by Fgf/Erk signaling. The addition of 2i helps to prevent ESCs from differentiation (Nichols and Smith, 2009). Interestingly, it has been recently demonstrated that 2i+LIF indeed maintains ESCs at a similar to totipotent state, and that they have potential to give rise to both embryonic and extraembryonic lineages (Morgani et al., 2013). In addition, ESCs are heterogeneous and this heterogeneity appears to reflect the cell states found in the blastocyst from which they are derived (Boroviak et al., 2014). As a result, a number of the factors expressed in the early epiblast progenitors in the blastocyst are expressed heterogeneously in ESCs (Canham et al., 2010; Hayashi et al., 2008; Niwa et al., 2009; Toyooka et al., 2008; van den Berg et al., 2008). ESC culture also contains a fraction of cells primed toward PrEN differentiation (Price et al., 2013).

1.3.2 Mouse epiblast stem cells

Pluripotent cells can also be generated from the post-implantation epiblast. These cells are known as Epiblast Stem cells (EpiSCs) (Tesar et al., 2007). As these cells are derived from a later developmental stage, they are closer to differentiating and express some germ layer as well as pluripotency markers and therefore they are referred to as “primed.” EpiSCs can be generated from pre-gastrulation (E5.5) to late-bud (E8.25) stages (Kojima et

al., 2014; Nichols and Smith, 2009). EpiSCs have capacity to give rise to all functional soma and germ cells including teratoma formation but the chimera contribution is limited to a specific subpopulation, in particular Oct4 high subpopulation of the EpiSCs (Han et al., 2010). Unlike naïve ESC or blastocysts, one of the X chromosomes in female embryos is randomly inactivated in primed epiblast and EpiSCs (Guo et al., 2009; Heard, 2004).

In addition, EpiSCs require bFgf and Activin for the self-renewal maintenance, and the inhibition of the JAK/STAT and/or the Rho-associated kinase signaling pathway helps to increase the survival of EpiSCs (Chenoweth et al., 2010). Moreover, EpiSCs can be derived from ESCs under EpiSC culture condition, while the reversion of this primed to naïve state requires genetic manipulation (e.g. exogenous expression of Klf4) (Guo et al., 2009). Recent findings of Kojima and colleagues reveal that EpiSCs derived from different stages of post-implantation have similar transcriptomes (Kojima et al., 2014). This suggests that specific populations of epiblast are selected to expand under the EpiSC culture (activin+bFgf). Like ESCs, EpiSCs are heterogeneous and contain populations with different differentiation biases or potency. Thus EpiSCs are not merely epiblast/ectoderm, but exhibit high levels of activity in genes associated with Fgf/MAPK, TGF-β and Wnt signaling. EpiSCs express genes related to anterior mesendoderm/AME and definitive endoderm/DE (*Lefty1*, *Cited2*, *Cer1*, *Lbh*, and *Sox17*) and anterior primitive streak (*Sfrp1*, *Foxa2*, *Chrd*, *Acvr1b* and *Frd8*) than the epiblast/ectoderm (Kojima et al., 2014). Based on in-depth proteomic analysis, Rugg-Gunn et al., 2012 reveals the unique and common surface protein of ESCs and EpiSCs, and found 60 cell-surface proteins unique to ESCs and 256 to EpiSCs. They also confirmed the expression of these genes by immunofluorescent staining of Pecam1, Pvrl2 and Cd81 specific in ESC, while Notch3, Cd40, Cdh10, Sirpa, Cd47 and Cdh2 specific to EpiSCs. Some of these markers are used in this thesis to help distinguish different aspects of pluripotency.

Hayashi et al., 2011 and Kojima et al., 2014 have compared the global transcriptome of gastrulation stage epiblast to epiblast-derived EpiSCs and ESCs using microarray. The list of genes upregulated in these cells are shown in figure 1.3D. Core pluripotency gene like *Oct4*, *Sox2* and *Nanog* are expressed in these cells, while ICM-related genes (e.g. *Dppa2*, *Dppa4*) are downregulated. In addition, epiblast and EpiSCs express many genes related to lineage commitment e.g. *Id2*, *Foxb1*, *Meis2*, *Tbx3*, *T-brachyury*, *Irx2*, *Gbx2*, *Fgfr2*, *Sp5*, *Cdx2*, *Wnt3a*, *Wnt5a*.

While recent work suggests that naïve cells can be generated from human embryos (Ware et al., 2014), the majority of existing human ESC lines resemble mouse EpiSCs in both cell culture requirements and gene expression profiles (Brons et al., 2007; Rossant, 2008; Tesar et al., 2007). Hence, despite derivation from human blastocysts, these embryos have a tendency to progress to the primed epiblast state when cultured *in vitro*.

1.3.3 Mouse embryonic germ cells

Murine germ cells originate during gastrulation before all germ layers are set. The early cells specified with germ cell programs are called primodial germ cells (PGCs), which during development migrate through several parts of the embryo to the genital ridge. During this route of germ cell specification and migration, pluripotent stem cells called embryonic germ cells (EGCs) can be derived. The successful derivation of EGCs depends on the supplement of Stem Cell Factor (SCF) and LIF. Oncostatin M, interleukin-6 and ciliary neurotrophic factor also help to increase the self-renewal capacity of EGCs (Durcovahills and Mclaren, 2004). Interestingly, EGCs have a similar gene expression as naïve state ESCs. Recently, Leitch and colleagues found that ESCs and EGCs have a similar global DNA methylation status. Both pluripotent cells can be cultured in 2i/LIF (MEK and Gsk3b

inhibitor) and show a similar induction of naïve pluripotency markers. Based on whole genome expression analysis of ESCs and EGCs, their gene expression was found to cluster based on culture conditions rather than cell types, suggesting that EGCs and ESCs were indeed very similar (Leitch et al., 2013). As mentioned above the ICM cells of the blastocyst and PGCs express a similar set of transcription factors, so that it is possible that the networks supporting these two cell types are closely related.

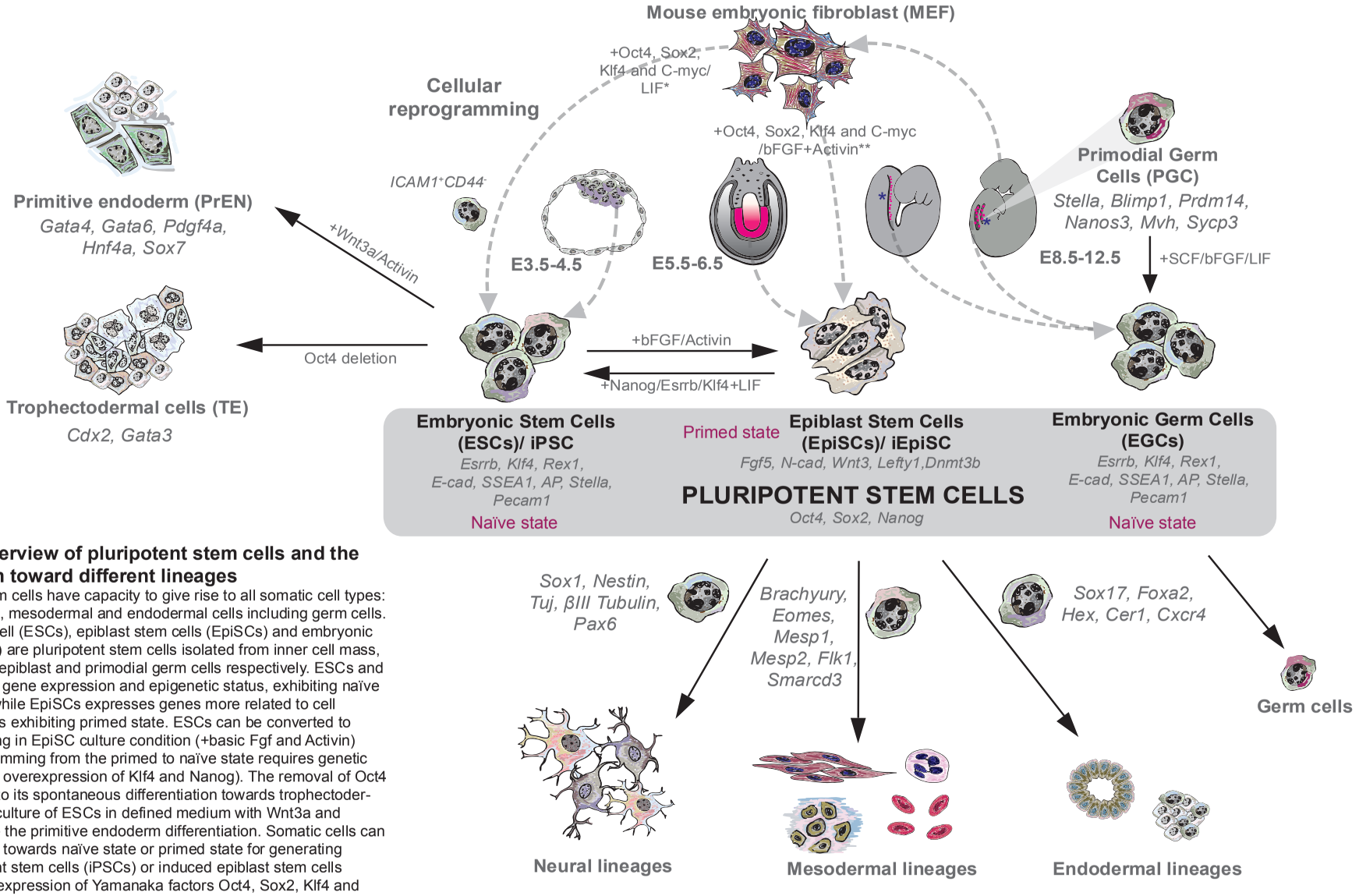


Figure 1.4 Overview of pluripotent stem cells and the differentiation toward different lineages

All pluripotent stem cells have capacity to give rise to all somatic cell types: ectodermal/neural, mesodermal and endodermal cells including germ cells. Embryonic stem cell (ESCs), epiblast stem cells (EpiSCs) and embryonic germ cells (EGCs) are pluripotent stem cells isolated from inner cell mass, post-implantation epiblast and primodial germ cells respectively. ESCs and EGCs share lot of gene expression and epigenetic status, exhibiting naïve pluripotent state while EpiSCs expresses genes more related to cell differentiation, thus exhibiting primed state. ESCs can be converted to EpiSCs by culturing in EpiSC culture condition (+basic Fgf and Activin) while the reprogramming from the primed to naïve state requires genetic manipulation (e.g. overexpression of Klf4 and Nanog). The removal of Oct4 from ESCs leads to its spontaneous differentiation towards trophectodermal lineage. The culture of ESCs in defined medium with Wnt3a and Activin can induce the primitive endoderm differentiation. Somatic cells can be reprogrammed towards naïve state or primed state for generating induced pluripotent stem cells (iPSCs) or induced epiblast stem cells (iEpiSCs) by overexpression of Yamanaka factors Oct4, Sox2, Klf4 and c-Myc. During reprogramming towards naïve state, fully reprogrammed cells exhibit ICAM1^{+}CD44^{-} expression profile.

1.3.4 Intrinsic factors regulating embryonic stem cell identity

The maintenance of pluripotency is controlled by both extrinsic regulators and complex network of transcription factors. While the list of key regulators is rapidly expanding, I will focus on the historical core of Oct4, Sox2 and Nanog (OSN) as the literature about their function is the most extensive. OSN together bind extensively to both pluripotency-related/ differentiation-related genes. These core factors form protein complexes with other transcription factors on enhancers of genes, leading to activation of pluripotency genes or repression of lineage specific genes. Autoregulatory loops formed by OSN balance their own expressions at appropriate levels.

Oct4 plays essential roles during early embryonic development and early lineage decisions. Oct4 mutant embryos exhibit a degenerated inner cell mass and a defect in the formation of embryonic epiblast and extraembryonic endoderm. ESCs cannot be derived from Oct4 mutant embryos (Nichols et al., 1998). In addition, a precise level of Oct4 is required to maintain the ESC state. Reduction or overexpression of Oct4 in ESCs leads to the differentiation toward trophectoderamal cells and primitive endoderm/mesoderm, respectively (Niwa et al., 2000; Radzisheuskaya et al., 2013; Yeom et al., 1996). Oct4 is also required for PGC specification and differentiation (Kehler et al., 2004; Scholer et al., 1990). Oct4 is described in more detail in section 1.8.

Octamer binding proteins like Oct4 are known to bind cooperatively with HMG proteins of the SRY-related HMG box family, and in the case of Oct4, its partner in the regulation of pluripotency is Sox2. Sox2 is expressed in all pluripotent cell types and then later in neural progenitors. Sox2 mutant embryos exhibit epiblast defects and die at peri-implantation stages, although Sox2 mutants also exhibit abnormalities in the placental

lineages, especially in the chorion. Moreover, like the phenotype of Oct4$^{-/-}$, ESCs cannot be derived from Sox2 mutant embryos (Avilion et al., 2003). Sox2 and Oct4 form heterodimer that binds to a cognate Oct/Sox element at many important downstream regulators of pluripotency alongside genes repressing lineage differentiation in ESCs. In addition, the dose of Sox2 is critical for naïve pluripotency maintenance. The reduction and overexpression of Sox2 also induce rapid ESC differentiation, similar to Oct4 (Kopp et al., 2008; Li et al., 2007b; Masui et al., 2007). As mentioned above, Sox2 is expressed in various progenitors of the central nervous system, including retinal progenitor cells, and neural stem cells (NSC), suggesting additional roles in lineage specific self-renewal and multipotency (Miyagi et al., 2008; Taranova et al., 2006).

Nanog is a homeobox transcription factor and an essential mediator of the choice between epiblast and PrEN (Mitsui et al., 2003; Ralston and Rossant, 2009). In the early epiblast, Nanog inhibits the induction of the PrEN program. Nanog is expressed in epiblast progentors from morula stages, but its expression is shut down around the time of implantation. Later, Nanog is expressed in the posterior proximal epiblast around the PS, and then during the differentiation of PGCs. Nanog is essential for the formation of germ cells, both the maturation and migration of PGCs toward genital ridges. Nanog-null ESCs retain their self-renewal ability and can contribute to all lineages in chimeras except the germ cells (Chambers et al., 2007; Mitsui et al., 2003). In addition, ESCs express Nanog heterogeneously similar to what is seen in the blastocyst. The knockout of Nanog makes ESCs more likely to differentiate; whereas, its overexpression enables ESCs to self-renew in the absence of LIF (Chambers et al., 2007; Hatano et al., 2005). While Nanog expression is more homogeneous in 2i culture than it is in serum and LIF, ESCs in 2i culture also contain a subpopulation co-expressing Nanog and Hex, which exhibits totipotency (Morgani et al., 2013).

The three factors discussed above are the three most well known members of the ESC gene regulatory network; however, a number of high-throughput screens have revealed a more exhaustive list of ESC regulators. Additional factors known to be important for supporting pluripotency include (1) transcription factors (Smad1, Stat3, Tcf3, c-Myc, Esrrb, Sall4, Tbx3, Zfx, Ronin, Klf4, Prdm14), (2) transcriptional co-factors (cohesion, PAf1 complex, Dax1, Cnot3, Trim28), (3) chromatin regulators (polycomb group, SetDB1/ESET, esBAF, Chd1, Chd7, Tip60-p400) and (4) non-coding RNAs (miRNAs, GC-rich ncRNAs) (Young, 2011). Remarkably, Smad1, Stat3 and Tcf3, downstream effectors of BMP, LIF and Wnt signaling respectively, co-occupy many genes coordinately with OSN to directly link extracellular signals to the regulation of the pluripotency circuitry (Chen et al., 2008a).

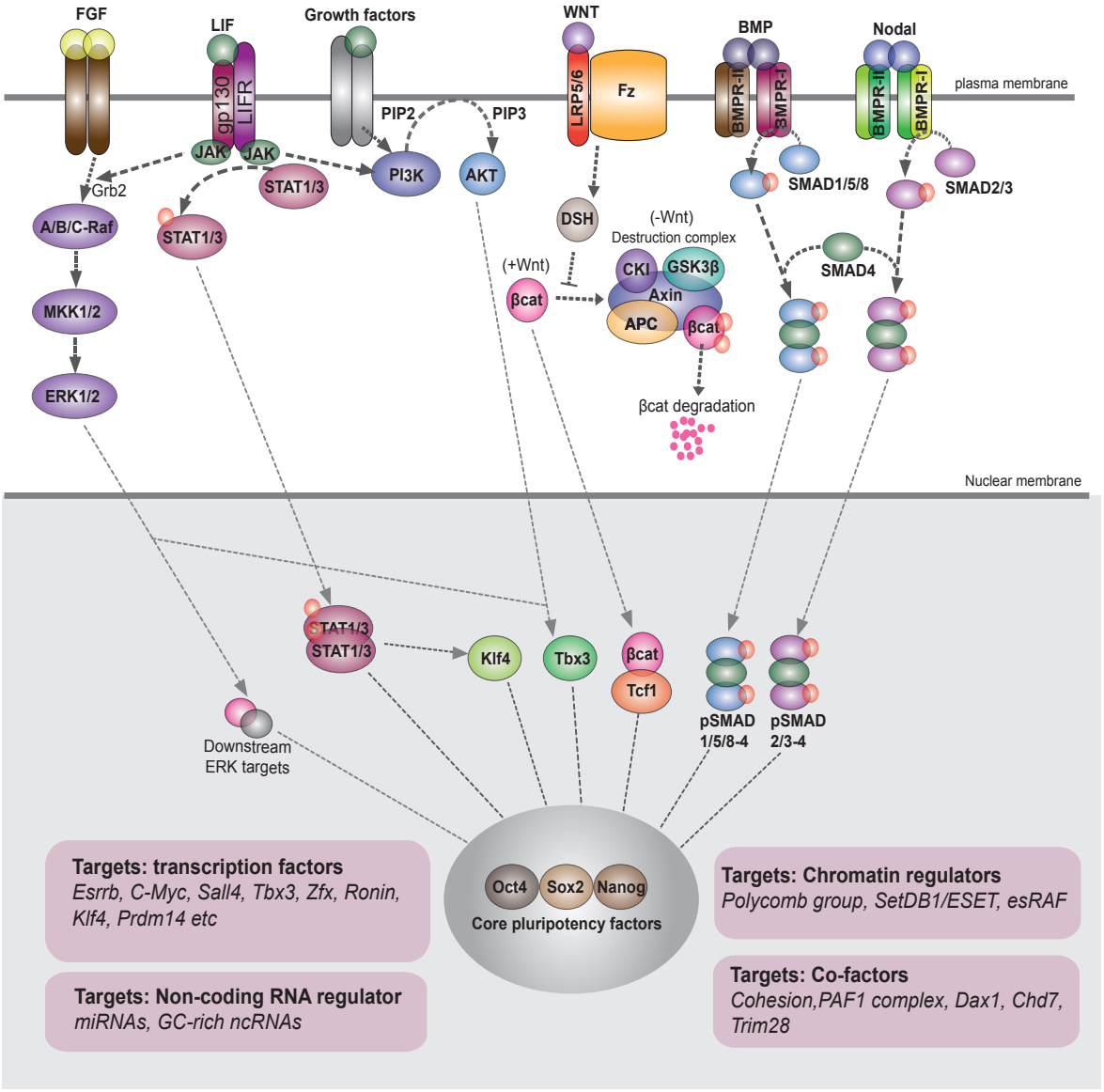


Figure 1.5 Extrinsic and intrinsic factors regulating pluripotency

Oct4, Sox2 and Nanog (OSN) together bind extensively to both pluripotency-related/ differentiation-related genes. These core factors form protein complexes with other transcription factors on enhancers of genes, leading to activation of pluripotency genes or repression of lineage specific genes. Downstream effectors of BMP, LIF and WNT signaling co-occupy many genes coordinately with OSN to directly link extracellular signals to the regulation of the pluripotency circuitry. Details of how each extrinsic signaling is linked to core pluripotent circuitry are described in section 1.3.5.

1.3.5 Extrinsic factor regulating embryonic stem cell identity

a) LIF-STAT pathway

Leukemia inhibitory factor or LIF is a member of the IL6 cytokine family. LIF was identified as the factor produced by feeders required to support ESC expansion. LIF is required for the maintenance of an undifferentiated state of naïve ESCs. However, there are a number of conditions such as Nanog or Klf4 overexpression, or the combination of Erk and Gsk3β inhibition (2i), where LIF is not absolutely required to maintain self-renewal (Chambers et al., 2007; Ying et al., 2008). In a number of these cases, the inclusion of LIF in the culture medium still augments ESC expansion. LIF induces three downstream signaling pathways: (1) Jak/Stat3 (2) PI3K/Akt (3) Grb2/MAPK (Niwa et al., 2009). ESC self-renewal has been shown to be regulated by Jak/Stat3 signaling, and Stat3 has been shown to be required for ICM survival (Do et al., 2013). The LIF receptor gp130 is also required for the maintenance of pluripotency during diapause. Diapause is a process in certain mammals, in which the mother can suspend embryonic development at the blastocyst stage when in the process of weaning or nutrient starvation. When conditions improve, the pluripotent cells of the blastocyst can resume development. Interestingly, induced diapause has been used to improve ESC derivation (Brook and Gardner, 1997; Evans and Kaufman, 1981).

In Jak/Stat3 signaling, phosphorylated Stat1/3 proteins localise to the nucleus and act as a transcription factor, and together with other pluripotency regulatory complexes, they induce genes involved in self-renewal and pluripotency (Chen et al., 2008a). Several Stat3 target genes are identified, including *Klf5*, *Klf4*, *Smad7*, *Gbx2*, *Icam1*, *Sall4*, *Stat3* itself (Bourillot et al., 2009). In particular, LIF has been shown to induce the expression of Klf4, which appears to be essential for murine cellular reprogramming (Stuart et al., 2014).

b) BMP/Nodal-Smad pathway

BMPs (Bone morphogenetic proteins) and Nodal are both members of the TGF-β superfamily of secreted growth factors. TGF-β binds to transmembrane type I and type II serine/threonine kinase receptor complexes, in which the type II receptor is specific for different classes of cytokines. Binding of the cytokines cause phosphorylation of the type I receptor through the kinase activity on the type II receptor, which leads to the recruitment of specific R-Smads to transduce signals to the nucleus. BMP signaling involves Smad1, Smad5, and Smad8, while the interaction of Nodal with its receptor leads to the activation of Smad 2 and 3. These phosphorylated R-Smads form a heteromeric complex with Smad4 (Co-Smad). The R-Smad/C-Smad complex translocates to the nucleus and regulates target gene expression in conjunction with different sequence specific DNA binding proteins. The signaling mediated by R-Smad is also negatively regulated by I-Smads (Smad6 and Smad7).

In mouse ESCs, Bmp4 together with LIF maintains ESC pluripotency and self-renewal. BMP signaling induces the expression of bHLH proteins known as ID proteins, that bind to, and suppress, the activity of differentiation-promoting transcription factors. However, BMP/Smad signaling is dispensable for self-renewal when cells are cultured in 2i (Ying et al., 2008). Several BMP-Smad1/5 target genes are enriched in neural differentiation, which are silenced in naïve ESCs. Nodal, another member of the TGF-β family also has a role in early epiblast expansion. However, it is also a dose-dependent inducer of mesoderm and endoderm. Another TGF-β signaling molecule called activin can substitute Nodal and supports EpiSC self-renewal in the presence of Fgf (Vallier et al., 2009).

c) Wnt-β-catenin signaling

Wnt ligands can transduce the signaling through several pathways, but here I focus solely on the canonical pathway because it has been implicated in the maintenance of pluripotency. Canonical Wnt/β-catenin signaling is regulated through post-translational modification of β-catenin. In the absence of Wnt ligands, β-catenin is bound to a destruction complex, composed of the APC and Axin scaffolding proteins and the glycogen synthase and casein kinases (GSK3β and CK1). Gsk-3β activation leads to a ubiquintin-mediated proteolytic degradation of β-catenin. Upon stimulation of the Wnt pathway through the Frizzled receptor, Gsk-3β activity is inhibited through the activity of Dishevelled (Dsh or Dvl) and the expression of β-catenin is stabilised, enabling it to translocate to the nucleus. In the nucleus, β-catenin interacts with the HMG-domain, containing repressors of the Tcf family, and converts them to be activators by displacing Groucho family co-repressors. In mouse ESCs, canonical Wnt signaling supports self-renewal primarily through blocking differentiation towards-neuroectodermal lineages. Tcf3 is a major component of the OSN network, and it acts to shut down the pluripotency network in the absence of Wnt signaling (Atlasi et al., 2013). In the case of Tcf3, Wnt stimulation does not appear to convert it to an activator, but rather induces Tcf1 to bind with β-catenin and compete Tcf3 away from its targets. In this way, the complex of Tcf1-β-catenin supports ESC self-renewal (Yi et al., 2011).

The intrinsic and extrinsic regulating pluripotency is illustrated in figure 1.5.

Section 1.4 Induced pluripotent stem cell and mechanism of cellular reprogramming

ESC are remarkable cell lines derived from the mammalian embryo with the capacity to both expand indefinitely in culture and differentiate into all adult cell types, a property known as pluripotency. A major goal of regenerative medicine has been to develop the capacity to transform somatic cells to an ESC like state, enabling the generation of patient specific stem cells. Initial efforts to reprogram differentiated, mature somatic cells to a pluripotent embryonic state, was driven by the nuclear reprogramming. Nuclear reprogramming involves the transfer of embryonic factors from an oocyte or ESC to enable the reprogramming of a differentiated somatic nucleus. This was achieved by either transferring the somatic nucleus to an oocyte or ESC, or through cell fusion and the generation of heterokaryons (Gurdon et al., 1958; Hochedlinger and Jaenisch, 2006; Wilmut et al., 1997). A more targeted alternative was developed by Shinya Yamanaka and colleagues and involves the overexpression of specific transcription factors in somatic cells. They identified candidate transcription factors normally expressed in ESCs and used retrovirus to force their expression in differentiated mouse embryonic fibroblasts. Starting with 24 ESC specific transcription factors, they were able to induce an ESC like pluripotent state through the prolonged culture of these infected MEFs under conditions that normally support ESC growth. Of these original twenty-four factors, they found that Oct4, Sox2, Klf4 and C-myc (also called together as "Yamanaka factors") were sufficient to reprogram mouse embryonic fibroblasts to pluripotent cells, which they termed induced pluripotent stem cell (iPSCs). These iPSCs expressed ESC markers such as SSEA-1 and alkaline phosphatase and exhibited comparable level of pluripotency genes such as Oct4, Sox2, Nanog and etc (Takahashi and Yamanaka, 2006). With time, this protocol was improved, such that iPSCs were shown to be able to contribute to the germ line and generate cloned animals. Human

iPSCs have also been generated with either the original set of Yamanaka factors or a combination of Oct4, Sox2, Nanog, and Lin28. Human iPSCs express the same set of transcription factors as human ESCs, exhibit a similar methylation status and differentiation potential (Takahashi et al., 2007; Yu et al., 2007). While initially developed in mouse and human, the protocol with original four factors has been shown to partially induce the fibroblasts into induced pluripotent stem cell like states also in other vertebrates (chicken, quail, zebra finch, zebrafish) and invertebrate (fruit fly) (Rossello et al., 2013). This suggests the conservation of stem-cell like phenotypes and the reprogramming gene network in animal kingdom.

Reprogramming occurs as a result of the progressive acquisition of new cell identities, culminating in the stable establishment of the network of transcription factors normally expressed in pluripotent cells, the pluripotency network. Following viral induction, the four factors induce a mesenchymal-epithelial transition (MET) that is essential for successful reprograming and produces a pre-iPS state, in which some components of the pluripotency network are expressed, although not stably. This is followed by the establishment of a stable state with the capacity to differentiate in chimeras or fully reprogrammed iPSCs that have established naïve pluripotency. Fully reprogrammed iPSCs can both produce germ line transmission and yield iPSC-derived embryos in tetraploid complementation. Moreover at fully state of cellular reprogramming, retrovirally expressed reprogramming factors are efficiently silenced by the induced ESC-like transcriptome, and this enables the faithful testing of these cell lines in a variety of differentiation assays. Full reprogramming is generally associated with the activation of Dio1-Dlk3 locus (Stadtfeld et al., 2010).

In the initial phase of reprogramming, Oct4, Klf4 and Sox2 act as pioneer factors and bind to distal regulatory elements on genes with in DNase resistant closed chromatin.

With time, this binding needs to be converted to recognition of the proximal promoter regions of these genes and this will require co-factors such as the cohesin complex (Taberlay et al., 2011). C-Myc acts as an amplifier and requires OSK binding for its recruitment to specific promoters on active chromatin for transcriptional induction of its target genes. Early targets include many genes associated with promoting reprogramming such as GLISI, mir-302/367, MET promoting factors and apoptosis regulators. Interestingly, although OSK induce abundance of genes at this early state of cellular reprogramming, they do not recognise the gene targets normally expressed in the naïve ESC state. The time taken to convert these immediate early genes from silent to transcribed may be reflected in the observation that the early phase of the reprogramming is thought to be stochastic process (Buganim et al., 2012). In addition, the early upregulation of apoptosis regulators by OSK also explains the high levels of apoptosis known to occur during this early phase of reprogramming and this is consistent with the finding that mutant apoptotic regulator (e.g. p53) increase the chance of successful iPSC generation (Kawamura et al., 2009).

All four reprogramming factors facilitate the induction of MET (figure 1.6). As TGF-β signaling is known to promote EMT (or suppress MET) and TGF-β inhibitors can replace either Sox2 or C-myc. Oct4 and Sox2 appear to inhibit the expression of TGFβR3 and TGFβ3 while stimulating the expression of miR-200, a specific inhibitor of the EMT regulator, Zeb2. Klf4 also inhibits TGFβ3 and stimulates the expression of E-cadherin. C-myc both inhibits TGFβ1 and TGFβR2 expression and induces expression of additional miRNA that target TGFβR2 (Li et al., 2010). Interestingly, overexpression of the MET regulator E-cadherin can replace Oct4 in the reprogramming cocktail (Redmer et al., 2011).

The establishment of the endogenous pluripotency network is a long process, which involves the initial activity of OSK as pioneer factors, binding their targets in silent

chromatin and their slow activation. However, despite the pioneer activity of OSK, they are unable to recognise their endogenous ESC targets in differentiated cells, where large segments of the genome are inaccessible to OSK (e.g. regions containing pluripotency genes Nanog, Sox2, PRDM14) as they are marked with the H3K9me3 histone modification. In somatic cells, these inaccessible targets are marked by H3K9me3. The H3K9 methyl transferase SUV39H1/2 is required for maintenance of these domains and its inhibition in reprogramming increases the access of OSKM to sites within these H3K9me domains. The loss of this mark appears to be an inefficient process that is influenced by culture conditions including response to the reprogramming block induced by Bmp4 (Chen et al., 2013) and these regions include nearly all the hotspots of aberrant DNA methylation that appear, when iPSCs are compared to ESCs. The H3K9 demethylases Kdm3/4 are also essential for the transition from pre-iPSCs to iPSCs. Moreover, Oct4 itself is a potent activator of two additional H3K9 demethylases: Jmjd1a and Jmjd2c.

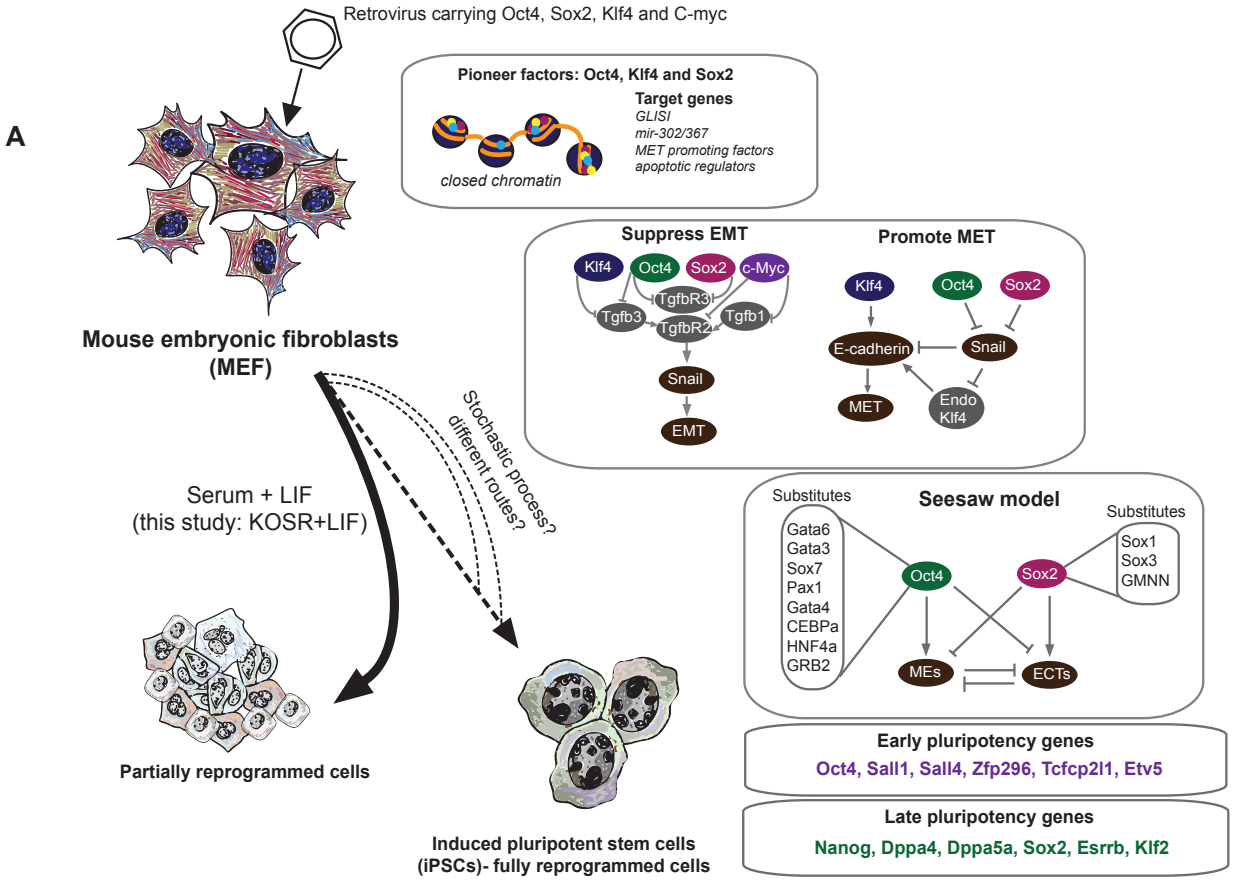
The transition between pre-iPSCs and fully reprogrammed iPSCs is not rapid, and a combination of single cell analysis and the identification of specific precursor populations suggest a degree of heterogeneity in reprogramming in the favor of stochastic processes. The pluripotency network itself is activated in stages with the transcription of endogenous Sox2 and Esrrb coming very late in the process. The expression of late pluripotency genes marks cells as close as possible to being fully reprogramming state (Buganim et al., 2012; O'Malley et al., 2013).

One of the current views of reprogramming and the establishment of pluripotency involves a conflict between distinct lineage specific transcription factors in the same cell. Pluripotent cells are also maintained or induced by conflicts between transcription factors specifying different lineages and this implies that a number of early lineage specifiers should be able to induce pluripotency, as it was recently shown. In addition, this model suggests that

the reprogramming factors (O, S, K) are also acting on their normal developmental programs. For example, Oct4 is believed to induce mesendodermal (ME) and inhibit ectodermal (ECT) genes, while Sox2 induce ECT and inhibit ME genes. The balance between ME versus ECT genes leads to a "seesaw model ", in which the contrasting activities of these factors creates a road that leads to the acquisition of pluripotency. Interestingly, ME-specific transcription factors (Gata6, Gata3, Sox7, Pax1, Gata4, CEBPa, HNF4a, and GRB2) can replace exogenous Oct4 in the iPSC induction; whereas ECT-specific transcription factors (Sox1, Sox3, and GMNN) can replace exogenous Sox2 (Shu et al., 2013).

There have been several efforts to identify both transcription factors and surface markers to follow the process of reprogramming in a stepwise manner. Oct4, SSEA1 and alkaline phosphatase (AP) staining are commonly used to determine the acquisition of pluripotency. However, these markers, in particular Oct4, are upregulated early during the emergence of the network and the expression of these markers does not indicate the acquisition of fully reprogrammed state (Buganim et al., 2012). James O'Malley and colleagues have identified two additional surface markers ICAM1 and CD44 that enable the sorting of cell populations during the reprogramming process (O'Malley et al., 2013). In addition to the previously established Nanog-GFP reporter, these markers enable a better characterization of the reprogramming mechanism. They found a series of changes in these markers, that could be used to plot the trajectory of reprogramming to a final iPSC state that was $ICAM1^{+}CD44^{-}$, similar to ESCs. This $ICAM1^{+}CD44^{-}$ population contains both a Nanog positive and a Nanog negative population, reflecting the heterogeneity in iPSC clonal lines. In addition, they followed the reprogramming behavior based on these two markers and found that reprogramming occurs via an $ICAM^{-}CD44^{+}$, $ICAM^{-}CD44^{-}$ and finally to $ICAM^{+}CD44^{-}$ and that gene expression changes could be tracked in these populations. These changes commence with the downregulation of mesenchymal gene expression (e.g. *N*-

cadherin (*Cdh2*), *Snail*, *Slug*, *Zeb1*, and *Zeb2*), transient upregulation of epidermal genes (e.g. *Krt6a*, *Ngfr*, *Evpl*), upregulation of early pluripotency genes (e.g. *Oct4*, *Sall1*, *Sall4*), and culminate with the upregulation of late pluripotency genes (e.g. *Nanog*, *Dppa4*, *Sox2*, *Esrrb*, *Klf2*), see figure 1.6.



B

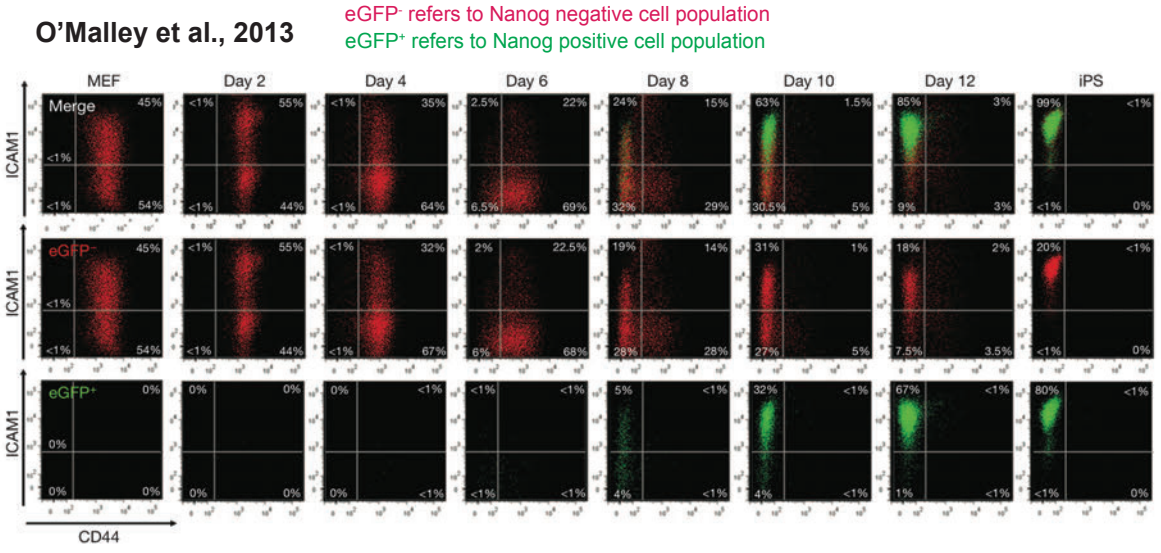


Figure 1.6 Mechanism of murine cellular reprogramming

A) Somatic cells (e.g. mouse fibroblast cells) can be reprogrammed toward induced pluripotent stem cells by overexpression of Yamanaka factor Oct4, Sox2, Klf4 and c-Myc. In the initial phase of reprogramming, Oct4, Klf4 and Sox2 act as pioneer factors and bind to distal regulatory elements on genes with closed chromatin. The four factors also induce a mesenchymal-epithelial transition (MET) that is essential for successful reprograming and produces a pre-iPS state. One of the current views of reprogramming and the establishment of pluripotency involves a conflict between distinct lineage specific transcription factors in the same cell, called "seesaw model". Oct4 is believed to induce mesendodermal (ME) and inhibit ectodermal (ECT) genes, while Sox2 induce ECT and inhibit ME genes. The balance of ECT versus ME gene network requires for pluripotency acquisition and interestingly the Oct4 and Sox2 can be substituted by ME and ECT genes respectievly to induce reprogramming. Some reprogrammed cells later upregulate early pluripotency genes and only a small fraction of cells culminate with the upregulation of late pluripotency genes and become fully reprogrammed cells.

B) James O'Malley and colleagues have identified two additional surface markers ICAM1 and CD44 that enable the sorting of cell populations during the reprogramming process. The fully reprogrammed iPSC cells exhibit ICAM1^{+}CD44^{-}, similar to ESCs. This ICAM1^{+}CD44^{-} population contains both a Nanog positive and a Nanog negative population.

Section 1.5 Roles of Oct4 in the early mouse embryonic development

Neither maternal nor zygotic Oct4 is required for the initiation of zygotic transcription, early cleavage or the initiation of the expression of early embryonic gene expression relevant to pluripotency (e.g. Nanog and Sox2) (Wu et al., 2013). One of the first requirements for Oct4 involves the specification of ICM and TE at the 16-cell stage. At this point, Oct4 expression in the ICM inhibits Cdx2, whereas Cdx2 expression in the TE inhibits Oct4. Both Oct4-mutant ICM cells and ESCs differentiate to trophoblasts (Nichols et al., 1998). Oct4 also regulates the specification of PrEN, both by inducing the transcription of Fgf4 and potentially acting cell-intrinsically within the PrEN cells to potentiate ERK signaling (Frum et al., 2013). Oct4 also reinforces PrEN fate by feed-forward mechanism via partnership with its own downstream target, Sox17 (Aksoy et al., 2013).

Oct4 is expressed throughout the post-implantation epiblast prior to the onset of gastrulation and is successively confined to proximal posterior epiblast during gastrulation (Fuhrmann et al., 1999; Pelton et al., 2002). Deletion of Oct4 from post-implantation epiblast during gastrulation leads to severe phenotypes including craniorachischisis, random heart tube orientation, failed turning, defective somitogenesis and posterior truncation. In addition, Oct4 is required for cell viability and proper cell proliferation within the primitive streak (DeVeale et al., 2013). Moreover, based on studies in *Xenopus* and EpiSCs, Oct4 appears essential to maintain intact adherens junctions in the epiblast (Livigni et al., 2013). Oct4 expression is repressed in the epiblast during the late gastrulation and by E7.5 it emerges again and is restricted to only PGCs (Sabour et al., 2010; Stebler et al., 2004). The expression of Oct4 is maintained in PGCs migrating towards genital ridges (Scholer et al., 1990; Stebler et al., 2004; Yeom et al., 1996). Conditional ablation of Oct4 in the PGCs results in premature apoptosis prior to the completion of migration to genital ridges. Thus,

Oct4 might not be required for PGC specification but is crucial for the survival of PGCs and differentiation (Kehler et al., 2004). The role of Oct4 in PGC survival might be regulated by an oxygen-regulated transcription factor HIF-2α as it has been shown that HIF-2α mutant embryo severely decreased the number of PGCs (Covello, 2006). However, once germ cells are allocated to genital ridges, the deletion of Oct4 from oocytes does not impair the initiation of totipotency/pluripotency after fertilization (Wu et al., 2013).

Section 1.6 Oct4 roles and its regulation in embryonic stem cells

The level of Oct4 in ESCs is essential for both maintaining the pluripotency network and to support efficient differentiation. Loss of Oct4 expression results in ESC differentiation to trophoblast (Nichols et al., 1998). However, iPSC and ESCs expressing lower levels of Oct4 have an enhanced capacity to self-renew and cell lines can be isolated with as low 15% of normal Oct4 expression (Karwacki-Neisius et al., 2013; Radzisheuskaya et al., 2013). Moreover, in experiments where Oct4 mutations are rescued by expression of mutant gain and loss of function proteins expressed from a transgene, the level of Oct4 expressed in different clonal lines supported by these transgenes is remarkable similar and depends on the activity of the Oct4 transgene (Hammachi et al., 2012; Morrison and Brickman, 2006). This observation suggests that ESC growth selects for a precise level of Oct4 expression and activity. These observations are also consistent with the observations that ESCs with elevated Oct4 levels are prone to differentiate to primitive endoderm and mesoderm (Niwa et al., 2000; Radzisheuskaya et al., 2013).

Oct4 is predominantly a transcriptional activator and under standard ESC conditions it is predominantly involved in supporting the pluripotency network (Hammachi et al., 2012). There are currently several ChIP-seq datasets and this has enabled us to compile an annotated list of gene targets from several studies (Livigni et al., 2013; Loh et al., 2006; Matoba et al., 2006; Sharov et al., 2008). Some of these targets are listed in Figure 1.7 These include chromatin modifiers; for example *Jmjd1a* and *Jmjd2c*, encoding histone H3 lysine 9 demethylases and Jarid2 and Mtf2, the components of polycomb repressive complex 2 (PRC2) (Chen et al., 2008b; Kim et al., 2008; Loh et al., 2006). In addition, this Oct4 network is linked to non-coding RNA network. For example, Oct4, synergistically acting with Nanog and Sox2, represses X chromosome inactivation. Oct4 binds to *Tsix* and *Xite*

(noncoding RNA genes) and enhances their transcriptions, in turn leading to repression of non-protein coding gene *Xist*, which coats and inactivates the X chromosome. Hence, Oct4 prevents X chromosome inactivation through indirectly repressing *Xist* (Baker, 2009; Navarro et al., 2008). Oct4 also induces the expression of large intergenic noncoding RNAs (lincRNAs, e.g. *lincRoR*) and miRNA network regulating the maintenance and establishment of pluripotency and cell cycle (Greer Card et al., 2008; Lichner et al., 2011; Loewer et al., 2010; Wang et al., 2013a). Part of this non-coding RNA network (e.g. miR-296, miR470 and miR134 in mouse; Lin28 and miR-145 in human) is involved in Oct4 autoregulation as they post-transcriptionally regulate Oct4 expression through direct binding to the Oct4 mRNA (Qiu et al., 2010; Tay et al., 2008; Xu et al., 2009).

There have also been some ChIP-seq studies done on Oct4 targets in mouse EpiSC or transient epiblast like cells (EpiLCs) and human ESCs. These targets in human ESCs include *OCT4*, *SOX2*, *NANOG*, *LEFTY2/EBAF*, *CDX2*, *HAND1*, *DPPA4*, *FOXO1A*, *CRIPTO/TDGF1*, *ZIC3*, *ESX1l*, *HOXB1*, *MEIS1*, *PAX6*, *LHX5*, *LBX1*, *MYF5* etc. Some of these genes are important for differentiation into three germ layers (Boyer et al., 2005). Some of this gene list is also found specifically in EpiLCs but not in the naïve ESCs, suggesting human ESCs are more similar to mouse EpiSCs than the naïve ESCs. In addition, Oct4 occupancy are substantial around the genes associated with post-implantation epiblast e.g. Fgf5, Oct6 and Wnt8, in mouse EpiLCs. Interestingly, most of the occupancy sites of Oct4 are distal and far from the transcriptional start site (TSS), suggesting Oct4 acts on the enhancers of these genes to regulate their expressions (Buecker et al., 2014).

Oct4 protein has been shown to be associated with a number of other pluripotency regulators and that also recognise overlapping sets of targets. However, the Oct4 protein complex also includes a number of molecules associated with other cellular processes. While the significance of these has yet to be demonstrated, the Oct4 core protein interactome

includes transcriptional coactivators and repressors, cell cycle associated proteins, metabolic processing proteins, ribonucleoprotein complexes, RNA and factors involving protein transport and localization, signal transduction mediators, and translational regulators (Campbell et al., 2007). Recently, it was also shown that Oct4 forms a complex with cyclinB-Cdk1, which inhibits Cdk1 activation. Oct4-mediated Cdk1 inhibition ensures normal G2 progression and prevents premature mitotic entry (Zhao et al., 2014).

Oct4 expression is regulated by a TATA less minimal promoter driven by two upstream regulatory elements: proximal enhancer (PE) and distal enhancer (DE). These enhancers are essential drivers of Oct4 expression at distinct embryonic stages. The Oct4 proximal enhancer plays a role in the epiblast and EpiSC, while the distal enhancer regulates Oct4 expression in the morula, the ICM, ESCs and PGCs (Okazawa et al., 1991; Yeom et al., 1996). The distal enhancer of Oct4 is bound by many transcription factors already associated with the pluripotency network, including Oct4 itself, Sox2, Nanog, Sall4, Tcf3, Smad1, Stat3, Esrrb, Klf4, Klf2, Klf5, E2f1, n-Myc, and Zfx (Chen et al., 2008a; Chew et al., 2005; Ng and Surani, 2011). Those regulatory elements are hypomethylated in ESCs and highly methylated in trophoblast and other differentiated cell types. Given its ability to reprogram somatic cells, it is essential that it remains off in differentiated cell types and it is therefore usually found sequestered away in heterochromatin (Deshpande et al., 2008; Feldman et al., 2006; Li et al., 2007a).

Oct4 is also post-translationally modified, although the significance of these modifications has yet to be shown. Post–translational regulation of Oct4 protein is mediated by phosphorylation, the addition of monosaccharide O-linked β-N-acetylglucosamine (O-GlcNAc), ubiquitination and sumoylation (Jang et al., 2012; Saxe et al., 2009; Wei et al., 2007; Xu et al., 2004). The detail of Oct4 regulation is summarised in figure 1.7.

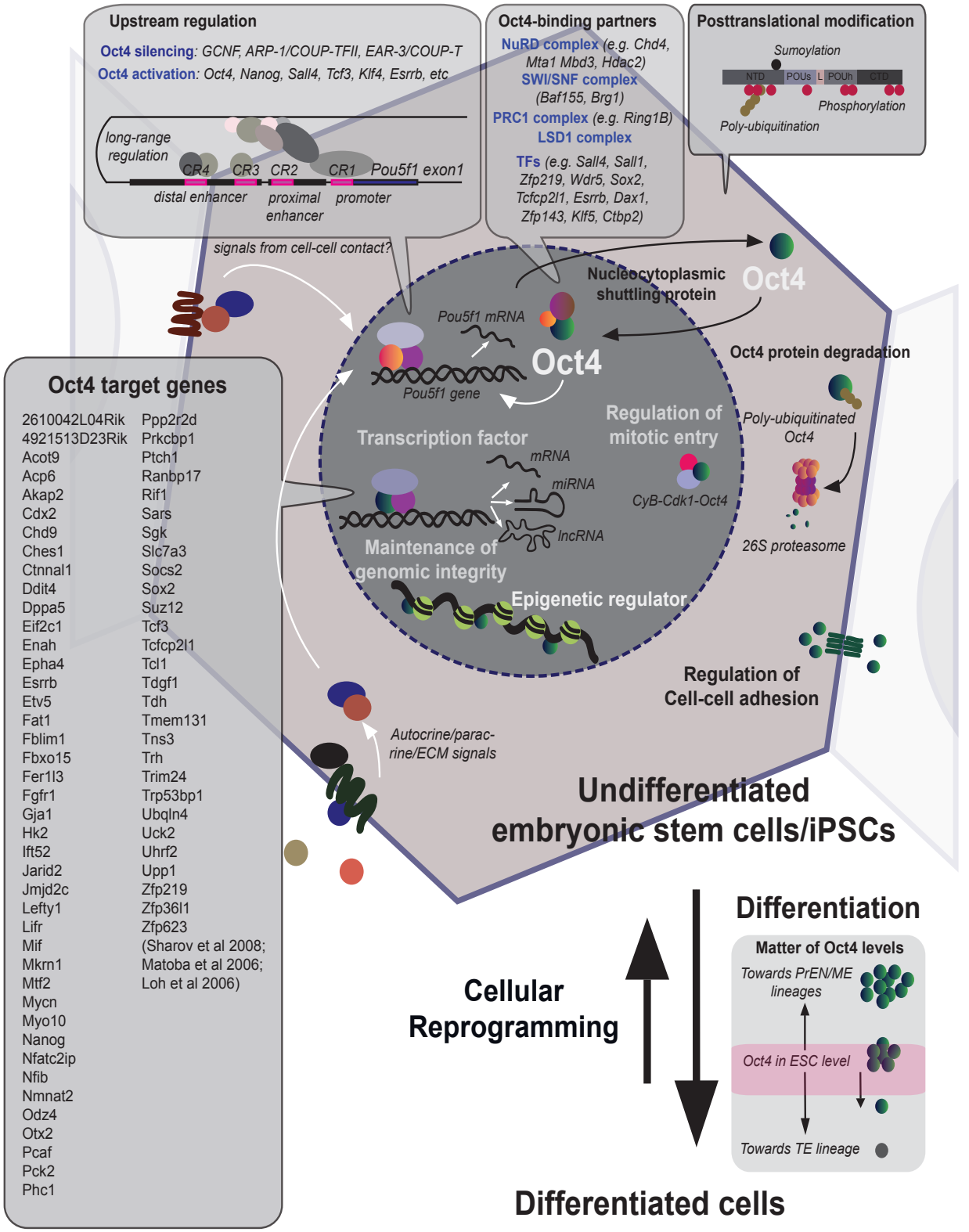


Figure 1.7 Transcriptional network of Oct4 and regulation of Oct4
Oct4 target genes are obtained from Sharov et al 2008, which are found in both the studies of Matoba et al 2006 and Loh et al 2006. The detail of figure is described in section 1.6

Section 1.7 Introduction to vertebrate evolution: relevance to gene regulatory network, pluripotency and germ cell specification

1.7.1 Introduction to vertebrate evolution (Figure 1.8)

All chordates possess shared characters that include a hollow neural tube dorsal to a notochord, pharyngeal gill slits, an endostyle, and a post-anal tail. While these features may not be ubitquitous, they are all present at some point in the organisms' life cycle. The presence of these characteristics is requisite for the membership of the groups. Animals in phylum Chordata include tunicate (salps, sea squirts), cephalochordate (lancelets or amphioxus) and vertebrates. Vertebrates have also evolved a complicated gene regulation that enables them to produce complex and novel morphologies including neural crest and their derivatives, neurogenic placodes, elaborated segmented brain, and endoskeleton (Shimeld and Holland, 2000). The phylum Vertebrata includes several evolutionary clades that mostly diverged by the Precambrian and Paleozoic Eras (600-360 Mya), see figure 1.8 (Hedges and Kumar, 2009). The vertebrates can be broadly divided into (1) the organisms lacking a hinged jaw, termed as agnathans (cyclostomes) including lamprey and hagfish, and (2) the organisms with hinged jaw, termed as gnathostomes including cartilaginous fish (chondrichthyes), and bony vertebrates (osteichthyes). Below are listed the different vertebrate evolutionary clades:

(1) Cyclostomota: jawless fish (agnathans) including lamprey and hagfish,

(2) Gnathostomata:

(2.1) Chondrichthyes: cartilaginous fish including sharks, rays and chimaeras,

(2.2) Osteichthyes:

(2.2.1) **Actinopterygii**: ray-finned fish including bichirs, sturgeons, paddlefish, gars, bowfins, and teleost fish

(2.2.2) Sarcopterygii

(2.2.2a) **Actinistia**: coelacanths

(2.2.2b) **Dipnoan**: lungfish

(2.2.2c) Tetrapoda

(a) **Lissamphibia**: frogs, toads, salamanders, and caecilians

(b) **Amniota**: mammals, tuataras, squamates (lizards, snakes, and amphisbaenians), turtles, crocodilians, and birds.

Here I summarise the detail of each clade that are relevant to this thesis and the classification and timescale are based on the review of (Hedges and Kumar, 2009).

(1) CYCLOSTOMATA: jawless vertebrate

Cyclostomata comprises the two extant jawless fish orders: (1) Myxinidae: hagfish and (2) Petromyzonidae: lampreys, representing the most basal lineage of vertebrates. Both lampreys and hagfish belong to this group because the presence of a jawless mouth armed with retractable horny teeth and the lack of numerous key traits of gnathostomes, including the paired fins/limbs and a mineralised skeleton. Unlike lamprey, the adult form of hagfish lack some phenotypic traits of vertebrates including vertebrae, heart innervation and eye lens, suggesting jawed vertebrates are more closely related to lampreys than to hagfish. However, recent studies on the craniofacial development of hagfish and lamprey embryos

confirms that there are pan-cyclostome patterns of development that are not shared by jawed vertebrates (Oisi et al., 2013). This embryological data combined with other morphological and molecular evidence strongly support the monophyly of these cyclostome's orders and suggest that the deviated traits of hagfish are just secondary losses in this lineage (Oisi et al., 2013; Shimeld and Donoghue, 2012).

Lamprey and hagfish have been studied in order to understand the origins of vertebrate patterning in evolution and in particular the origin of gnathostomes. The developmental biology of lampreys has been studied in two lamprey species including sea lamprey and Japanese lamprey, which account for the majority of lamprey studies. In addition, both have complete genome assemblies. Hagfish eggs are more difficult to find and until now there has been no available hagfish genome assembly. As a result, the embryo studies in this basal group of vertebrate are limited to mostly lamprey (Shimeld and Donoghue, 2012). The genome assemblies of lampreys reveal some interestingly uncommon events called programmed genome rearrangement (PGR). PGR is the irreversible process of elimination of portions of chromosomes (chromosome diminution) or the loss of entire chromosomes (chromosome elimination) during embryonic development. PGR events happen only in somatic cells; thus the somatic genome is different from the germline genome in species that have PGR. In lampreys, PGR occurs at mid-blastula stage (MBT), which is the beginning of zygotic transcription. Hence, cells in early cleavage and blastula stages are the only somatic cells that have the same genome to that of the germline. This PGR is also found in hagfish, leading to the idea that this event is conserved in cyclostomes (Sémon et al., 2012).

(2) GNATHOSTOMES: jawed vertebrates

Features distinguishing gnathostomes from cyclostomes include (1) the presence of pair appendages (pelvin fins and pectoral fins in fish, limbs in terapods) and (2) extensive neural crest-derived and mineralised dermal skeleton, and (3) immunoglobin-based adaptive immunity (Shimeld and Donoghue, 2012; Venkatesh et al., 2014). Gnathostomes are comprised of two major clades: Chondrichthyes (cartilaginous fish) and Osteichthyes (bony vertebrates).

(2.1) CHONDRICHTHYES: cartilaginous fish

Extant Chondrichthyes or cartilaginous fish comprises two subclasses; (1) Holocephalii: ratfish, rabbitfish, and elephantfish and (2) Elasmobranchii: sharks, rays, skates, sawfish, and guitarfish. Unique characteristics of cartilaginous fish that differ from Osteichthyes are (a) the possession of a skeleton of cartilage, (b) internal fertilization via modified male pelvic fins (claspers), (c) possession of placoid (tooth like) scales, and (d) heterocercal tail fin in many lineages. Based on evidence from morphological and molecular analysis, the monophyly of Holocephalii and Elasmobranchii is well supported (Arnason et al., 2001; Mallatt and Winchell, 2007).

(2.2) OSTEICHTHYES: Bony vertebrate

After splitting from the chondrichthyan lineage, the osteichthyan ancestor evolved to have a highly complex process of endochondral ossification. This is likely the result of a tandem gene duplication of genes in the secretory calcium binding phosphoprotein (SCPP) family (Venkatesh et al., 2014). Osteichthyes are comprised of five clades: Actinopterygii,

Actinistia, Dipnoi, Lissamphibia, and Amniota. The first three are bony fish and the latter two are tetrapods.

(2.2.1) Actinopterygii

Extant Actinopterygii or ray-finned fish diversified in the lower Denovian (416-397 Mya) and consist of five major clades that had diversified by the end of the Carboniferous (300 Mya). These five clades include: (1) Polypteriformes: bichirs, (2) Acipenseriformes: sturgeons and paddlefish, (3) Lepisosteiformes: gars, (4) Amiiformes: bowfin, (5) Teleostei: zebrafish, medaka, puffer. Based on molecular and morphological data, the relationship between these five fish clades is controversial, in particular whether (1), (2), (3), and (4) form an ancient fish clade or the monophyletic Holostei of gars and bowfin (Broughton et al., 2013; Hurley et al., 2007).

Teleostei are the most species rich and diversified group of all vertebrates. Teleostei is typically grouped together with Lepisosteiformes and Amiiformes to form the subclass Neopterygii. There are four subdivisions of extant teleosts: (1) Osteoglossomorpha (e.g. mooneyes and bonytongues) (2) Elopomorpha (e.g. eels, tarpons and bonefish) (3) Otacephala (e.g. ostariophysan and clupeomorph teleosts) (4) Euteleostei, the remaining teleosts (e.g. Argentiniformes, Osmeriformes, Salmoniformes and Neoteleostei) (Hedges and Kumar, 2009).

(2.2.2) SARCOPTERYGII

(2.2.2a) Actinistia

Actinistia is a subclass of mostly lobe-finned fish identified in the fossil record. There are only two extant species of this subclass including West Indian Ocean coelacanth (*Latimeria chalumnae*) and the Indonesian coelacanth (*Latimeria menadoensis*). These fish are closer relatives of lungfish and tetrapods than actinopterygians (Amemiya et al., 2013; 2014). It has been held for long time that the most exciting feature of this fish is its prehistoric appearance, which resembles fossils of their ancestors that date back at least to the early Denovian period around 400 Mya (Johanson et al., 2006). This leads to the idea that coelacanth is a very slowly evolving vertebrate lineage and is commonly mentioned as a potential "living fossil." However, the concept of "living fossil", "basal lineage" and "primitive extant species" has been challenged in the field of evolution. Didier Casane and Patrick Laurenti recently reviewed numerous coelacanth studies and concluded that coelacanth should not be mentioned as a living fossil, based on some growing controversial points e.g. low intra-specific molecular diversity, low substitution rates and morphological stability of coelacanth (Casane and Laurenti, 2013). The studies of Chris Amemiya and colleagues provide some insight into how coelacanth protein coding genes might evolve. They examined 251 protein coding genes of coelacanth compared to their orthologs in 22 vertebrates including cartilaginous fish, teleosts, amphibians, birds, reptiles and mammals, and found that coelacanth has a relatively slow rate of protein evolution, compared to other vertebrates (Amemiya et al., 2013). This leads to the idea that coelacanth might be a good reference point for studying the evolution of protein structure. In addition, genome-wide analysis of coelacanth (*Latimeria chalumnae*) provides insight into the genes and regulatory elements that changed during the transition from water to land in the tetrapod ancestor. Coelacanth gene classes absent in tetrapods include fin development, otolith, ear

development, kidney development, trunk and tail development. However, a significant number of homeobox genes (e.g. LIM, POU, TALE, ZF, HNF) responsible for the body plan show only minor variation when coelacanth is compared to ray-finned fish and tetrapods, but the number of these genes is approximately doubled when compared to amphioxus. Thus, during this transition, genes unnecessary for land survival were deleted, while tetrapods retained the key early developmental programs (Amemiya et al., 2013).

(2.2.2b) Dipnoi

Lungfish are living representatives of the Subclass Dipnoi, which belongs to one of extant Sarcopterygii, along with tetrapods and coelacanths. Lungfish are further subdivided in to three families: (1) Lepidosirenidae: South American lungfish, (2) Protoperidae: African lungfish, and (3) Ceratodontidae: Australian lungfish.

(2.2.2c) Tetrapods

Tetrapods are composed of two major groups: amphibians (frogs, axolotl and caecilians) and amniote (reptiles, birds and mammals). Non-amniote tetrapods lay eggs into aquatic environments, similar to teleost fish. Their early embryonic body plans develop quickly after fertilization and the embryo's nutritional source is from intracellular yolk. The successful transition to a completely terrestrial environment in amniotes is part of a revolution in embryonic development that involved the development of extraembryonic tissues and the presence of an egg shell (reptiles, birds and monotremes). In particular, hypoblast/primitive endoderm is one of the amniote innovations although some of the gene regulatory network specifying PrEN is found in deep anterior endoderm of frog embryos and yolk syncytial layer (YSL) of fish embryos (Stern and Downs, 2012).

(a) Lissamphibia

Amphibia are subdivided into three subclasses: Labyrinthodontia, Lepospondyli and Lissamphibia. Only the latter subclass comprises all modern amphibians. Lissamphibians can be further divided into three distinct orders: (1) Caudata/Urodela: salamanders and newts, (2) Anura: frogs and toads, and (3) Gymnophiona/Apoda: caecilians. Based on phylogenetic analyses, Caudata and Anura are grouped in the clade Batrachia, which is monophyletic (Hedges and Maxson, 1993; Laurin and Reisz, 1997).

(b) Amniota

The Amniota comprises six terminal taxa: (1) Testudines: turtles, (2) Sphenodontia: tuatara, (3) Squamata: lizards and snakes, (4) Crocodylia: alligators and crocodiles, (5) Aves: birds, and (6) Mammalia: mammals. The relationships between these taxa are not entirely clear, several conflicting lines of evidence influence the nature of Amniota phylogeny. In particular Testudines complicate things. Based on early examinations of the mammalian and reptile skull, these taxa were classified based on holes or openings near the temples, termed temporal fenestra. The skull of turtles appeared to lack these fenestra (animals with this trait were referred to as Anapsids) and as a result were classed as a more primitive taxa. Other species appeared to have evolved to have one temporal fenestra in Synapsids (or theropsids, mammals) and two fenestra in Diapsids (crocodiles, lizards, snakes, tuatara and birds). Thus, this morphology of skull had originally placed turtle as paraphyletic group to mammals and birds (Gauthier et al., 1988; Lee, 1993; Lyson et al., 2010). However, there is now growing evidence based on in-depth molecular data that has prompted a re-evaluation of classification based on the skull morphology of reptiles. This alternative view places turtles closer to birds and crocodiles and suggests that turtles underwent a secondary loss of skull fenestration or a reversal to an ancestral condition in turtles (Chiari et al., 2012; Crawford et al., 2012; Wang

et al., 2013b). The study of Hugall et al., 2007 based on phylogenetic analysis of long nuclear gene RAG-1 provides additional evidence to place turtles closer to a monophyletic Archosauria (groups of Aves and Crocodylia, including extinct dinosaurs) (Hugall et al., 2007). Based on a combination of fossil calibration together with molecular data, the molecular time estimates predict that turtles diverged from archosauria in the early Triassic around 230 Mya and crocodylia later split from birds around 220 Mya (Kumar and Hedges, 1998; Paton et al., 2002). Squamates and tuatara diverged earlier around 270 Mya (Hugall et al., 2007).

The mammalian ancestor probably diverged from reptiles some 320-350 Mya (Carboniferous period) (Blair and Hedges, 2005; Pereira, 2006). Mammals would have evolved alongside dinosaurs during the Mesozoic era and continued to evolve following the mass extinction that wiped out the dinosaurs in the Cretaceous-Paleocene mass extinction some 65 Mya, that opened new ecological opportunities. Mammal species radiation has expanded ever since, making this lineage one of most successful vertebrates showing great eco-morphological specialization (Luo, 2007). Mammals consist of three major clades: Monotremes (Prototherian mammals, e.g. platypus), Marsupials (e.g. tammar wallaby, Tasmanian devil, koala, opossums, wombats), and Eutherian (e.g. mouse, human). Marsupials and Eutherians together form a clade called Therian. All mammals have the following features: the presence of homeothermy, lactation and hair. The key characteristics of monotremes are the presence of venom, electroreception, meroblastic cleavage, oviparity and a unique reproductive system with combined reptile/bird and therian features, while those of therians are the presence of holoblastic cleavage, placentation, viviparity, testicular descent and trophectoderm (Frankenberg et al., 2013; Warren et al., 2008). Key features of Marsupials are the presence of pouch, short gestation and prolonged lactation. The placenta of marsupials does not support much fetal growth as it does in eutherians. Marsupial infants, called joey, are born in fetal state and have to make their own way to mothers pouch to

receive nutrient via a teat. Eutherians have prolonged gestation, in which embryos develop inside their mother's for longer periods of time. The novel traits of the eutherian lineage include the emergence of novel structures in pre-implantation development including an inner cell mass (Warren et al., 2008) and probably better supportive trophectoderm. The combination of these novel mechanisms might have facilitated the extremely successful radiation of eutherian lineages on Earth.

Living eutherians or placental mammals consist of four major groups of orders: Afrotheria, Xenathra, Euarchontoglires, and Laurasiatheria. The former two are Gondwanan origin and form a clade called Atlantogenata. The latter two are Laurasian origin and form a clade called Boreoeutheria. Fossil records reveal that eutherian mammals originated in the early Cretaceous around 125 Mya (David Archibald, 2003; Ji et al., 2002). Four major clades of eutherians form a monophyletic group and each clade contains subdivisions into several orders as follows:

(a) Afrotheria can be further subdivided into 6 orders: Macroscelidea (e.g. cape elephant shrew), Afrosoricida (e.g. tenrec, golden moles), Tubulidentata (e.g. aardvark), Proboscidea (e.g. elephant), Hyracoidea (e.g. hyrax or shrewmouse or dassy), and Sirenia (e.g. manatee or sea cow).

(b) Xenathra consists of extant placental mammals including anteater, tree sloths, and armadillos. Based on divergence time estimates, it has been found Xenathra have diverged from Afrotheria around 100 Mya which is coincident with the separation of Africa and the South American continent (David Archibald, 2003; Murphy et al., 2007). It is still debated whether Afrotheria or Xenathra is the basal placental lineage. Many traits of Xenathra are not found in other eutherians e.g. extra articulations in vertebral joints, ischiosacral fusion, and internal testicles.

(c) Euarchontoglires consists of five orders: Rodentia (e.g. mouse, deermouse, jerboa, hamster, squirrel, mole rat), Lagomorpha (e.g. rabbit, pika), Primates (e.g. human, monkey), Dermoptera (e.g. colugos or flying lemur), and Scandentia (e.g. treeshrew).

(d) Laurasiatheria consists of six orders: Carnivora (e.g. ferret, cat, tiger, dog, walrus, seal), Pholidota (e.g. pangolin), Perissodactyla (e.g. horse, rhinoceros), Chiroptera (e.g. bats), Cetartiodactyla, and Eulipotyphla. Cetartiodactyla can be further divided into two clades: Artiodactyla (e.g. antelope, alpaca, sheep, pig, buffalo, goat, cow, wild yak) and Cetacea (e.g. whale, dolphin). Eulipotyphla is composed of four further families: Solenodontidae (e.g solenodons), Erinaceidae (e.g. hedgehogs, gymnures), Talpidae (e.g. desmans, moles, shrew moles) and Soricidae (e.g. true shrews). Both Talpidae and Soricidae form a monophyletic clade called Soricomorpha (Roca et al., 2004).

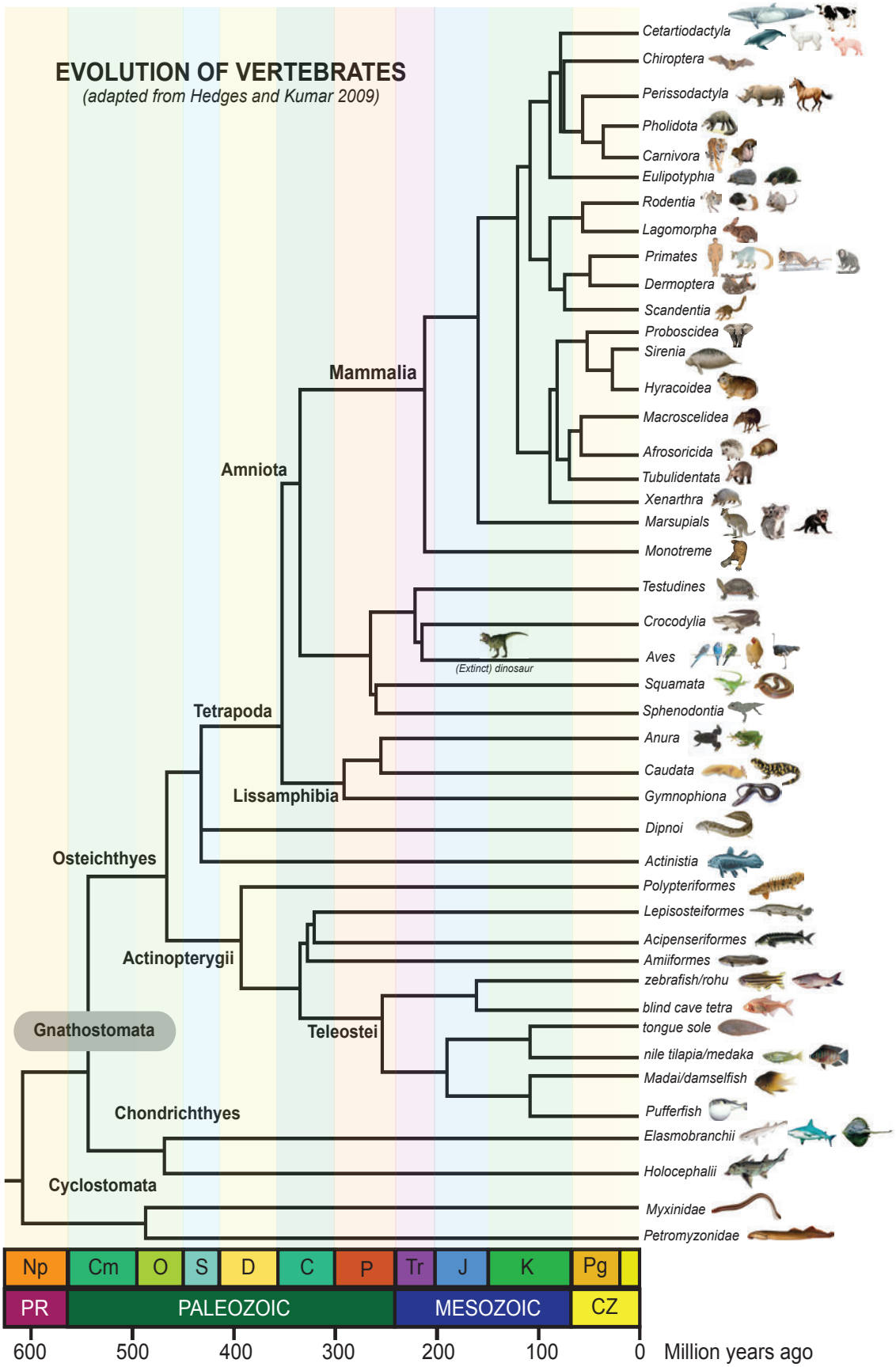


Figure 1.8 Evolution of vertebrates The figure is adapted from Hedges and Kamar 2009. The numbers in each node are refered to divergence time in table 3.1 Abbreviations: C (Carboniferous), Cm (Cambrian), CZ (Cenozoic), D (Devonian), J (Jurassic), K (Cretaceous), Np (Neoproterozoic), O (Ordovician), P (Permian), Pg (Paleogene), PR (Proterozoic), S (Silurian), and Tr (Triassic).

1.7.2 Origin of vertebrates

The origin of vertebrates has been associated with elemental embryological innovation and genome duplication (Donoghue and Keating, 2014). According to recent models, two ancient whole genome duplications (WGD) are responsible for the expansion of the regulatory gene repertoire in vertebrates (e.g. HOX cluster, TGF-β signaling, FGF signaling, insulin receptors, nuclear receptors, neural crest genes etc). In other areas of evolution WGDs have also been suggested to drive the emergence of rich transcription factor repertoires in some plants, fungi and protozoa (de Mendoza et al., 2013). The increase in gene complexity driven by WGDs facilitates species diversification and faster adaptation to novel conditions, increased biological complexity and provides an origin for evolutionary novelties (Van de Peer et al., 2009). Following WGD, the fate of individual genes is not always clear. According to Ohno's classical view, one member of a pair of duplicated genes sustains the original function while its paralog loses function (non-functionalization) or acquires new functions (neo-functionalization). In the DDC (Duplication, Degeneration, Complementation) model, a third possible outcome is suggested, that the ancestral structural and regulatory subfunctions of a parental single-copy gene are partitioned into two duplicate genes (subfunctionalization), leading to the subsequent preservation of both paralogs (Cañestro et al., 2007).

Based on the identification of genomic-wide sets of putative regulatory regions for five vertebrates (human, mouse, cow, stickleback and medaka), three broad trends of regulatory innovation were found that provide a stepwise account of vertebrate evolution. The first period involved the innovation of novel regulatory networks, interfacing and modifying the regulatory elements of transcription factors and key developmental genes and this likely occurred during the evolution of early vertebrate ancestors and continued until the

divergence of the mammal lineage from birds and reptiles (around 300 MYA). A second stage began between 300-100 MYA and involved perturbations of receptor and extracellular signaling interactions to gain novel signaling networks, although the changes to transcriptional regulation had begun to decline during this period. The third period is found in placental mammals, when regulatory innovation for genes implicated in post-translational protein modification increases while the first two trends in innovation (signaling and transcription) decrease to the background levels (Lowe et al., 2011).

1.7.3 Evolution of core pluripotency in vertebrates

Oct4, Sox2, and Nanog (OSN) and the other transcription factors regulating pluripotency have evolved based on their roles in development. Thus OSN have different roles in different compartments of the vertebrate embryo, following the principles of "Mosaic pleiotropy" and "Heterotopy". Mosaic pleiotropy means that the same proteins contribute to different developmental processes and body structures, while Heterotopy means that changes in spatial regulation are associated with morphological divergence. Conserved domains such as the HMG-box (in Sox2) and POU domain (in Oct4) are found across metazoans, but thought to have originated deep in unicellular eukaryotic evolution (de Mendoza et al., 2013). The interaction of POU-SOX proteins is also ancient and a part of ancestral genetic complexity involved in bilaterian development (e.g. *Hydractinia*, planaria, fruit flies and vertebrates). In *Hydractinia*, its POU protein is called Polynem (Pln) and it functions as a block to differentiation in the i-cells, the progenitor population capable of differentiation into somatic and germ cells. Interestingly, Pln has the capacity to re-establish a "stem cell-like," state when reintroduced into differentiated epithelial cells (Millane et al., 2011; Plickert et al., 2012). This could reflect the conserved roles of POU proteins in the maintenance and establishment of progenitor cell phenotypes in the animal kingdom. In this thesis, evolution of POU proteins, in particular POUV/Oct4 is described in more detail

(section 1.8, chapter 3 and chapter 4). Moreover, Nanog might also be a vertebrate innovation, originating with the Bsx gene family at early vertebrate evolution. Surprisingly, Cephalochordate Bsx is incapable of replacing mouse Nanog in murine cellular reprogramming, indicating that the gene regulatory network that supports naïve pluripotency might come from developmental innovations that originate at the beginning of vertebrate evolution (Theunissen et al., 2011). Identification of a naïve pluripotency toolkit in jawless vertebrates (e.g. lamprey and hagfish) might help provide an understanding of why the gene regulatory network supporting naïve pluripotency first emerged and why.

1.7.4 The evolution of germ cell specification

Germ cells give rise to both the sperm and oocyte, which enable the transfer of genetic information from one generation to the next and as a result germ cells themselves have significant roles in evolution. In the animal kingdom, a review of 28 taxa (Extavour, 2003) of both invertebrates and vertebrates, indicate that their germ cells have a lot of gene expression in common. For example, two key germ cell determinants, Nanos and Vasa, are conserved in flies, worms, frogs, fish, chicken, and mammals. This suggests that all organisms in the animal kingdom retain the ancestral machinery for germ cell specification. There are two distinct modes for germ cell development: preformation or the predetermined mode and epigenesis or the inductive mode. Predetermined modes of germ cell specification rely on the localization of maternally inherited determinants as germ plasma to drive germ cell specification. Inductive modes for germ cell specification employ integrative signaling from surrounding tissues for the specification of PGCs during gastrulation. While less common, it appears that the inductive mode is ancestral trait, while preformation is a secondary derived trait. There is little evidence of reversion of performation to epigenesis in any of the branches of the animal kingdom (Extavour, 2003). In this thesis, I explore the

activity of POUV proteins in vertebrates and how that activity might be related to germ cell specification; hence, here I summarise the recent findings regarding the mode of germ cell specification observed amongst several vertebrate taxa below:

a) Inductive mode of germ cell specification

Cyclostomata (lampreys and hagfish), Dipnoi (lungfish), Urodeles (axolotl), Testudines (turtles), Monotremes (platypus), Marsupials (tammar wallaby), Eutherians (mouse and human)

b) Predetermined mode of germ cell specification

Actinopterygians (sturgeon, zebrafish, medaka), Anura (frogs), Archosauria (birds and crocodiles)

Germ cell specification in some vertebrate taxa is still controversial, including cartilaginous fish and other reptiles. Andrew Johnson and colleagues (Bachvarova et al., 2009) suggest that the PGCs of snake are observed at the onset of gastrulation in the lower hypoblast layer similar to chicken, thus snake might have a predetermined mode. While lacertoids exhibit localization of their PGCs similar to that observed in turtle, suggesting an inductive mode. But these inferences are based only on the location of PGCs at later stages and implies that future work needs to be done to resolve the different stages of PGC formation to definitively identify modes of germ cell specification in these reptiles.

Section 1.8 Evolution of Oct4 homologs in vertebrates

1.8.1 Identification and the origin of vertebrate Oct4 homologs (Figure 1.9)

Available genome assemblies across several classes of vertebrate species enable a comparison of class V POU proteins in different species, as there are not many members of this category of POU domain protein, it is relatively straight forward to assign orthologous and syntenic relationships. Two types of orthologous families of *POUV* genes have been identified, namely *POU5F1* and *POU5F3* (Frankenberg and Renfree, 2013). Both genes are believed to have originated as a result of a single genome duplication; hence *POU5F1* is paralogous to *POU5F3* and vice versa. The flanking genes of all vertebrate *POU5F3* are *NPDC1* and *FUT7*. In the coelacanth and turtle genomes, their *POU5F1* genes are flanked by the paralog of *NPDC1*, namely *NPDC1L* (*NPDC1-like*), indicating that the original duplication giving rise to *POU5F1* and *POU5F3* was multigenic and presumable occurred sometime in early vertebrate evolution. Frankenberg et al 2013 examined the conservation of *NPDC1* and *NPDC1L* to locate *POUV* genes in the available vertebrate genomes. They found that some vertebrates including coelacanth, turtle, axolotl, monotremes, and marsupials retain both *POU5F1* and *POU5F3* genes. Birds and crocodiles have lost the *POU5F1* gene, indicating the extinction of *POU5F1* might be common in all archosaurians (including dinosaurs). Frogs, but not axolotl, lost *pou5f1*, but in this case, *pou5f3* has undergone further tandem gene duplication, giving rise to three paralogs: *xlpou91* (*pou5f3.1*)*, xlpou25* (*pou5f3.2*) *and xlpou60* (*pou5f3.3*). The absence of *POU5F3* in both the genomes from various lizards and snakes also indicates a single extinction event in a common ancestor of squamate reptiles. During early mammal evolution, *NPDC1L* was deleted and *POU5F1* was instead flanked by *DDX39B*. This pattern persists in monotremes while the therian ancestor underwent further addition of H2 major histocompatibility

complex between *POU5F1* and *DDX39B*. Finally, the *POU5F3* gene was eliminated from the eutherian ancestor.

The presence of both POUV paralogs in the coelacanth genome led to the previous view that *POU5F1* originated from genome duplication in a sarcopterygian ancestor; however, putative chondricthyan *POU5F1* and *POU5F3* were recently identified using the genomic location and conservation of *NPDC1* and *NPDC1L* genes. Partial sequence for both *POUV* genes was found in the little skate genome while only the *POU5F3* gene is found in the elephantfish genome. This suggests that the origin of *POU5F1* occurred at least as early as a gnathostome ancestor. Recent views of vertebrate evolution support this, as the current 2R hypothesis argues that two rounds of whole genome duplication occurred during early vertebrate evolution and this may explain the origin and duplication of *POUV* genes. It is well accepted that urochordates, not cephalochordates, are sister group to vertebrates or closest living relatives to vertebrates (Delsuc et al., 2006). The first round of genome duplication probably occurred after the split between urochordates and vertebrates and the second appears to occur prior to gnathostome radiation (osteichthyan-chondrichthyan split) or after the split between gnathostomes and cyclostomes. However, the accurate timing of second duplication relative to the gnathostome-cyclostome split remains ambiguous (Holland et al., 2008; Kuraku, 2008; Kuraku et al., 2009). It is unclear which of the ancient whole genome duplications was responsible for the birth of *POU5F1* and *POU5F3*. Identification of *POUV*-like genes among jawless vertebrates (e.g. lamprey and hagfish) might resolve this ambiguity and identify the point of origin of this class of transcription factor.

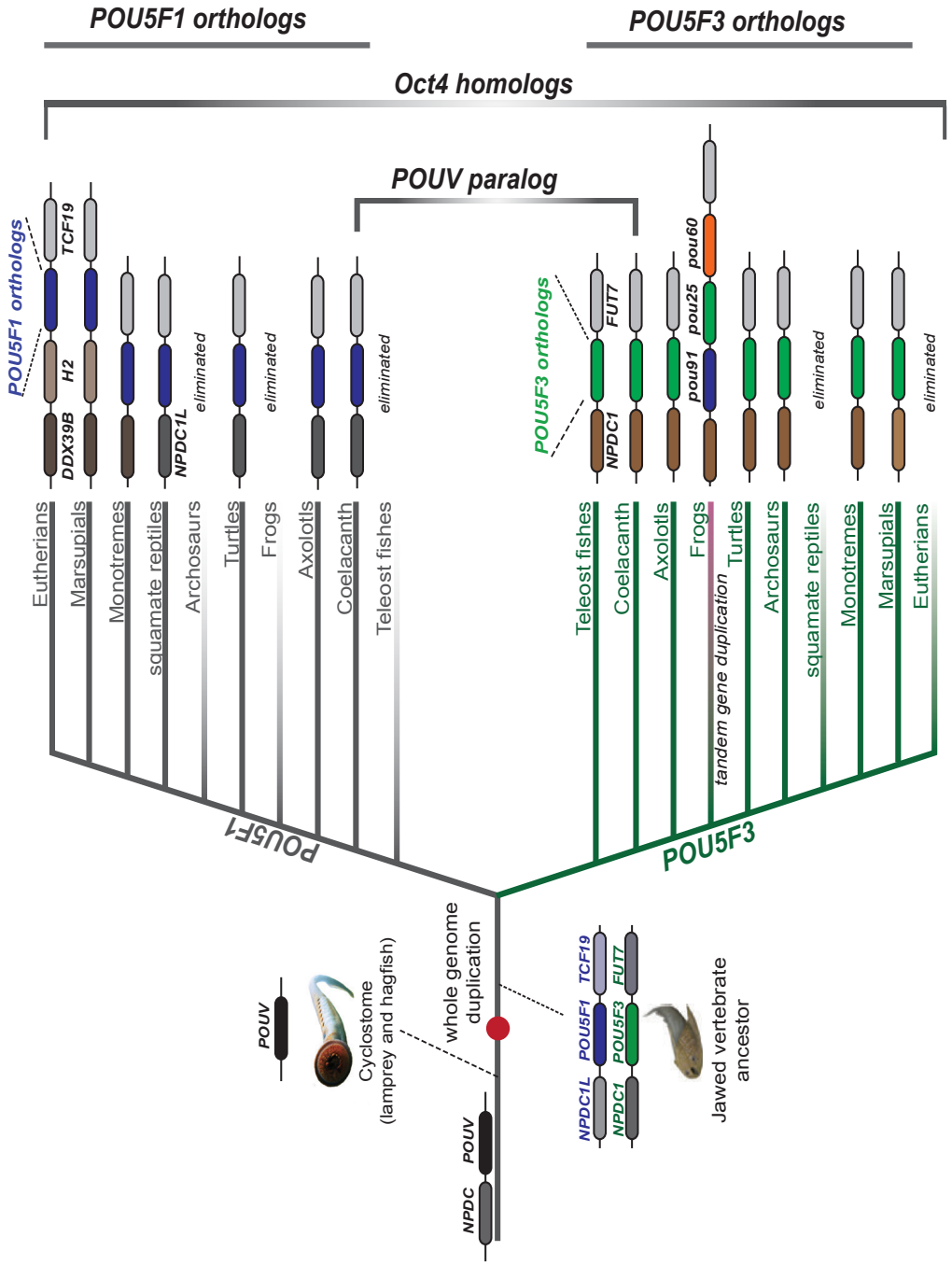


Figure 1.9 The evolution of Oct4 homologs in vertebrates
(described in section 1.8.1) The simplified vertebrate tree depicts the evolution of duplicated *POUV* genes (*POU5F1* and *POU5F3*) from a single *POUV* gene. Some branches of vertebrates show transparent lines, meaning that their *POUV* genes have been eliminated during evolution and marked as "eliminated". The cartoons on the right side of the tree indicate the synteny of *POUV* genes marked in blue color for *POU5F1* and green color for *POU5F3*. In *Xenopus* lineage, tandem duplicated *pou91* (*pou5f3.1*) and *pou60* (*pou5f3.3*) are marked with blue and orange boxes indicating their predicted activities as Oct4-like activity and diverged activity, respectively. The color boxes next to the right and left sides of *POUV* genes mark the flanking genes which are used to confirm the identity of *POUV* homologs, as described in Frankenberg and Renfree 2013.

1.8.2 Comparative early vertebrate embryogenesis and relevance to expression of Oct4 homologs

a) Medaka fish versus zebrafish (Figure 1.10A)

Medaka fish and zebrafish belong to the families Cypriniformes and Beloniformes, which diverged during early teleost fish evolution about 250-300 MYA (Yamanoue et al., 2006). Both fish have similar early embryonic development. Their oocytes have a large amount of yolk and after fertilization the blastomeres divide on the top of the yolk, so called meroblastic cleavage. The large yolk region is formed based on cytoplasmic segregation of yolk and as a result of early meroblastic cleavages producing a single yolk cell with embryonic blastomeres on top. The first embryonic movements begin at the blastula stage. At this point the embryo contains three distinct cell types, an epithelial monolayer on top of the embryo known as the enveloping layer (EVL), deep cells of the blastoderm (DEL) are located underneath these and then a large multinucleated yolk known as the yolk syncytial layer (YSL). Radial intercalation of the DEL leads to epibody movements and production of the embryonic epiblast. At the same time (512 cells) zygotic transcription is initiated. These epiboly movements eventually result in the embryonic cells completely surrounding the yolk cell. The descendants of the DEL cells will make up all the germ layers and the EVL only periderm. Gastrulation occurs at 50% epiboly with the formation of a marginal region referred to as the marginal ring that is the equivalent to the PS. DEL cells involute through this region to make the mesoderm and endoderm, and shortly after the onset of these movements, a thickening at one side of the embryo produces the embryonic shield, the fish equivalent of the node in mouse (Grubb, 2006). Neurulation, and organogenesis of both fish are also well conserved and are quite well conserved with other vertebrates. Germ cell specification in both zebrafish and medaka fish is based on a predetermined mode of

specification (presence of germ plasm), similar to other actinopterygian fish (Ijiri et al., 1996; Iwamatsu, 2004).

Zebrafish Pou5f3 (Spg2/Pou2, here I referred to as Drpou5f3) and medaka fish Pou5f3 (Medaka Oct4, here I referred it as Olpou5f3) are true orthologs of other vertebrate POU5F3 (Frankenberg and Renfree, 2013). Both fish POUV proteins localise to all cells of blastomeres at cleavage stage and are later restricted to the epiblast and marginal ring during gastrulation. Both fish Pou5f3 proteins are later restricted to the posterior tip of the embryonic body (Belting et al., 2001; Sánchez-Sánchez et al., 2009; Takeda et al., 1994). However, there are also some different features of the Pou5f3 expression and function in PGC biology and brain development between zebrafish and medaka fish. In zebrafish, Drpou5f3 is expressed at the midbrain-hindbrain boundary (MHB) where it is required for the establishment and maintenance of progenitors in this region, regulating their competence to respond to Fgf8 (Belting et al., 2001; Reim and Brand, 2002). It is also expressed in the anterior neural plate early during neural induction (Belting et al., 2001). Similar expression patterns of POU5F3 during MHB formation have been shown in *Xenopus* (Xlpou25), axolotl (AmPOU5F3) chick (GgPOU5F3) embryos (Cao, 2004; Lavial et al., 2007; Tapia et al., 2012). Unlike zebrafish, Olpou5f3 is expressed in PGCs, undifferentiated spermatogonia and germ plasm of the oocyte, but not detected at MHB (Belting et al., 2001; Sánchez-Sánchez et al., 2009; Takeda et al., 1994). Zebrafish Pou5f3 has also been shown to be important for the initiation of zygotic transcription and this is unlikely to be conserved in mammals, as Oct4 is not expressed at the 2C stage, the point at which the zygotic genome is activated in mouse.

It's noteworthy that bichir (*Polypterus*) and sturgeon (*Acipenser*) undergo holoblastic cleavage, similar to *Xenopus*. Thus these fish exhibit complete cleavage of the early blastomere segregating yolk and animal cytoplasm at the same time until the first

horizontal cleavage events. These fish lineages diverged from teleost fish prior to a teleost specific WGD (Takeuchi et al., 2009), suggesting that the expression and function of different POUV proteins in these species might resemble that found in frogs and axolotl. In addition, transitional species on the way to teleosts, but still prior to the WGD like gar (*Lepisosterus*) and bowfin (*Amia*) display the transitional forms of cleavage from a holoblastic to meroblastic type. Perhaps these species could shed light on the evolution of the teleost Pou5f3 activity that has lost some functional activity when it is tested for its capacity to support murine ESCs in place of Oct4 (described in detail below).

b) Axolotl versus frog (Figure 1.10B, C)

Most frogs and salamanders undergo holoblastic cleavage, which is the primitive form of cleavage (also found in lamprey and echinoderm) (Takeuchi et al., 2009). The fertilised egg already contains yolk localised to what is referred to as the vegetal hemisphere. Cells in the upper, non-yolky, animal hemisphere will produce the ectoderm and mesoderm. As amphibians possess large amounts of yolk which is an impediment to the cleavage, cleavage occurs asymmetrically (so called unequal holoblastic cleavage), producing a morula with small animal cells and larger yolky vegetal cells. The cells on the future dorsal side are smaller and also tend to cleave faster. By the 128-cell stage the amphibian embryo starts to form a blastocoel in the centre of the animal hemisphere and zygotic transcription initiates around the 10-12th cleavages at the mid-blastula transition. Gastrulation begins as a result of the marginal zone (analogous to the PS in the mouse) involuting through the dorsal blastopore and migrating up and inside the blastocoel. The structure formed at the beginning of gastrulation at the edge of where the dorsal blastopore will form, is the organiser, and can induce secondary axis formation in heterotropic grafting experiments (Hamburger, 1988). This structure is the equivalent to the node in amniotes and shield in fish. As in zebrafish, involution expands from the initial site on the dorsal side to involute the entire marginal zone

or the circumference of the embryo. In *Xenopus*, the lineage fate is determined very early and fates can be mapped at either the 32-cell stage or the blastula (Moody, 1987). As in mouse, the first cells to involute will form the anterior endoderm and prechordal plate, then notochord. The location of cells relative to the site of involution prior to gastrulation will determine their fate (Grubb, 2006; Zorn and Wells, 2009).

Axolotl has two Oct4 homologs and here I refer to them as AmPOU5F1 and AmPOU5F3. Both proteins localise to animal and marginal zones of the gastrulating embryo, which is thought to be an equivalent structure to post-implantation epiblast and PS in the mouse embryo. Interestingly, both homologs are also found in the gonad. Only AmPOU5F3 protein was expressed in the MHB during brain regionalization, similar to zebrafish Pou5f3 (Tapia et al., 2012).

In *Xenopus*, the frog ancestor lost *pou5f1* and retained only *pou5f3*. However, *pou5f3* underwent further tandem gene duplication, giving rise to three paralogs in the same genomic region: *xlpou25* (*pou5f3.2*), *xlpou91* (*pou5f3.1*) and *xlpou60* (*pou5f3.3*).

Xlpou60 is expressed maternally and its transcript is localised to the animal hemisphere in unfertilised oocytes and during early cleavage stages up to MBT. Its expression remains in the animal hemisphere and marginal zone of the embryo until early gastrulation (Morrison and Brickman, 2006).

Xlpou25 is expressed and peaks during gastrulation. Like Xlpou60 it is localised to the animal and marginal zones of the embryo. Later Xlpou25 is found in MHB during brain regionalization and the posterior tip of the neural tube, similar to zebrafish Pou5f3 (Cao, 2004; Hinkley et al., 1992). Xlpou25 is also expressed in the anterior neural plate during the early phase of neural induction.

Xlpou91 is first activated in the animal and marginal zones during mid-blastula stage (onset of zygotic transcription), but not as highly as Xlpou25. Interestingly, only Xlpou91 is strongly expressed in PGCs at neurula stage when PGCs initiate zygotic transcription, suggesting the role of Xlpou91 in PGC biology equivalent to other PGC-specific POUV homologs (Venkatarama et al., 2010). Otherwise its expression is identical to Xlpou25. Either Xlpou25 or Xlpou91 is required to maintain the MHB region and to sustain the expansion of the forebrain region (Morrison and Brickman, 2006).

c) Chicken (Figure 1.10D)

Chicken embryos develop from a disc of cytoplasm sitting on top of a massive yolk. From fertilization to laying takes approximately 20 hours during which a lot of developmental changes take place. Early meroblastic incomplete cleavages produce cells that remain open to the yolk. By the end of cleavage generating a disc of cells, the blastoderm, is generated sitting on top of the yolk. In the centre, this disc is thick with up to six layers of cells and these deeper cells are shed into the subgerminal cavity (that separates the blastoderm from the yolk) leaving a single layer known as the area pellucida, that will make the majority of the embryo. At the edge of the area pellucida, where cells were not shed, is the area opaca and the cells between the two regions of the blastoderm are known as the marginal zone. Cells at the posterior edge of the blastoderm in the region of a structure known as Koller's sickle grown under the blastoderm at the same time as isolated epiblast cells delaminate from the area pellucida, and these two populations will join to make the hypoblast (chick equivalent of VE). There are now two layers, an upper layer, from which most of the embryo will be derived, referred to as epiblast and lower extra-embryonic endoderm layer, the hypoblast. The area between these layers is the blastocoel and they are

joined at their edges by the area opaca. Thus while the geometry of the embryo is different the principles are the same as other vertebrates.

Gastrulation in the chick begins with the formation of the PS just anterior of Koller's sickle. Epiblast cells migrate towards the PS, undergo EMT and migrate along the underside of the epiblast, initially mixing with the hypoblast, but later sorting out, so that the ingressing endoderm forms the embryo and the hypoblast forms the yolk sac. At the anterior end of the PS, a structure known as the Hensen's node forms. It has the same function as the node in mouse and is similar to the organiser in amphibian, or shield in fish. Similar to teleost fish and frogs, chicken germ cell specification is based on maternally deposited germ plasma that dictates the site of germ cell specification.

Chicken has only *POU5F3* gene, also known as *cPOUV* (here I called it *Gallus gallus* POU5F3 or GgPOU5F3). GgPOU5F3 is expressed initially in all blastomeres of the area pellucida and area opaca, as well as in a speckled pattern in the hypoblast. Prior to gastrulation GgPOU5F3 expression is localised to the epiblast region of the area pellucida, During gastrulation, it is expressed around the PS and in the presumptive ectoderm, mesoderm but not endoderm. Similar to other POU5F3 orthologs, GgPOU5F3 is also expressed in neuroectoderm in the MHB region. In addition, GgPOU5F3 is co-expressed alongside germ cell markers VASA and SSEA-1 (germ cell markers) in the PGCs (Lavial et al., 2007).

d) Marsupials versus eutherians (Figure 1.10E, F)

Marsupials (e.g. tammar wallaby, koala, Tasmanian devil) and eutherians (e.g. human, elephant, mouse) belong to the superclass Therian, placental mammals. Therian eggs with no/little yolk undergo holoblastic cleavage while monotremes (e.g. platypus) eggs possessed large amount of yolk undergo meroblastic cleavage similar to other birds and reptiles (Leon Hughes and Hall, 1998). Marsupials and eutherians both have a unique blastula structure called the "blastocyst", which can implant into the mother's uterus through a novel cell type called trophectoderm. Unlike mouse, tammar wallaby has a unilaminar blastocyst that is formed by association of cells into a hemisphere, initially supported by adherence to the zona pellucida. Thus the blastocyst consists of a hemisphere of cells with pluriblast (future embryo) at the centre flanked by trophoblast at the periphery. The pluriblast is defined as the population that will eventually give rise to epiblast (equivalent to 4.5 dpc epiblast in mouse) and hypoblast (equivalent to PrEN). In some species this distinction is apparent morphologically (opossum) whereas others, the unilaminar hemisphere appears homogenous (wallaby) and the pluriblast cells do not become apparent until epiblast formation. The hypoblast will form from the pluriblast and it has been suggested that it may follow the pattern in the chicken, a population of posterior cells adjacent to the trophoblast combine with delaminating islands to form a coherent epithelium beneath the epiblast (Frankenberg et al., 2013). Following hypoblast formation, the remaining disc of pluriblast is referred to as epiblast. All cells in the unilaminar blastocyst express NANOG, POU5F1, POU5F3, CDX2, and SOX2 proteins, suggesting these cells remain totipotent, and equivalent to mouse morula. Like mice, marsupials can be induced into diapause at the unilaminar blastocyst stage when the mother has not finished the lactation period.

During the early blastocyst stage, tammar wallaby POU5F1 is found in all cell lineages. However, tammar POU5F3 is expressed in a more restricted fashion, only in $NANOG^+$ cells committed to epiblast, and not in either the $GATA6^+$ and $CDX2^+$ cells that mark hypoblast and trophectoderm, respectively (Frankenberg et al., 2013). During gastrulation, both tammar POU5F1 and POU5F3 can be found in the epiblast. Interestingly, during gastrulation, POU5F3 protein is localised to the nucleus while POU5F1 localised mainly to the cytoplasm of the epiblast cells. At later stages, only tammar POU5F1 is expressed in the PGCs and adult gonads, similar to mouse Oct4 (Frankenberg et al., 2010). This expression segregation of tammar POUV proteins suggests that tammar POU5F1 activity might be equivalent to the mouse Oct4 activity in the ICM and PGCs, while tammar POU5F3 activity might be equivalent to mouse Oct4 activity in post-implantation epiblast.

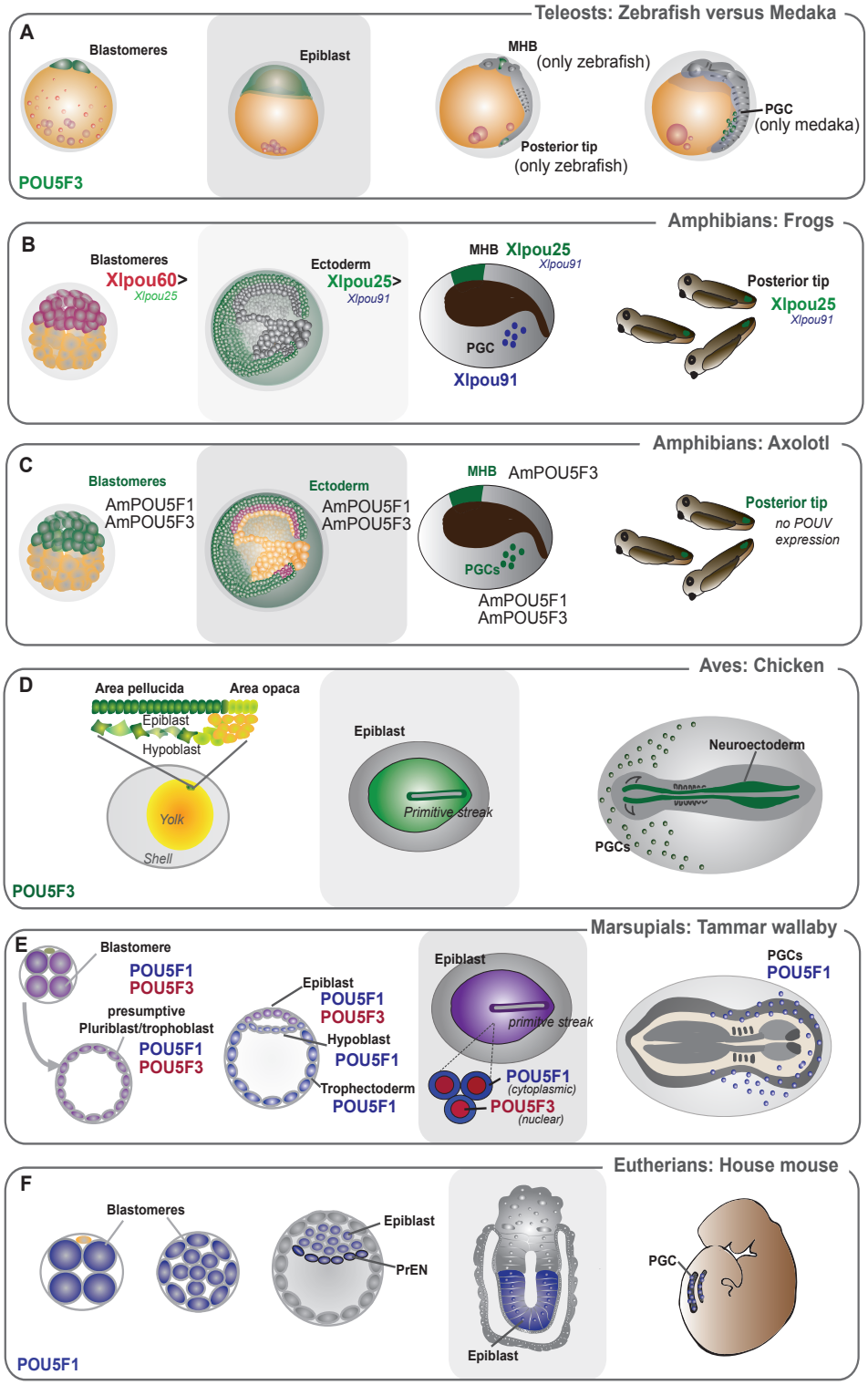


Figure 1.10 Comparative POUV expression among vertebrates A) Zebrafish (*Danio rerio*, Dr) and medaka fish (*Oryzias latipes*, Ol) have only single POUV protein called Drpou5f3 and Olpou5f3, respectively. Both POUV are expressed similarly in blastomeres around animal hemisphere and epiblast cells during epiboly process of gastrulation. Later Olpou5f3 is specifically expressed in the primodial germ cells (PGC) while Drpou5f3 is specifically expressed in the midbrain-hindbrain boundary (MHB) and posterior tip. B) African clawed frog (*Xenopus lavis*, Xl) has three POUV homologues; Xlpou25, Xlpou60 and Xlpou91. Xlpou60 is abundantly expressed from oocyte to cleavage. Later Xlpou25 and Xlpou91 are expressed in the ectoderm (epiblast-equivalent structure) during gastrulation. Xlpou91 is then specifically expressed in PGCs. Mexican axolotl (*Ambystoma mexicanum*, Am) has both POU5F1 and POU5F3, called here AmPOU5F1 and AmPOU5F3. Both are expressed in cleavage, gastrulation and PGCs. AmPOU5F3 is also expressed in MHB. C) Chicken/*Gallus gallus* POUV is a POU5F3 (called here GgPOU5F3). GgPOU5F3 is expressed in the epiblast of area pellucida , and less in area opaca and hypoblast layer. GgPOU5F3 is later expressed in the gastrulation stage epiblast, neuroectoderm and germ cells at later stage of development. D) Tammar wallaby (*Macropus eugenii*, Me) has both POU5F1 and POU5F3. Both are expressed in all blastomeres at cleavage and in presumptive pluriblasts and trophoblasts of the unilaminar blastocyst stage. Specified epiblast at blastocyst stage later expresses only MePOU5F3. After implantation, MePOU5F3 localizes to the nucleus while MePOU5F1 localizes to the cytoplasm of the epiblast. Only MePOU5F1 is later expressed in the PGCs. E) Oct4 or Pou5f1 in mouse (*Mus musculus*) embryo is expressed in all blastomeres at cleavage and morula stages, and later is restricted to the epiblast and primitive endoderm (PrEN) at the late blastocyst stage. During gastrulation, Oct4 is expressed in the epiblast, and later is restricted to only posterior epiblast and PGCs successively.

1.8.3 Conserved network of Oct4 homologs in regulating pluripotency and differentiation

Several studies have shown that POUV proteins across vertebrate evolution have the capacity to support and induce pluripotency in murine ESCs in place of Oct4. We have previously shown that *Xenopus* Pou91 (Pou5f3.1) could replace mouse Oct4 efficiently while zebrafish Pou5f3 (Drpou5f3/Pou2) had no such activity. Other laboratories also examined the ability of other vertebrate POUV proteins, including chick POUV (GgPOU5F3/cPOUV), platypus POU5F1, opossum POU5F3, axolotl POU5F1 and axolotl POU5F3, to rescue mouse Oct4-null ESCs (commonly used ZHBTc4 ESC cell lines). Different POUV proteins show different abilities to rescue Oct4-null phenotypes and in this thesis I examine the quantitative difference between POUV paralogs. It has also been shown that mouse Oct4 can rescue POUV-deleted *Xenopus* embryos, further emphasizing the conservation of POUV activities (Morrison and Brickman 2006).

In addition to its role in maintaining ESC self-renewal, Oct4 is one of the key reprogramming factors. Natalia Tapia and colleagues have investigated the substitution of mOct4 in murine cellular reprogramming with other vertebrate POUVs, including medaka Pou5f3, zebrafish Pou5f3, axolotl POU5F1, axolotl POU5F3, and Xlpou91. These POUVs have the ability to reprogram mouse embryonic fibroblasts into an ESC-like state with varied degrees of success (Tapia et al., 2012). Zebrafish Pou5f3 lacks the capacity to induce reprogramming, fitting well with the rescue assay shown in Morrison and Brickman, 2006 whereas medaka Pou5f3 had some activity, although it exhibited less than 10% of the activity normally demonstrated by Oct4. Interestingly, Xlpou91 was almost as effective as Oct4 (approximately 80% Oct4 activity) at reprogramming human fibroblasts, which is consistent with its ability to support Oct4 null murine ESCs. However, this study did not

point to any logical pattern for the evolution of the capacity of these proteins to reprogram human cells, e.g. both axolotl POU5F3 and POU5F1 were not able to induce iPSCs, but *Xenopus* Pou5f3 (Xlpou91) was. In an attempt to explain the differences in these proteins I examine their activity in both ESCs (chapter 3) and reprogramming (chapter 4) and attempt to try and shed some light on the evolutionary basis for the capacity of some of these proteins to support pluripotency in mammals.

A number of developmental phenotypes observed in different species in response to gain and loss of function of POU5F1 and POU5F3 appear similar. The lack of Oct4 in the post-implantation mouse embryo results in a failure to support continued axis extension (DeVeale et al., 2013) and this may be because of a loss of PS progenitor populations. Similarly, in *Xenopus* gastrulation, partial depletion of all three POU5F3 homologs leads to a failure in axis extension as a result of the immediate early differentiation of marginal zone progenitors to endoderm (Morrison and Brickman, 2006). In addition, zebrafish mutants lacking both maternal and zygotic Pou5f3 proteins (MZspg mutant) and more efficient knock down of *Xenopus* Pou5f3 proteins exhibit gastrulation failures (Lachnit et al., 2008; Lunde et al., 2004; Reim et al., 2004). Thus, the phenotypes of XlpouV deletion and MZspg mutant have some commonalities. Moreover, comparisons between *Xenopus* and mouse POUV target genes suggest that Oct4 and its homologs regulate cell adhesion and this may be linked to blocking differentiation in progenitor cells. Interestingly, some of these conserved Oct4/POUV targets (*Xenopus* Xlim5, Xcad2 and Xsal1) and direct regulators of adhesion, such as E-cadherin, can rescue XlpouV depleted embryos (Livigni et al., 2013).

Oct4 regulatory networks are likely to be conserved. In addition to the regulation of adhesion mentioned above there are two examples of potential conserved networks that are worth introducing; the link between Oct4 and Cdx, and a role for Oct4 homologs in neural development. Oct4 is an essential regulator of early lineage choices in mammalian

development. While these choices involve extra-embryonic development and as such are unique to mammals, how are they linked to conserved networks in other vertebrates? When Oct4 inhibits trophoblast fate in favor of ICM, its acts through inhibiting Cdx2 expression. Interestingly, this regulatory relationship is also found in *Xenopus*, despite the absence of trophoblast. The homolog of Cdx2 in *X.laevis* is Xcad3, which is expressed in posterior neural tube at larval stages. XlPouV depletion in *Xenopus* leads to the expansion of Xcad3 expression (Morrison and Brickman, 2006), indicating that a POUV-CDX network is likely conserved among vertebrates.

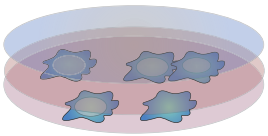
POU5F3 proteins have also been shown to have roles in both forebrain specification and supporting a progenitor population at the MHB in both fish and frog (Morrison and Brickman, 2006; Reim and Brand, 2002). How conserved is this function? Based on expression it appears that POU5F3 homologs are also expressed in these regions in both chick and axolotl embryos (Lavial et al., 2007; Tapia et al., 2012). There is no evidence for mouse Oct4 expression in the MHB region and epiblast depletion of Oct4 via conditional mutagenesis does not appear to create a MHB phenotype (DeVeale et al., 2013) although overexpression of an activator form of Oct4 in ESCs induces neural and MHB gene expression (Hammachi et al., 2012). Hence, while this regulatory loop may still be encoded in the mammalian genome, Oct4 itself could have been replaced by another octamer protein that acts to sustain both forebrain and MHB development.

Section 1.9 Aims of this study

Oct4 homologs are conserved among vertebrates in their activities to maintain and establish pluripotency. In this thesis, I explore the links between a pattern of evolution of these proteins and their functional properties. As there are two clear assays for function, I employ them both and explore different aspects of POUV conservation.

The specific aims of this thesis include:

1. Better define the evolutionary relationships of POUV proteins. In chapter 3, I explore the evolutionary relationships between the genes encoding different POU5F1 and POU5F3 proteins. I attempt to better describe their divergence and evolutionary origins. I also assess the functional activity of these proteins in ESCs and compare their activity to their evolutionary relationships, what is known about their expression pattern and whether they are derived from a species with POU5F1, 3 or both.
2. To better understand the diversification of POU5F3 activity that occurred in *Xenopus* as a result of tandem duplication using a reprogramming assay (chapter 4).



CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2 Material and Methods

Section 2.1 Reagents-Materials, Plasmids, Antibodies, and Primers

2.1.1 Reagents and materials used in this study

1. Dulbecco's Modified Eagle Medium (DMEM)-High Glucose, Invitrogen, 41965-039

2. 100X MEM Non-essential Amino Acids solution (NEAA), Sigma-Aldrich, M7145

3. 200 mM L-Glutamine, Invitrogen, 25030024

4. Fetal Bovine Serum (FBS), Invitrogen, 10270-106

5. Knockout Serum Replacement (KOSR), Invitrogen, 10828-028

6. Leukemia Inhibitory Factor (LIF)-Homemade LIF produced by COS cells (Dora Papp)

7. Glasgow Minimum Essential Medium (GMEM), Sigma-Aldrich, G5154-6X500 ML

8. β-mercaptoethanol, Sigma-Aldrich, M6250

9. Sodium pyruvate, Invitrogen, 11360039

10. Penicillin-Streptomycin, Invitrogen, 15140-122

11. 2.5% Trypsin (10X), Invitrogen, 15090-046, 5 mL of 2.5% Trypsin was dissolved in 500 mL of PBS with 0.186 gram of EDTA and 5 mL of chicken serum to obtain 1X Trypsin-EDTA for ESC/iPSC culture.

12. Dimethyl sulfoxide (DMSO), Merck Millipore, 102952

13. Dulbecco's Phosphate Buffered Saline (PBS), Sigma-Aldrich, D8537-100 ML

14. Lipofectamine ® 2000 transfection reagent, Invitrogen, 11668019

15. Restriction enzymes and buffers, New England Biolab (NEB)

16. T4 ligase, New England Biolab (NEB), M0202S

17. Phusion® High-Fidelity PCR Master Mix with HF Buffer, New England Biolab (NEB), M0531L

18. UltraPure^{TM} Agarose, Invitrogen, 16500-500

19. Tris-acetate EDTA (TAE), Homemade

20. Zymoclean™ Gel DNA Recovery Kit, Zymo Research, D4008

21. QIAquick® PCR Purification Kit, Qiagen, 28104

22. Qiagen Plasmid Mini/Maxi Kit, Qiagen, 12125/12362

23. Illustra™ PlasmidPrep Mini Spin Kit, GE Healthcare, 28-9042-70

24. SyBr ® Safe DNA Gel Stain, Invitrogen, S33102

25. 1 kb Plus DNA Ladder, Invitrogen, 10787018

26. Carbenicillin, GE Healthcare, P31-020

27. Kanamycin, Sigma-Aldrich, K4000-25G

28. LB broth, SUND core facility

29. LB agar, SUND core facility

30. Retro-X™ Concentrator, Clontech, 631453

31. Retro-X™ qRT-PCR Titration Kit, Clontech, 631456

32. Tetracycline Hydrochloride, Sigma-Aldrich, T-7660, Dissolved in 70% ethanol to obtain 10 mg/mL

33. Nuclease-free water, Ambion, AM9937

34. Puromycin dihydrochloride from *Streptomyces alboniger*, Sigma-Aldrich, P8833

35. Triton® X-100, AppliChem, A13388-0500

36. Donkey serum for immunofluorescence, Sigma-Aldrich, D9663-10ML

37. Leukocyte Alkaline Phosphatase Kit, based on napththol AS-BI and fast red violet LB, Sigma-Aldrich, 86R-1KT

38. RNA extraction kit: RNeasy® Mini Kit, Qiagen, 74106

39. Trizol® Reagent, Invitrogen, 15596018

40. DirectPCR (Tail) Lysis Reagent for Genotyping Using Crude Lysates, Viagen, 102-T

41. GeneArt® Seamless PLUS Cloning and Assembly Kit, Invitrogen, A13288

42. Agilent's microarray kit, Agilent Technologies

    42.1 SurePrint G3 Mouse GE 8X60K Kit, G4852A

42.2 LowInput QuickAmp Labeling Kit One-Colour, 5190-2305

42.3 RNA Spike In Kit-One Colour, 5188-5282

42.4 Gene Expression Hybridization Kit, 5188-5242

42.5 Pack 5 Backings 8 arrays per slide, G2534-60014

42.6 Gene Expression Wash Pack, 5188-532

43. DNA Clean & Concentrator™, Zymo Research, D4029

44. RNase-Free DNase set, Qiagen, 79254

45. Hexadimethrine bromide (polybrene), Sigma-Aldrich, H9268-10G

46. SuperScript® III Reverse Transcriptase (200 U/μL), Invitrogen, 18080-044

47. TOPO TA Cloning® Kit for Sequencing, Invitrogen, 45-0030

48. NucleoSpin® RNA Virus, Macherey-Nagel, 740956-50

49. Mitomycin C from *Streptomyces caespitosus*, Sigma-Aldrich, M4287

50. L-Ascorbic acid, Sigma-Aldrich, A4403-100MG, Dissolved in water to obtain 10 mg/mL stock solution

51. Alk5 inhibitor (Alk5i), Tocris, A83-01, Dissolved in DMSO to obtain 10 mM stock solution

52. Accutase® solution, Sigma, A6964-100 mL

53. 25/75/150 cm^2 cell culture flask, Corning Incorporated

54. Cryotube, Thermo Scientific

56. 100 mm X 20 mm style cell culture dish, Corning Incorporated

57. 6, 12, 24, 96 (flat/V bottom) well cell culture plate, Corning Incorporated

58. 25 mL disposable, Pre-sterile, polystyrene reservoir

59. Universal, Thermo Scientific

60.BD Falcon™ 15 mL high clarity polypropylene conical tube, BD Bioscience

61. 5 mL polystyrene round-bottom tube with cell strainer cap, Corning Incorporated

62. Serological pipettes, VWR

63. 15 μ-Slide 8 well, Ibidi, 80826

64. WhatmanTM FP30/0.45 CA-5 Filter unit (0.45 μm), GE Healthcare, 10462100

65. BBD6220 CO_2 incubator, Thermo Scientific

66. GlycoBlue™ Coprecipitant, Applied Biosystems, AM9515

2.1.2 Plasmids used in this study

For chapter 3: Rescue experiment

Expression vectors carrying POUV genes were generated from pPCAG575-SiP (called shortly pCAG in this thesis). The expression of POUV gene is driven by CAG promoter (murine pDR10, Rathjen et al., 1990).

1. **pCAG-3xflag mOct4**, Generated by Fella Hammachi, Source of cDNA: *Mus musculus* (house mouse)/ mOct4 (Genbank: MGI:101893; ENSMUSG00000024406), referred to here as **mouse Oct4, mouse Pou5f1 or mOct4**

2. **pCAG-3xflag Xlpou91**, Generated by Fella Hammachi, Source of cDNA: *Xenopus laevis* (African clawed frog)/Xlpou91 or Pou5f3.1 or Oct91, Genbank: M60077; NP_001081342.1, referred to here as **Xlpou91, Pou5f3.1 or X91**

3. **pCAG-3xflag Xlpou25**, Generated by Fella Hammachi, Source of cDNA: *Xenopus laevis* (African clawed frog)/Xlpou25 or Pou5f3.1 or Oct25, Genbank: ABH07383, referred to here as **Xlpou25, Pou5f3.2 or X25**

4. **pCAG-3xflag MePOU5F1**, Generated in this study, Source of cDNA: *Macropus eugenii* (tammar wallaby)/POU5F1, Genbank: FJ998419, cDNA was obtained from Stephen Frankenberg, referred to here as **tammar POU5F1 or tammar P1**

5. **pCAG-3xflag MePOU5F3**, Generated in this study, Source of cDNA: *Macropus eugenii* (tammar wallaby)/POU5F3, Genbank: FJ998420, cDNA was obtained from Stephen Frankenberg, referred to here as **tammar POU5F3 or tammar P3**

6. **pCAG-3xflag CpPOU5F1**, Generated in this study, Source of cDNA: *Chrysemys picta bellii* (painted turtle)/POU5F1, cDNA sequence was obtained from Stephen Frankenberg and synthesised by Invitrogen, referred to here as **turtle POU5F1 or turtle P1**

7. **pCAG-3xflag CpPOU5F3**, Generated in this study, Source of cDNA: *Chrysemys picta bellii* (painted turtle)/POU5F3, cDNA sequence was obtained from Stephen Frankenberg and synthesised by Invitrogen, referred to here as **turtle POU5F3 or turtle P3**

8. **pCAG-3xflag AmPOU5F1**, Generated in this study, Source of cDNA: *Ambystoma mexicanum* (axolotl; Mexican salamander)/AxOCT4, Genbank: AY542376, cDNA sequence was obtained from Elly Tanaka, referred to here as **axolotl POU5F1 or axolotl P1**

9. **pCAG-3xflag AmPOU5F3**, Generated in this study, Source of cDNA: *Ambystoma mexicanum* (axolotl; Mexican salamander)/AxPOU2, Genbank: KF020689, cDNA sequence was obtained from Elly Tanaka, referred here as **axolotl POU5F3 or axolotl P3**

10. **pCAG-3xflag LcPOU5F1**, Generated in this study, Source of cDNA: *Latimeria chalumna* (coelacanth)/POU5F1, cDNA sequence was obtained from Stephen Frankenberg and synthesised by Invitrogen, referred here as **coelacanth POU5F1 or coelacanth P1**

11. **pCAG-3xflag LcPOU5F3**, Generated in this study, Source of cDNA: *Latimeria chalumnae* (coelacanth)/POU5F3, cDNA sequence was obtained from Stephen Frankenberg and synthesised by Invitrogen, referred here as **coelacanth POU5F3 or coelacanth P3**

For chapter 4: iPSC experiment

Retroviral expression vectors carrying POUV genes were generated from pMXs-gw (Kitamura et al., 2003). POUV coding sequences were firstly subcloned into pENTR-2B2 (Invitrogen). POUV gene in pENTR-2B2 was then subcloned into pMXs-gw vector through Gateway ® cloning LR recombination reaction (Invitrogen).

1. **pMXs-mOct4**, Generated by Shinya Yamanaka laboratory (provider: Addgene), Source of cDNA: *Mus musculus* (house mouse)/ mOct4, Genbank: MGI:101893; ENSMUSG00000024406, referred to here as **mouse Oct4, mouse Pou5f1 or mOct4**

2. **pMXs-Sox2**, Generated by Shinya Yamanaka laboratory (provider: Addgene), Source of cDNA: *Mus musculus* (house mouse)/ Sox2, Genbank: AL606746.1, referred to here as **Sox2 or S**

3. **pMXs-Klf4**, Generated by Shinya Yamanaka laboratory (provider: Addgene), Source of cDNA: *Mus musculus* (house mouse)/ Klf4, Genbank: RP23-322L22.2, referred to here as **Klf4 or K**

4. **pMXs-c-Myc**, Generated by Shinya Yamanaka laboratory (provider: Addgene), Source of cDNA: *Mus musculus* (house mouse)/ c-Myc, Genbank: AU016757, referred to here as **c-Myc or M**

5. **pMXs-dsRED**, Generated by Shinya Yamanaka laboratory (provider: Addgene), Source of cDNA: *Discosoma* sp. (sea anemone)/ dsRED, referred to here as **dsRED or R**

6. **pMXs-3xflag Xlpou91**, Generated by this study, Source of cDNA: *Xenopus laevis* (African clawed frog)/Xlpou91 or Pou5f3.1 or Oct91, Genbank: M60077; NP_001081342.1, referred to here as **Xlpou91, Pou5f3.1 or X91**

7. **pMXs-3xflag Xlpou25**, Generated by this study, Source of cDNA: *Xenopus laevis* (African clawed frog)/Xlpou25 or Pou5f3.1 or Oct25, Genbank: ABH07383, referred to here as **Xlpou25, Pou5f3.2 or X25**

2.1.3 Antibodies used in this study

For flow cytometry analysis and FACS sorting

1. Mouse CD54 (ICAM-1) conjugated with Biotin, eBioscience, 13-0541-81, Dilution 1:200

2. Human/Mouse CD44 conjugated with APC-eFluor® 780, eBioscience, 47-0441-82, Dilution 1:200

3. Alexa Fluor® 647 Streptavidin (used together with ICAM-1), BioLegend, 405237, Dilution 1:1000

4. Mouse CD31 (PECAM-1) conjugated with APC, BD Pharmingen, 551262, Dilution 1:100

5. Mouse SSEA1 conjugated with Alexa Fluor® 647, BD Pharmingen, 560120, Dilution 1:1000

6. Mouse CD324 (E-cadherin) conjugated with eFluor® 660, eBioscience, 50-3249, Dilution 1:400

7. Mouse c-kit (CD117) conjugated with APC, BD Pharmigen, 561074, Dilution 1:500

For immunofluorescence

Primary antibody

1. Mouse Oct3/4 (Host: mouse), Santa Cruz, sc-5279, Dilution 1:200

2. Mouse Sox2 (Host: goat), Santa Cruz, sc-17320, Dilution 1:200

3. Mouse Klf4 (Host: goat), R&D, AF3158, Dilution 1:200

4. Mouse C-myc (Host: rabbit), Abcam, ab32072, Dilution 1:200

5. Flag (Host: mouse), Sigma, F3165, Dilution 1:1000

6. Mouse SSEA1 conjugated with Alexa Fluor® 647, BD Pharmingen, 560120, Dilution 1:50

7. Mouse Nanog (Host: rat), eBioscience/14-5761, Dilution 1:400

8. Mouse p120 catenin (Host: mouse), BD Pharmingen, 610134, 1:250

9. Human Gata6 (Host: Goat), R&D, AF1700, Dilution 1:200

10. Mouse Cdx2 (Host: mouse), BioGenex, MU392A-UC, Dilution 1:100

11. Mouse E-cadherin (Host: Goat), R&D, AF748, Dilution 1:200

Secondary antibody

1. Donkey anti goat Alexa 488 (IgG), Molecular Probes, A121055, Dilution 1:800

2. Donkey anti goat Alexa 568 (IgG), Molecular Probes, A11057, Dilution 1:800

3. Donkey anti goat Alexa 647 (IgG), Molecular Probes, A11057, Dilution 1:800

4. Donkey anti rabbit Alexa 568 (IgG), Molecular Probes, A10042, Dilution 1:800

5. Donkey anti rabbit Alexa 647 (IgG), Molecular Probes, A31573, Dilution 1:800

6. Donkey anti mouse Alex 488 (IgG), Molecular Probes, A21202, Dilution 1:800

7. Donkey anti mouse Alexa 568 (IgG), Molecular Probes, A10037, Dilution 1:800

2.1.4 Primers used in this study

Table 2.1 Primers used in this study

Gene	UPL Probe	Forward	Reverse
Cdh1	18	ATCCTCGCCCTGCTGATT	ACCACCGTTCTCCTCCGTA
Cdh2	67	GGTGGAGGAGAAGAAGACCAG	GGCATCAGGCTCCACAGTAT
Cdx2	34	CACCATCAGGAGGAAAAGTGA	CTGCGGTTCTGAAACCAAAT
Dlx3	18	GCAAGTCGAAAGAGGGATGT	CTCCTTCACTTCCCACGAAA
Esrrb	93	AACTGGGCCAAGCACATC	ATCTCCATCCAGGCACTCTG
Fgf5	95	GCGAAACTTCAGTCTGTACTTCACT	ACCGGTGAAACCAAAGGTG
Gata6	40	GGTCTCTACAGCAAGATGAATGG	TGGCACAGGACAGTCCAAG
IRES-PAC	41	TGGCTCTCCTCAAGCGTATT	CCCCAGATCAGATCCCATAC
Klf4	62	CGGGAAGGGAGAAGACACT	GAGTTCCTCACGCCAACG
Lhx5	55	TGTGCAATAAGCAGCTATCCA	TTGCACACAAACTTGTTCTCG
Nanog	25	CCTCCAGCAGATGCAAGAA	GCTTGCACTTCATCCTTTGG
Oct4	95	GTTGGAGAAGGTGGAACCAA	CTCCTTCTGCAGGGCTTTC
Prdm14	73	GGCCATACCAGTGCGTGTA	TGCTGTCTGATGTGTGTTCG
Sox7	97	GCGGAGCTCAGCAAGATG	GGGTCTCTTCTGGGACAGTG
Utx	25	CTGATGCAAGTCTATGACCAATTT	CAAGATGAGGCGGATGGT
TBP	97	GGGGAGCTGTGATGTGAAGT	CCAGGAAATAATTCTGGCTCA
GAPDH	52	GGGTTCCTATAAATACGGACT	CCATTTTGTCTACGGGACGA
*MXs-Oc4	39	GTGGTGGTACGGGAAATCAC	TCTGAAGCCAGGTGTCCAG
*MXs-X91	60	GTCGTGTCCAAGCCTTTACC	CTCCACGGGGTCACATTTA
*MXs-X25	49	CCAATGGGGCAATTAATGA	CCCCATCAAGCATCTCTCC
*MXs-Sox2	107	CGCCCAGTAGACTGCACA	CAAGAAAGCTGGGTTTCACAT
*MXs-C-myc	109	TCGAAACTCTGGTGCATAAGG	TGACACCAGACCAACTGGTAA
*MXs-Klf4	106	TGGTACGGGAAATCACAAGTT	GACGCGAACGTGGAGAAG
**CAG-POUV	-	CAGCTCCTGGGCAACGTGCTGG	CTTCGGCCAGTAACGTTAGG
**MXs-POUV	-	PBMN: GCTTGGATACACGCCGCCC	-
***NanogWT	-	Nanog genoF: TAACTCTTCTTTCTA TGATCTTTCCTTC	Nanog geno R: GCATCTCAGT AGCAGACCCTT G
***NanogGFP	-	IC200Rodda3F: GGGTCACCTTACAG CTTCTTTTGCATTA	GFP geno R2: TCGTGCTGCTTC ATGTGGTC

* primers used for screening of retroviral silencing in iPSC experiment

** primers used for sequencing

***primers used for Nanog-GFP genotyping (section 2.3.2), NanogWT primer set is used for detected the Nanog wild type allele and Nanog-GFP primer set is used for detecting the presence of GFP after Nanog promoter.

Section 2.2 DNA/RNA manipulation techniques

2.2.1 DNA isolation from bacteria

Bacteria containing plasmids of interest were inoculated in LB broth at volume 4 mL and 100 mL for small-scale and large-scale amplifications, respectively. 100 µg/mL of ampicillin/carbenicllin and 50 µg/mL of kanamycin were added to LB broth, depending on the resistance gene presenting in the plasmids. The bacterial culture was incubated for overnight at 37 ^{o}C with agitation at 225 rpm.

After overnight incubation, the bacteria cells were harvested by centrifuging at 5000 rpm for 10-20 minutes at 4 ^{o}C. Supernatant was discarded and plasmids were extracted using QIAGEN kits, following the manufacturer's instruction. For small-scale plasmid preparation or "mini-prep" by QIAprep Spin Miniprep Kit, the DNA was eluted with 30 µL of nuclease free water. In addition, the plasmid extraction was carried out by Illustra^{TM} PlasmidPrep Mini Spin Kit. For large-scale plasmid preparation or "maxi-prep" by the QIAGEN Plasmid Maxi kit, DNA pellet was re-suspended in 300 µL of nuclease-free water and incubated overnight at 4 ^{o}C and later stored in -20 ^{o}C.

2.2.1 DNA isolation from mammalian cells

- *DNA isolation from embryonic stem cell or induced pluripotent stem cells*

Cells were grown in appropriate culture medium until 80% confluency (6-well plates), which contained approximately more than one million cells. Genomic DNA was extracted using the Qiagen-DNeasyTM Blood and Tissue Kit according to the manufacturer's protocol.

- DNA isolation from mouse-tail for genotyping

Tail samples were lysed with 100 µL of DirectPCR (Tail) Lysis Reagent (Viagen) supplement with 2.5 µL of proteinase K (20 mg/mL) and incubated overnight at 55 ^{o}C and then inactivated at 85 ^{o}C for 45 minutes. The samples were centrifuged briefly at 4000 rpm for 3 minutes and 1 µL of crude lysate was used directly for PCR.

2.2.2 Polymerase chain reaction (PCR)

All PCR reactions of cloning and genotyping were performed with Phusion® High-Fidelity PCR Master Mix with HF Buffer (PhuHF). The primers were manually designed and optimised in Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) to prevent the primer dimer and other complications. Primer annealing temperatures were calculated using NEB-web tool (http://tmcalculator.neb.com/). All primers were synthesised by Integrated DNA Technologies Company (IDT). 20-µl volumes of PCR reactions contained following recipe: 10 µL of 2X PhuHF master mix (already containing dNTPs), 1 µL each of 10 µM forward and reverse primers, 1-100 ng of DNA template and nuclease-free water. All PCR reactions were performed as following steps in PCR machine: an initial denaturation step (30 seconds at 98 ^{o}C), followed by successive 34 cycles of denaturation (10 seconds at 98 ^{o}C), annealing (10 seconds at T_a calculated by NEB-web tool), and extension (1 minutes/kb to be amplified at 72 ^{o}C). In the end a final incubation at 72 ^{o}C 5-10 minutes took place. Some PCR primers had annealing temperature at 72 ^{o}C, thus the annealing step was omitted.

2.2.3 Restriction enzyme digestion

Restriction enzymes were used to generate compatible ends capable of joining PCR products containing DNA of interest into expression vectors or sub-cloning one gene from one vector to another vector or linearizing the DNA for further application. Briefly, 1 μg of purified DNA was digested in 50 μL volumes containing 5 μL of 10X NEB buffer, 0.5 μL of 100X BSA (the addition of BSA depended on restriction enzymes), 10 units of restriction enzyme, and nuclease-free water. The DNA/solution mixture was incubated for 2 hours. For linearization of plasmid in the rescue experiment, 100 μg of purified plasmid DNA was digested in 200 μL volumes containing 20 μL of 10X NEB buffer, 1000 units of restriction enzyme, and nuclease-free water. The DNA/solution mixture was incubated overnight. Reactions were performed at optimal temperature depending on the enzyme being used.

2.2.4 Dephosphorylation of DNA fragment ends

To prevent self-ligation of vector and facilitating the success of cloning, 5' end of linearised vector was removed by using 1 unit of Calf Intestinal Phosphatase (CIP) for 1 μg of digested DNA. The reactions were incubated at 37 ^{o}C for 30 minutes.

2.2.5 DNA fragment ligation

Digested PCR products or digested DNA fragments from other vectors (DNA inserts) were joined with destination vectors by using T4 DNA ligase. The amounts of DNA inserts and vectors were used at a molar ratio of 3:1. The calculations of ligation reaction were performed by http://nebiocalculator.neb.com/#!/ligation. The 20-μL volume ligation reaction contains 2 μL of 10X T4 DNA ligase buffer, 1 μL of T4 DNA ligase, appropriate amount of DNA insert and DNA vector, and nuclease-free water. The reactions were carried

out at room temperature (cohesive ends) for 20 minutes or at 16 ^{o}C (blunt-ends) for overnight. After incubation, 1 μL of the ligation mixture was immediately added to competent cells for transformation.

2.2.6 GeneArt Seamless cloning

The GeneArt Cloning is based on the homologous recombination to assemble adjacent DNA fragments sharing end-terminal homology. To generate the homology arms for assembly, DNA sequences of inserts containing gene of interest and vector were used for designing primers in web-based design tool (http://www.lifetechnologies.com/order/oligoDesigner). The primers were synthesised by Integrated DNA Technologies. Phusion HF-based PCR were performed to generate DNA insert with 15 bp homology at each end. Generally, the DNA inserts were joined to linearised vector pUC19L supplied by the GeneArt Assembly kit. The amounts of DNA inserts were calculated following manufacturer's instruction. The reactions were carried out in 20 μL-volume containing 100 ng of insert, 100 ng of linear pUC19L vector, 4 μL of 5X Reaction buffer, 2 μL of 10X enzyme mix and nuclease-free water. The reactions were incubated for 1 hour at room temperature and 1 μL of the reaction was used for transformation into appropriate competent cells.

2.2.7 DNA electrophoresis

The PCR products or digested DNA fragments were run on 1% agarose gel in TAE buffer. To prepare the agarose gel, 1 g of agarose powder was mixed with 100 mL of TAE buffer, and dissolved in microwave. The DNA staining chemical, SyBr-Safe, was used for 1:10 volume of agarose gel (e.g. 10 μL of SyBr-Safe in 100 mL gel) and was added directly into the gel before setting it on the apparatus. 50X TAE (Tris-Acetate-EDTA) buffer is stock

solution composed of 242 g of Tris base, 57.1 mL glacial acetic acid, and 100 mL of 500 mM EDTA (pH8.0) in total volume of 1 L. For using as running buffer, the stock solution was diluted 50:1, thus 1X of TAE contained 40 mM Tris, 20 mM acetic acid and 1 mM EDTA. For running electrophoresis, the voltage was set at 100-110 V and running time was 25-30 minutes. After this, the DNA was visualised under the UV or blue light for further imaging or gel excision/DNA purification.

2.2.8 DNA clean-up from agarose gels

DNA was extracted from gel using the Zymo Research-Zymoclean^{TM} Gel DNA Recovery Kit. The weight of excised agarose gel containing DNA was firstly measured, and the gel was dissolved in dissolution buffer with 3X volume of gel and incubated at 50 ^{o}C for 10-15 minutes. The dissolved gel was purified by column following manufacturer instruction. DNA was eluted with nuclease-free water.

2.2.9 DNA clean-up from solutions by Zymo-Research DNA kit

The DNA was purified by Zymo-Research-DNA Clean and Concentrator kit. The volume of DNA mixture solution was adjusted to 100 µL and 3 volumes of the binding buffer was added. The reaction/buffer mixture was added to the column, and processed according to the manufacturer's instruction. The DNA was eluted with nuclease-free water.

2.2.10 DNA clean-up from solution by precipitation

Firstly, DNA sample volume was adjusted to 200 µL with nuclease-free water. Next, 22 µL of 3M sodium acetate at pH 5.2 and 470 µL of ice cold 96% ethanol were added and mixed well. To pellet the DNA, the mixture was centrifuged at max speed/4 ^{o}C for 30

minutes. Then the DNA pellet was washed twice with 1000 µL of ice cold 70% ethanol and centrifuged at max speed/4 ^{o}C for 10 minutes. Then the DNA pellet was air-dried for 10 minutes under the laminar hood. The dried DNA pellet was re-suspended in 200 µL of PBS or nuclease-free water.

2.2.11 DNA quantification

DNA quantification was performed using NanoDrop 2000 (Thermo Scientific). 1 µL of samples were used for the measurement. Absorbance ratios of 260/280 (A260/280) were observed to indicate the DNA purity. An A260/280 of approximately 1.8 is generally accepted for DNA with good purity (Sambrook and Russell, 2001).

2.2.12 Bacterial transformation

These bacterial strains were used for transformation.

a) One Shot® Stbl3™ Chemically Competent *E. coli* (Invitrogen)

Genotype: F-*mcr*B *mrrhsd*S20(rB-, mB-) *rec*A13 *sup*E44 *ara*-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(StrR) *xyl*-5 λ-*leumtl*-1

b) One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen)

Genotype: F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G

c) MAX Efficiency® DH5α™ Competent Cells (Invitrogen)

*Genotype***:** F- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rk-, mk+) *pho*A *sup*E44 λ-*thi*-1 *gyr*A96 *rel*A1

Transformation: Approximately 1 μL of 100 ng/μL DNA was added to cold competent cells. The bacteria were then incubated on ice for 30 minutes, subjected to heat shock at 42 ^{o}C for 30 seconds (Stbl3) or 45 seconds (TOP10 and DH5α), followed by short incubation on ice. 250 μL of SOC medium was then added, followed by one hour incubation with agitation at 37 ^{o}C. 50-100 μL of the incubated culture was inoculated on LB agar (on Petri dish) and incubated overnight at 37 ^{o}C.

2.2.13 DNA sequencing

DNA sequencing was performed by GATC Biotech. Samples for sequencing were prepared by mixing 2.5 μL of 10 μM primers with 7.5 μL of DNA sample (300-600 ng). The results were analyzed by APE plasmid editor software.

2.2.15 RNA isolation from mammalian cells

For gene expression analysis, RNA from embryonic stem cells and induced pluripotent stem cells was isolated using either RNeasy^{TM} Mini Kit or Trizol reagent (Qiagen) according to manufacturer's instructions. Briefly, the cells were grown until 70-80% confluency in 6-well plates. The cells were washed once with PBS to remove dead cells and culture medium. Then, 400 μL of lysis buffer or Trizol reagent was immediately added to the cells, and the cell lysate was transferred to eppendof tube for extraction as following:

For RNA extraction by RNeasy^{TM} Mini Kit, the lysate was loaded onto RNeasy MinElute Spin Columns for RNA capture, following by DNaseI treatment (Qiagen), washing step and elution step as described in the protocol. The RNA was eluted in 20 μL of RNase-free water.

For RNA extraction by Trizol reagent, the lysate was mixed with 104 ul of chloroform and vortexed vigorously. The chloroform/lysate mixture was centrifuged at max speed/4 ^{o}C for 30 minutes. The upper aqueous layer containing RNA were collected and mixed with 0.6 ul of 15 mg/mL GlycoBlue. RNA then was isolated by isopropanol precipitation by adding 267 µL of isopropanol into the collected 200-µl aqueous solution. The mixture was then centrifuge at max speed for 10 minute at 4 ^{o}C. After this centrifugation, the RNA pellet (appeared blue from GlycoBlue) were wash twice with 550 µl of 75% EtOH, air dried and resuspended in 20 µL of RNase free water. To avoid the contaminated DNA in RNA samples, the RNA was treated with DNaseI (Qiagen) for 30 minutes at 37 ^{o}C. The RNA from DNaseI treatment was re-extracted with Trizol reagent.

2.2.16 RNA quantification and quality control

RNA quantification was performed using NanoDrop 2000 spectrophotometer (Thermo Scientific). 1 µL of samples was used for the measurement. At absorbance (A) 260/280 ratio, the value around 2.0 for RNA indicated that the RNA was not contaminated with organic compounds e.g. residual phenol from the extraction procedure.

For gene expression analysis by microarray, RNA quantification was performed firstly with NanoDrop and secondly with Bioanalyzer 2100 (Agilent). The Bioanalyzer provides the detail of RNA purity, quality, and concentration. The RNA was prepared and loaded onto RNA Chip according to manufacturer's instruction. The analysis was based on the assay of Eukaryote Total RNA Nano Series II. Results of RNA concentration and RNA Integrity Number (RIN) were recorded. The best quality of RNA without degradation shows RIN number 10. Only RNA samples with RIN 10 were used for microarray-based global transcriptome analysis.

2.2.17 First strand cDNA synthesis

For gene expression analysis by qRT-PCR, the cDNA were synthesised using SuperScript® III Reverse Transcriptase (Invitrogen) from high quality RNA. Briefly, 1 µg of total RNA was mixed with 5 µL of 50 ng/µL random hexamers and 1 µL of 10 mM dNTP mix and incubated at 65 ^{o}C for 5 minutes, followed by short incubation on ice. Then 4 µL of 5X First Strand (FS) buffer, 1 µL of 0.1 M DTT, 1 µL of 40,000 unit/mL RNaseOUT, and 1 µL of Superscript III enzyme were added. For control, the addition of superscript III was omitted. The reactions were incubated at 50 ^{o}C for one hour, followed by 70 ^{o}C for 15 minutes using PCR thermal cycler. The cDNA was diluted with nuclease-free water in ratio 1:100. The diluted cDNA was ready for qRT-PCR.

Section 2.3 Manipulation of transgenic mice and embryos

2.3.1 Transgenic mice

Nanog-GFP mice were obtained from Ian Chambers laboratory, University of Edinburgh. Nanog-GFP mice carried one wild type Nanog allele (NanogWT) and one Nanog allele with eGFP IRES puromycin resistance gene cassette inserted after the start codon of Nanog allele (Nanog-GFP). Homozygous embryo of Nanog-GFP/Nanog-GFP is lethal at early stage of development. Only heterozygous transgenic line (Nanog-GFP/NanogWT) can be maintained and used for all experiments. To maintain the line, the transgenic mice were crossed with transgenic or wild type 129sv (Jackson Laboratory) mice. The mice were maintained, bred, and manipulated at University of Copenhagen, SUND transgenic core facility under the project number 2012-15-2934-00142 and 2013-15-2934-00935.

2.3.2 Genotyping

Genotyping was used to confirm Nanog-GFP allele and Nanog wild type (WT) allele of transgenic mice and embryos/MEF used for iPSC experiments. Tail samples were collected by SUND transgenic core facility. The DNA isolation for genotyping is described in section 2.2.1. To confirm the presence of Nanog-GFP allele, forward primer IC200 Rodda3F and reverse primer GFP geno R2 were used in PhuHF-based PCR condition at an annealing temperature of 68 ^{o}C. To confirm NanogWT allele, forward primer Nanog geno F and reverse primer Nanog geno R were used in PhuHF-based PCR condition at an annealing temperature of 64 ^{o}C. See section 2.2.2 for PCR condition and section 2.1.4 for sequences of primers.

2.3.3 Mouse embryonic fibroblast isolation

Mouse embryonic fibroblasts (MEFs) used as feeders were derived from the C57BL/6NTac (Taconic) embryos at embryonic stage 13.5. MEFs for iPSC generation were derived from the embryo at embryonic stage 13.5. The embryos were from the cross of male Nanog-GFP mice with female 129S2/ScPasCrl (Charles Reiver). Embryos were dissected separately to avoid contamination of wild type ones. To screen the transgenic embryos, tail of each embryo was cut and place in eppendof tube for genotyping (see section 2.3.2). Briefly regarding the MEF isolation, the embryo head and internal organs were removed. The rest of embryo body was minced with scissor and incubated in 1X trypsin-EDTA for 30 minutes at 37 ^{o}C. During the incubation, the trypsinised embryos were physically separated by pipetting (with 5 mL-volume pipette) every 10 minutes. The dissociated cell suspension was re-suspended in MEF medium and centrifuged at 1800 rpm for 5 minutes. The cell pellet was then re-suspended with appropriate volume of MEF medium. For obtaining frozen stock, 500 μL of 4-8 million cells/mL cell suspension was added into cryotubes containing 500 μL of cold freezing medium composed of 20% of DMSO in FBS (see section 2.4B for MEF cell cultures and feeder preparation).

Section 2.4 Cell culture

All mammalian cells were cultured at 37°C, 5% CO_2 and 90% humidity (BBD 6220 incubators, Thermo Scientific), unless otherwise stated. All solutions were sterile and all cell manipulations were performed in a sterile laminar flow hood. The medium was warmed at 37 ^{o}C prior to use.

Section 2.4A Mouse embryonic stem cell culture

2.4A.1 The maintenance of mouse ESCs

Routinely, ESC were cultured in GMEM containing 0.1 mM non-essential amino-acids, 2 mM L-glutamine, sodium pyruvate, 0.1 mM β –mercaptoethanol, 10% Fetal Bovine Serum (FBS) and LIF. The flasks/dishes (Corning) for ESC culture were coated with 0.1% gelatin in PBS for 5-10 minutes. The ESCs were grown until 80% confluency before passaging or further applications. For passaging, the cells were washed once with warm PBS and detached by 1X trypsin-EDTA with 3 minutes incubation at 37 ^{o}C. ESC medium was then added to a flask/dish and the cell suspension was pipetted several times to generate a single cell suspension (neutralization step, FBS in ESC medium blocks the trypsin activity). Cells were collected by centrifugation at 1300 rpm for 3 minutes. Cells were then re-suspended in appropriate volumes of ESC medium. In general, 1:5 split ratio was used.

2.4A.2 Mouse ESC colony picking and expansion

10 μL of 1X Trypsin-EDTA was added to V-bottom 96 well plates. Medium was aspirated from the culture dish (100 mm^2) and cells were washed once with 10 mL of PBS.

12 mL of PBS was then added to the dish. The colonies were picked with a p10 pipette and transferred into trypsin solution and incubated for 5 minutes at room temperature. Then 200 µL of ESC medium was added into each well using multi-channel pipette and vigorously pipetted to generate single cell suspension. The cell suspension was then transferred to gelatin-coated flat-bottom 96-well plate and incubated overnight. The medium was changed the next day. About two days after picking when the ESCs grew until around 80% confluency, they were passaged to 48-, then 12- and 6-well plate for cell expansion and freezing. The procedures were described following:

Passaging from 96- to 48-well plate: washing with PBS (200 µL/well), cell detachment (10 µL of trypsin), neutralization/vigorous pipetting (100 µL of ESC medium) and transferring all cell suspension to gelatin-coated well containing 500 µL of ESC medium.

Passaging from 48- to 12-well plate: washing with PBS (500 µL/well), cell detachment (50 µL of trypsin), neutralization/vigorous pipetting (500 µL of ESC medium) and transferring all cell suspension to gelatin-coated well containing 1 mL of ESC medium.

Passaging from 12- to 6-well plate: washing with PBS (1 mL/well), cell detachment (100 µL of trypsin), neutralization/vigorous pipetting (1 mL of ESC medium) and transferring only 500 µL of cell suspension to gelatin-coated well containing 2 mL of ESC medium.

Passaging from 6- to 6-well plate: washing with PBS (2 mL/well), cell detachment (500 µL of trypsin), neutralization/vigorous pipette (3 mL of ESC medium), centrifugation at 1300 rpm for 3 minutes, and resuspension with 500 µL of ESC medium, (for passaging) transferring 100 µL of cell suspension to gelatin-coated well containing 2 mL of ESC medium, (for freezing) transferring 500 µL to cryotube containing 500 µL of ice cold ESC medium with 20% DMSO.

2.4A.3 Freezing mouse ESCs

Freezing medium composed of 10% DMSO (v/v) in ESC medium was prepared fresh and stored on ice before use. The cells were harvested after passaging. Cells were re-suspended in cold freezing medium to obtain 2 million cells/mL and transferred to cryotube (1 mL per vial). The tubes were stored at -80 ^{o}C overnight. On the following day, the cryotubes were transferred to liquid nitrogen cell bank.

2.4A.4 Thawing mouse ESCs

Frozen cells from liquid nitrogen were quickly thawed at 37 ^{o}C. The defrosted cell suspension was transferred to 9 mL of warm ESC medium. The cells were collected by centrifugation at 1100 rpm for 5 minutes. The cell pellet was gently re-suspended with appropriate volume of ESC medium and transferred to appropriate gelatin-coated vessel. The medium was changed on the following day.

2.4A.5 Transfection of mouse ESCs

a) Transfection by liposomes

Transfection was performed by using Lipofectamine® 2000 DNA transfection reagent (Invitrogen). Briefly, one million ESCs were seeded onto one well of gelatinised 6-well plate, followed by one-hour incubation. During this incubation, 3 μg DNA and 3 μL lipofectamine were added to 250 μL OptiMEM medium or serum-free ESC medium (in polystyrene tubes) and incubated at room temperature for 30 minutes. After one hour, the DNA/transfection agent mixture was added in dropwise manner to the culture and the cells

were incubated overnight. Medium was changed on the following day. The cells were collected for further applications after two days of transfection.

b) Transfection by electroporation

Transfection by electroporation was done as following:

1) The preparation of linearised plasmid

100 µg of plasmid was linearised with restriction enzymes as described in section 2.2.3. The linearised plasmids were then purified by DNA precipitation as described in section 2.2.10. The DNA pellet was air-dried and re-suspended in 200 µL PBS in a laminar flow hood.

2) ESC preparation

ESCs were grown until 80% confluency in 75 cm^2 flasks (containing around 20 million cells/flask). The cells were harvested as described in 2.4A.1. The cells were counted and 10 million cells were transferred to a universal tube. The cells were collected by centrifugation at 1300 rpm for 3 minutes, washed once with 10 mL of warm PBS, collected by centrifugation again. The cell pellet were then re-suspended in 600 µL of warm PBS, mixed with 200 µL of DNA and allowed to stand at room temperature for 10 minutes. The DNA/cell suspension was transferred to a 4 mm electroporation cuvette (Biorad). The electroporation was carried out by electroporator BioRad at 0.8 kV and 10 µF. The successful electroporation gave time constant around 0.1. After electroporation, the cells were allowed to rest at room temperature for 10 minutes. The cells were then gently transferred to a universal containing 10 mL of ESC medium. One million cells were plated

on gelatin-coated 100 mm culture dish containing 10 mL of ESC medium. For rescue experiment, the medium was changed after two days of electroporation.

Section 2.4B Mouse embryonic fibroblast cell culture

2.4B.1: MEF cell culture

Both MEF for feeder preparation and iPSC induction were cultured in MEF medium composed of DMEM (high glucose), 10% fetal bovine serum (FBS), 0.1 mM MEM non-essential amino acids (NEAA), 2 mM L-glutamine, 0.1 mM β-mecaptoethanol, and 1% penicillin-streptomycin (Invitrogen). For passaging, the cells were washed once with warm PBS and dissociated by adding 1X trypsin-EDTA and incubating for 5 minutes. After incubation, the trypisinised cells were re-suspended with MEF medium in ratio 1:10 (1 mL of trysinised cells with 9 mL of MEF medium). The cell suspension was centrifuged at 1800 rpm for 5 minutes. The cell pellet was then re-suspended in the appropriate volume of MEF medium.

2.4B.2: Feeder preparation

MEF cells from frozen stock were defrosted and expanded in MEF medium. The cells were cultured in 150 cm^2 flask until 90% confluency. The cells were harvested as passaging procedure described in section 2.3.4 and expanded in ten of 150 cm^2 flasks until 90% confluency. To inactivate the cell division, the MEF cells were treated with mitomycin-C (Sigma) at 10 ug/mL final concentration for 2 hours. Alternatively, the MEFs were irradiated with gamma radiation treatment at 3000 rad for 30 minutes. After treatments, the cells were harvested for obtaining frozen stock. 500 µL of 4-8 million cells/mL cell suspension was prepared and added into cryotubes containing 500 µL of cold freezing

medium composed of 20% of DMSO in FBS. The cells were stored in 80 ^{o}C and transferred to liquid nitrogen cell bank on the following day.

Section 2.4C Induced pluripotent stem cell culture

2.4C.1:The maintenance of iPSCs

Unless otherwise specified, all iPSCs were cultured on the irradiated feeders or mitomycin-treated feeders. The feeder preparation is described in section 2.4B.2. Two million of feeders were seeded onto gelatinised 6-well plate (around 300,000 cells/ well, 35 mm^2 surface area). The defined iPSC medium composed of 390 mL of DMEM high glucose, 100 mL of knockout^{TM} serum replacement (KOSR), 5 mL of non-essential amino acid, 5 mL of L-glutamine, 500 µL of B-mercaptoethanol, and 550 µL of LIF, 1000 µL of 10 mg/mL Vitamin C (L-ascorbic acid), 25 µL of 10 µM Alk5 inhibitor. The defined medium was used for both iPSC induction and maintenance.

2.4C.2: iPSC colony picking, expansion and freezing

The process of iPSC colony picking was the same as ESC colony picking described in section 2.4A.2, except that iPSCs were seeded onto flask/dishes coated with feeders. In general, the feeders were seeded one or two days before colony picking and expansion. The freezing stock preparation was the same as to ESC described in section 2.4 A.3. The cell pellet for freezing was re-suspended in defined iPSC medium to obtain 2-4 million/mL. 500 µL of cell suspension was added into cryotube containing cold freezing iPSC medium (20% DMSO in KOSR). The cells were stored at -80^{o}C and transferred to liquid nitrogen cell bank on the following day.

Section 2.4D Retrovirus production

The ecotropic retrovirus production was performed under GMO class I. The retrovirus strategy is illustrated in figure 2.1.

2.4D.1 Packaging cell lines for retrovirus production

293LTV (CellBioLab) cell line was used as packaging cell line for retrovirus production. The cells were maintained in the complete medium composed of DMEM (high glucose), 10% fetal bovine serum (FBS), 0.1 mM MEM non-essential amino acids (NEAA), 2 mM L-glutamine, and 1% Pen-strep. The cells were passaged in the same way as ESCs section 2.4A.1.

2.4D.2 Retroviral transfection

To produce retrovirus particle, packaging cell lines were transiently transfected with two expression vectors: pMXs-vector carrying gene of interest and pCL-ECO containing modified gene encoding retroviral components. pCL-ECO packaging vector consists of three retroviral genes: *gag*, *pol*, and *env* which are essential to establish complete retroviral particles capable of infection. Types of envelop protein, gp70 encoding by *env* determine the host of infection. The retrovirus produced by this pCL-ECO is ecotropic (MoMuLV based), meaning that it can infect only mouse and rat cells, but not human cells.

Transient transfection was performed using Lipofectamine LTX (Invitrogen). Briefly, the 293LTV cells were grown until 50% confluency. At the time of transfection, 50 μg of retrovirus expression vector pMXs, 25 μg of packaging vector and PLUS reagent (supplied with Lipofectomine LTX) were mixed in 1.50 mL of OptiMEM I (Invitrogen) to

obtain the “mixture 1”. Then, 100 µL of Lipofectamine LTX was mix with 1.50 mL of OptiMEM I to obtain the “mixture 2”, which was incubated at room temperature for 5 minutes. After the incubation, both mixtures were mixed together and incubated further for 20 minutes. The mixture of DNA/Lipofectamine was then added in dropwise manner onto the cells cultured in 25 mL of the 293 LTV medium. The cells with transfection reagents were incubated at the normal cell culture conditions overnight. The medium was changed on the following day. Retroviral expression vector with dsRED gene (pMXs-dsRED) was used to roughly monitor the transfection efficiency, and the strong expression of dsRED could be noticed in 293LTV after 24 hours. The retrovirus from pMXs-dsRED transfection was also collected for testing infection efficiency at the later step.

2.4D.3 Collection and concentrating of retrovirus

Retrovirus supernatant or medium containing virus particles was harvested at day 2 and day 3 after transfection to obtain two batches of retrovirus. The viral supernatant was filtered with WHATMAN 0.45 µM filter unit (made from cellulose acetate) to remove the contaminated 293LTV cells and debris. The retrovirus then was concentrated by Retro-Concentrator (Clontech) solution in ratio 1:3 and incubated at 4 ^{o}C overnight. On the following day, the virus supernatant/concentrator mixture was centrifuged at 1500 g for 45 minutes at 4 ^{o}C. The virus pellet was then re-suspended in small volume of MEF medium to obtain concentrated virus stock, which were aliquoted for single use and stored at -80 ^{o}C.

2.4D.4 Measurement of retrovirus titer

The titer of retrovirus was measure by Retro-X qRT-PCR Titration Kit (Clontech). Briefly, retroviral RNA were purified from fresh retrovirus with NucleoSpin RNA Virus Kit, following manufacturer instruction. The final elution was done with 50 μL of RNase-free water. To avoid the contamination of DNA from transient transfection of previous steps, the RNA samples were treated with DNase I (Clontech, 12.5 μL of RNA samples, 2.5 μL of DNaseI buffer, 4 μL of 5 units/ul DNase I, 6.0 μL of RNase-free water), and the RNA/DNaseI were incubated at 37 ^{o}C for 30 minutes and then at 70 ^{o}C for 5 minutes. The treated RNA was stored in -80 ^{o}C before the qRT-PCR step. The SyBrGreen-based qRT-PCR was performed according to the manufacturer's instruction (Clontech). Briefly, RNA samples were serially diluted in four concentrations: 1X, 0.1X, 0.01X and 0.001X. The RNA control template was serially diluted in five concentrations (copies/μL): $5X10^{7}$, $5X10^{6}$, $5X10^{5}$, $5X10^{4}$, and $5X10^{3}$. The qRT-PCR mastermix solution was composed of 5 μL of 2X Quant-X buffer, 0.2 μL of 10 μM retro-X forward primer, 0.2 μL of 10 μM retro-X reverse primer, 0.2 μL of Quant-X^{TM} enzyme, 0.2 μL RT-Enzyme Mix, 2.2 μL of RNase-free water and 2 μL of serially diluted RNA. The qRT-PCR reactions were carried out in a LightCycler480 (Roche) with following conditions: reverse transcription (RT) reaction (42 ^{o}C 5 minutes, 95 ^{o}C 10 seconds), 40 cycles of qPCR (95 ^{o}C 5 seconds, 60 ^{o}C 30 seconds), and dissociation curve (95 ^{o}C 15 seconds, 60 ^{o}C 30 seconds). The standard curve was constructed based on the serial diluted RNA template with known concentration of retroviral copies. The calculations of retrovirus RNA copy number were done in LightCycler480 software and EXCEL.

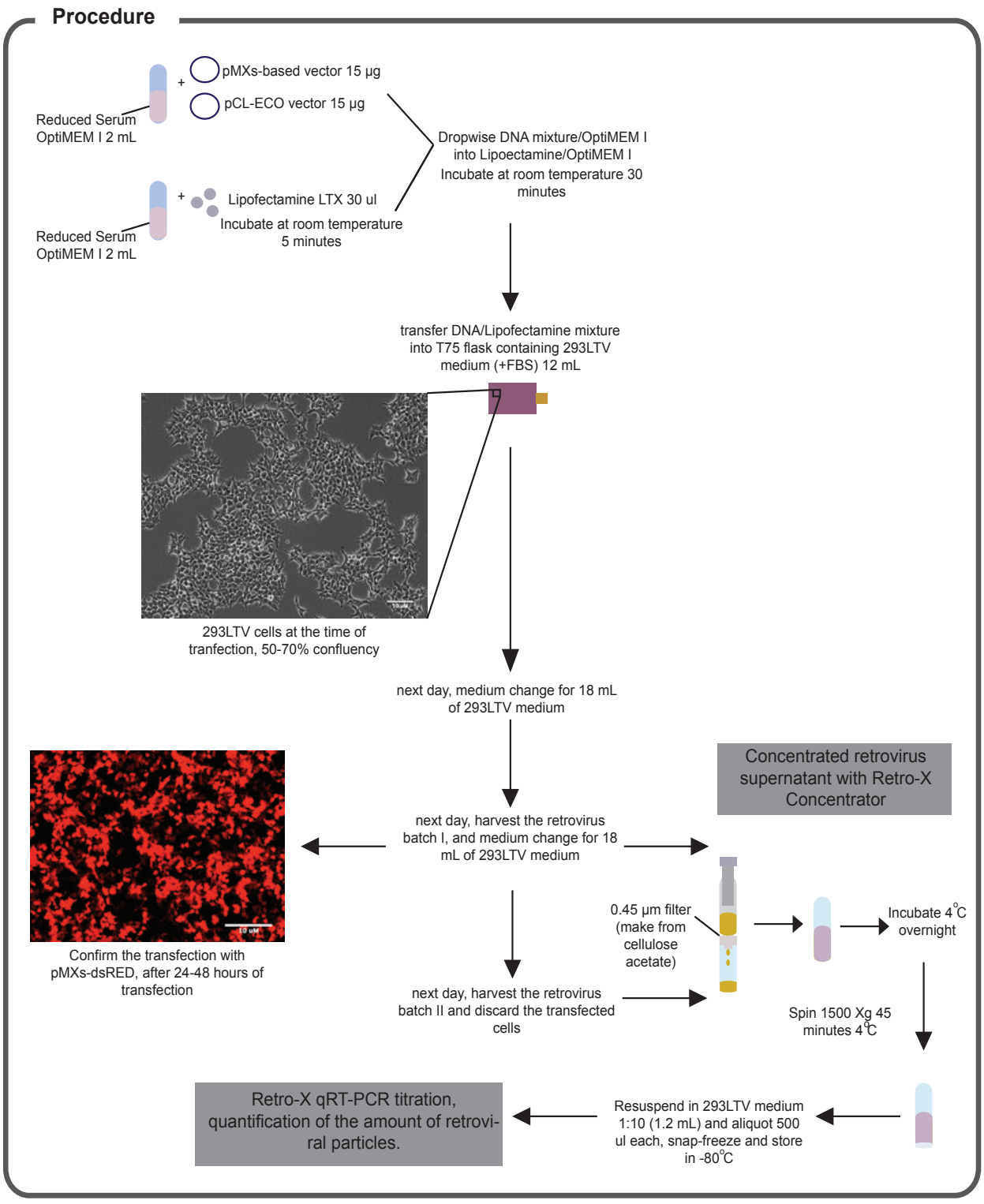


Figure 2.1 Retrovirus production for iPSC generation
see the description in section 2.4D

Section 2.5 Cell analysis and histological techniques

2.5.1 Immunocytochemistry

ESC cells were seeded onto 8 well 15µ-Slide (Ibidi) at a density 20,000 cells/well. The cells were grown for two days. The cells were washed once with warm PBS, and fixed with 150 µL of 4% paraformaldehyde (PFA) for 15 minute at room temperature. The fixed cells were washed three times with PBS and were stored in PBS at 4 ^{o}C in case immunostaining was not immediately performed after the fixation. The cells were incubated in the blocking solution (200 µL of TritonX 100, 2 mL of serum and 1 mL of 7.5% BSA in 100 mL of PBS) for one hour prior to the staining procedure. The types of serum used in the blocking solution were the same as the secondary antibodies' host. The primary antibodies were diluted in the blocking solution as described in section 2.1.3. The primary antibody solution was added into the fixed cells, followed by an overnight incubation at 4 ^{o}C. On the following day, the stained cells were washed three times with PBS prior the staining with secondary antibodies. Secondary antibodies were diluted 1:800 in the blocking solution. The solution was then added onto the cells with one-hour incubation at room temperature in the dark. The cells were washed three times with PBS, stored in PBS and visualised under microscrope. All immunofluorescence-stained cells were imaged in Leica AP6000 microscope. All images in each experiment were obtained using the same settings of the microscope.

2.5.2 Flow cytometry and Fluorescence-activated cell sorting (FACS)

ESCs/iPSCs were collected as described in 2.4A.1 and 2.4C.1. The cells were stained with appropriate dilution or primary antibodies (as section 2.1.3) in FACS buffer (10% FBS in PBS) with 15-minute incubation at 4 ^{o}C (dark). The cells were washed three times with FACS buffer and re-suspended in cold FACS buffer containing DAPI (100 ng/mL). If secondary antibodies were required, the cells were further stained with appropriate dilution of secondary antibodies for 15 minutes, washed three times with PBS, re-suspended in cold FACS buffer containing DAPI. If only one surface marker was examined, the primary and secondary antibodies were added at the same time for the staining. DAPI was used in all FACS analysis to exclude the dead cells. All experiments included unstained E14Tg2A ESCs as non-fluorescent control required for gating process. Flow cytometry analysis and cell sorting were carried out on a Fortessa (BD Bioscience), FACS Aria Cell Sorter II SORP (BD Bioscience). Flow cytometry data from Fortessa and FACS Aria were further analyzed in FCS Express (De Novo Software).

The amounts of cells used for sorting depended on the percentage of required subpopulations. The cells were sorted into FACS buffer-coated FACS tubes. The sorted cells were collected by centrifugation at 500 g for 10 minute at 4 ^{o}C. The supernatant was carefully removed by aspirator. For RNA extraction, 350 µL of buffer RLT was directly added to the cell pellet in FACS tubes and transferred to eppendofs for further RNA extraction procedure as described in 2.2.15 or stored at -80 ^{o}C.

2.5.3 Alkaline phosphatase staining

The Leukocyte Alkaline phosphatse Kit (Sigma, 86R-1KT) was used for the staining, Before starting the procedure, cells were washed once in PBS and fixed for one minute with a fresh solution containing 25 mL of citrate solution, 8 mL 37% formaldehyde and 65 mL acetone. The fixed cells then were washed twice with tap water and stained with fresh AP solution. 400 µL of FRV alkaline phosphatase solution and 400 µL of sodium nitrate solution were firstly mixed and incubated in dark for 2 minutes. Then the mixture was added into 18 mL of water and mixed well. 400 µL of naphthol was lastly added to the mixture. 5 mL of the mixture was immediately added to the fixed cells, followed by 25-minute incubation in dark at room temperature. After the incubation, the undifferentiated ESC colonies obtained red-violet colour. The stained cells were washed twice with tap water and air dried overnight. The photos of AP colonies were taken by microscope Leica-5500B (10X magnification). All AP images were then processed with FIJI, imaging processing software. The colonies were categorised into 3 classes: undifferentiated, mixed and differentiated colonies based on the intensity of AP staining. Undifferentiated colonies are AP positive containing less than 5% AP negative differentiated cells around periphery of the colony. Mixed colonies exhibit AP positive cells in the centre of the colony, but also contain from 10 to 80% of differentiated cells around periphery. Differentiated colonies show only faint AP staining and consist of at least 90% differentiated cells (figure 2.2). The rescue index was calculated by dividing (1) the number of rescued AP positive ES cell colonies obtained in the absence of endogenous Oct4 with (2) the number of rescued AP positive colonies obtained in the presence of endogenous Oct4 for a given transfection.

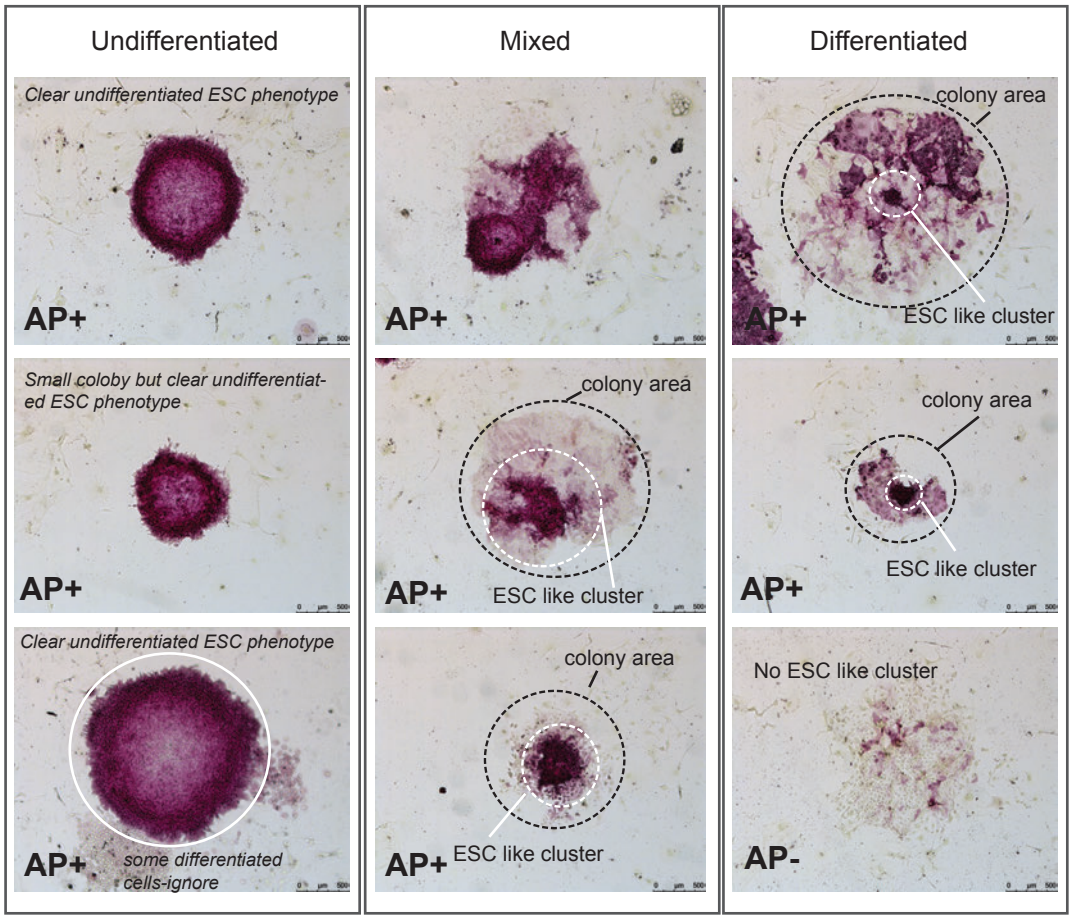


Figure 2.2 The classification of alkaline phosphatase (AP) stained ESC colonies
see the description in section 2.5.3.

2.5.4 Karyotyping

ESCs/iPSCs were grown until 50-70% confluence. The cells were incubated with ESC/iPSC medium containing 0.1 μg/mL demelcocine (colcemid) at 37 ºC for 40 minutes. Demecolcine depolymerises microtubules, leading to the cell arrest at metaphase. After incubation, the metaphase-arrested cells were collected and incubated in 2.5 mL of hypotonic solution (0.56% KCl) for further 6 minutes. Cells were fixed in 1 mL of fixative (3 methanol : 1 acetic acid) at 4 ºC for 30 minutes. Cells in fixative were dropped onto a cleaned glass microscope slide and air-dried. The slides were then stained with either 10% Giemsa (pH 7.2) or DAPI (1:10,000) for 20 minutes. The stained chromosomes can be visualised and counted under microscope (Leica-5500B at 63X magnification)

Section 2.6 Gene expression analysis

2.6.1 Quantitative RT-PCR (qRT-PCR)

Quantitative RT-PCR was performed using Roche Universal ProbeLibrary (UPL) System. UPL primers were designed by using Roche Assay Design Centre (http://lifescience.roche.com/shop/CategoryDisplay?identifier=Universal+Probe+Library). The list of UPL primers and probes are noted in section 2.1.4. Reactions were performed using the LightCycler® 480 Probes Master Mix (Roche) and UPL Set, Mouse (Roche) according to manufacturer's instructions. Briefly, 1 reaction of UPL qRT-PCR composed of 5 μL of Probes Master Mix, 0.45 μL of 10 μM forward/left primer, 0.45 μL of 10 μM reverse/right primer, 0.1 μL of specific probe, 2 μL of diluted first strand cDNA, and 2 μL of RNase-free water. The RNA and cDNA preparations are described in 2.2.15 and 2.2.17. The stand curve for each gene assay was constructed based on the serial dilution of cDNA pool (the combination of cDNA from all samples). The concentration of transcripts of each gene was calculated in LightCycler software based on the cDNA pool-derived standard curve. The values of concentration from each gene of interest were normalised to that of housekeeping gene TBP for obtaining the relative transcript level.

2.6.2 Agilent one-colour microarray

Global transcriptomes of POUV-rescued lines and iPSC lines were obtained by using Agilent one-colour microarray-based gene expression analysis, following manufacturer's instruction. Only samples with high quality RNA (RIN = 10) were used for the analysis to ensure the chance of successful and reliable outcomes. The RNA extraction and quality control are described in section 2.2.15 and 2.2.16. For both rescue and iPSC experiments, total RNA was labeled in Cyanine 3 CTP labeling reaction according to the manufacturer's instruction and purified by Qiagen's RNeasy Mini Spin Columns. The quantity of purified cRNA was measured by Nanodrop UV spectrophotometer. 600 ng of cRNA from each sample was used for fragmentation according to the manufacturer's instruction. The fragmented cRNA was then added to the 8X60K Agilent slides containing probes specific for mouse genes (Grid_GenomicBuild: mm9:NCBI37:Jul2007), followed by 17-hour hybridization step at 65 ^{o}C, and washing with Agilent wash buffer. The processed slides were scanned with Agilent Scanner (Agilent Technologies, G2600D SG12524268), following manufacturer's instruction. The data were analyzed by NIA Array Tool Analysis (http://lgsun.grc.nia.nih.gov/ANOVA/bin/login.cgi) for comparative global transcriptome analysis. Overexpressed genes and underexpressed genes from pairwise comparison were extracted to analyze the enrichment of gene categories based on several features e.g. molecular functions, biological functions, cellular functions, protein domains (Gene-annotation enrichment analysis or Gene ontology (GO) enrichment analysis). The GO analysis was done using NIA tools (http://lgsun.grc.nia.nih.gov/ANOVA/) and ExATLAS (http://lgsun.grc.nia.nih.gov/exatlas/), BINGO in Cytospace, and Panther (http://pantherdb.org). The lists of genes were also analyzed by GENEMANIA (http://www.genemania.org) to visualise the gene network. For GO analysis by BINGO, the parameters were including 1) Binomial statistical test 2) Bonferroni Family Wise Error Rate

(FWER) correction 3) significant level $P<0.005$ 4) visualization by overrepresented categories after correction and 5) GO_Biological_process as ontology file.

Section 2.7 ZHBTc4 rescue experiment

ZHBTc4 cell line designed by Hitoshi Niwa (Niwa et al., 2000) was used for Oct4 rescue experiment. The Oct4-null ESC rescue strategy is illustrated in figure 2.3B. In ZHBTc4 ESC cell line, both alleles of endogenous Oct4 are disrupted by homologous recombination and tetracycline-regulated Oct4 transgene is randomly integrated. Without adding tetracycline (Tc), Oct4 can be expressed from the transgene and the cell line can maintain its ESC self-renewal and pluripotency. In contrast, adding Tc blocks Oct4 expression from the transgene, leading to its dramatic reduction of Oct4 protein in 12 hours (Niwa et al., 2000). ZHBTc4 with Tc quickly differentiates toward trophectodermal lineage in a few days. Thus, we can test the ability of other POUV proteins in rescuing Oct4-null ESC by their constitutive expression in the same time of adding Tc to remove mouse Oct4 expression (figure 2.3A).

The sources of POUV genes used for rescue assay are listed in section 2.1.2. POUV coding sequences were subcloned into plasmid pPCAG575-SiP between CAG promoter and IRES-Puromycin resistant gene (PAC). The pPCAG575-SiP-POUV plasmids were linearized according to section 2.4A.5 (without removing Origin of replication). The 10 million of ZHBTc4 ESC cells were electroporated with 100 µg of linearised plasmids. One million of electroporated cells were then plated into gelatinised 100 mm culture dish containing 10 mL of ESC medium with and without tetracycline (Tc, 2 µg/mL). Thus, for each condition there were five plates with Tc and another five plates without Tc. In all experiments, 3XFlag mOct4 and empty vector were tested as the positive and negative control, respectively. At day 2 post electroporation, the medium was changed to ESC medium supplement with 1µg/mL puromycin (with and without Tc) to select the cells expressing transfected *POUV* genes. The medium was changed every two days. At day 9

post electroporation, the ESC colonies could be seen and were ready for alkaline phosphatase staining (section 2.5.3). At the same day of AP staining, 24 clones of POUV-rescued clones were picked and 12 clones were further expanded for obtaining frozen stocks (section 2.4A.1, 2.4A.2 and 2.4A.3) and other applications including immunofluorescent staining (section 2.5.1), flow cytometry (section 2.5.2), qRT-PCR (section 2.6.1), microarray-based transcriptome analysis (section 2.6.2). During passaging and cell preparation for all investigations, POUV-rescued clones were maintained in ESC medium supplement with puromycin and Tc all the time.

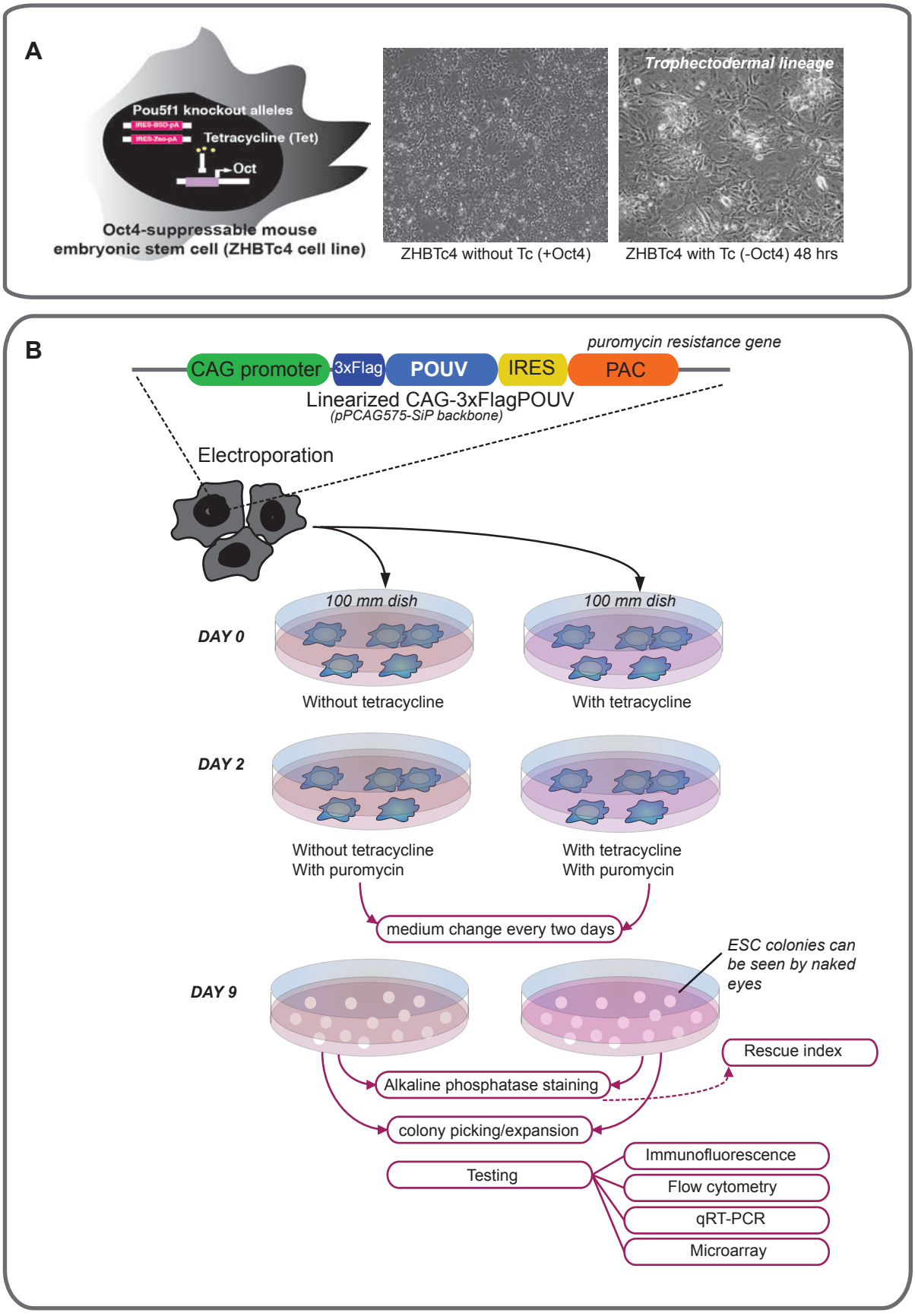


Figure 2.3 Procedures for Oct4-null ESC rescue experiment
see the description in section 2.7

Section 2.8 Procedure for the generation of iPSCs

The iPSC strategy is illustrated in figure 2.4. Induced pluripotent cells (iPSC) were established by ecotropic retrovirus-based induction of Yamanaka transcription factors: mouse Oct4, Sox2, Klf4 and c-Myc. In this study, mouse Oct4 in retroviral expression vector was replaced by *Xenopus* Pou91 (Xlpou91, Pou5f3.1) and *Xenopus* Pou25 (Xlpou25, Pou5f3.2) in the murine cellular reprogramming. The detail of plasmids is described in section 2.1.2. The retroviral expression system was based on pMXs-vector, which was originally used by Shinya Yamanaka laboratory for the first iPSC induction (Takahashi and Yamanaka, 2006).

The source of somatic cells for the reprogramming was Nanog-GFP mouse embryonic fibroblasts (Nanog-GFP MEFs), in which the cells carry green fluorescent protein (GFP) expression under the control of one *Nanog* promoter. The emergence of GFP indicated the Nanog expression, which was used to monitor the pluripotency acquisition and reprogramming behavior. The procedures for isolation and maintenance of Nanog-GFP MEFs are described in section 2.3.

Regarding the induction, 100,000 MEF cells were seeded onto gelatin-coated 6-well plate in 2 mL of MEF medium per well, followed by overnight incubation. On the following day, MEF cells were observed under microscope to ensure that MEF cells were grown until approximately 50-60% confluency. For infection, the following amounts of retrovirus were added to 1 mL of MEF medium containing 8 μg/mL polybrene.

(1) Negative control 1: no retrovirus added

(2) Negative control 2 - "SKM only":

$1X10^9$ retrovirus copies of Sox2, Klf4 and c-Myc

(3) Negative control 3 - "dsRED low SKM":

$1X10^9$ retrovirus copies of dsRED, Sox2, Klf4 and c-Myc

(4) Negative control 4 - "dsRED high SKM":

$5X10^9$ retrovirus copies of dsRED and $1X10^9$ retrovirus copies of Sox2, Klf4 and c-Myc

(5) Positive control/ POUV testing 1 - "mOct4 low dose SKM":

$1X10^9$ retrovirus copies of mouse Oct4, Sox2, Klf4 and c-Myc

(6) POUV testing 2 - "mOct4 high dose SKM":

$5X10^9$ retrovirus copies of mouse Oct4 and $1X10^9$ retrovirus copies of Sox2, Klf4 and c-Myc

(7) POUV testing 3- "Xlpou91 low dose SKM":

$1X10^9$ retrovirus copies of Xlpou91, Sox2, Klf4 and c-Myc

(8) POUV testing 4- "Xlpou91 high dose SKM":

$5X10^9$ retrovirus copies of Xlpou91 and $1X10^9$ retrovirus copies of Sox2, Klf4 and c-Myc

(9) POUV testing 5- "Xlpou25 low dose SKM":

$1X10^9$ retrovirus copies of Xlpou25, Sox2, Klf4 and c-Myc

(10) POUV testing 6- "Xlpou25 high dose SKM":

$5X10^9$ retrovirus copies of Xlpou25 and $1X10^9$ retrovirus copies of Sox2, Klf4 and c-Myc

The retrovirus were prepared immediately prior to the infection. To start infection, the medium on the MEF cultures was aspirated and immediately replaced by the medium containing retrovirus. The cells incubated overnight and the same procedure of infection was repeated on the following day. The day of the first infection was count as "Day 0". At day 2, cell media was replaced for 2 mL of MEF medium. On day 3, MEF medium was replaced with defined iPSC medium (section 2.4C.1). On day 4, the cells were collected by trypsinization (section 2.4 B.1). The cells were counted by haematocytometer and were then

collected by centrifugation at 1800 rpm for 5 minutes. The cells were re-suspended with defined MEF medium to obtain 10,000 cells/mL and reseeded onto irradiated feeder (10,000 cells/well). The cells were placed in the incubator without disturbing for two days. Medium was changed daily from day 6 to day 10 and every two-day from day 12-26. The experiments was stopped at day 27 post the first infection. On day 24-28, iPSC colonies were picked and expanded as described in section 2.4A.2 and 2.4C.2.

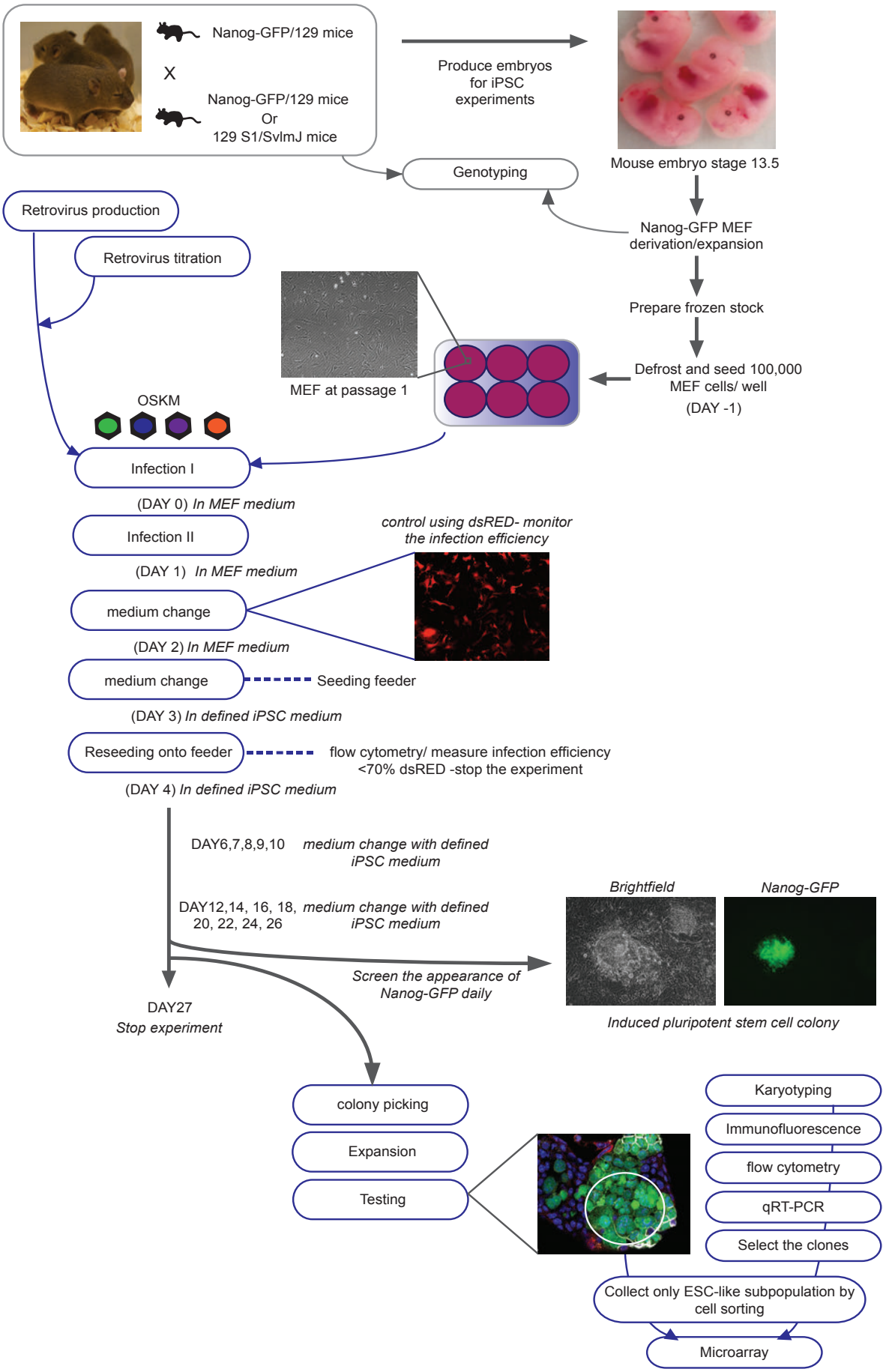


Figure 2.4 Procedure for iPSC generation and analysis
see the description in section 2.8

Section 2.9 Phylogenetic analysis and the estimation of rate of protein evolution

The procedure for phylogenetic analysis is illustrated in figure 2.5. The sources of POU5F1 and POU5F3 proteins used in this study are shown in table 2.2 and 2.3. Phylogeny of vertebrates was estimated based on a manually-curated POU protein alignment (only POU domains) under Bayesian phylogenetic framework by using BEAST 1.8.0. The best amino acid substitution model (JTT+G) determined by ProtTest 2.4 was applied. The procedure for BEAST analysis is illustrated in figure 2.5. Parameters used for BEAUTi v1.8.1 set up as following:

1) Substitution Model: JTT (Jones et al., 1992)

2) Site heterogeneity Model: Gamma

3) Number of Gamma Categories: 4

4) Clock Model: Lognormal relaxed clock (Uncorrelated)

5) Tree Prior shared by all tree models: Speciation/Yule Process

6) MCMC: Length of chain:10,000,000, Echo state to screen every: 10,000, Log parameters every: 10,000

7) Taxa and Tree Prior TMRCA was set as shown in Figure 2.6 and Table 2.4.

The XML files with following parameters and input protein sequences were generated by BEAUTi and subjected to further process with BEAST. I used online platform CIPRES SCIENCE GATEWAY (https://www.phylo.org) for BEAST analysis (Miller et al., 2010). Two outputs from BEAST analysis (.log and .tree file) were analyzed for statistical outcomes of the phylogenetic tree and constructed the phylogenetic tree by TreeAnnotator v.1.8.1 and visualised by FigTree (v.1.4.2). The rate of protein evolution was visualised as gradient colour in FigTree.

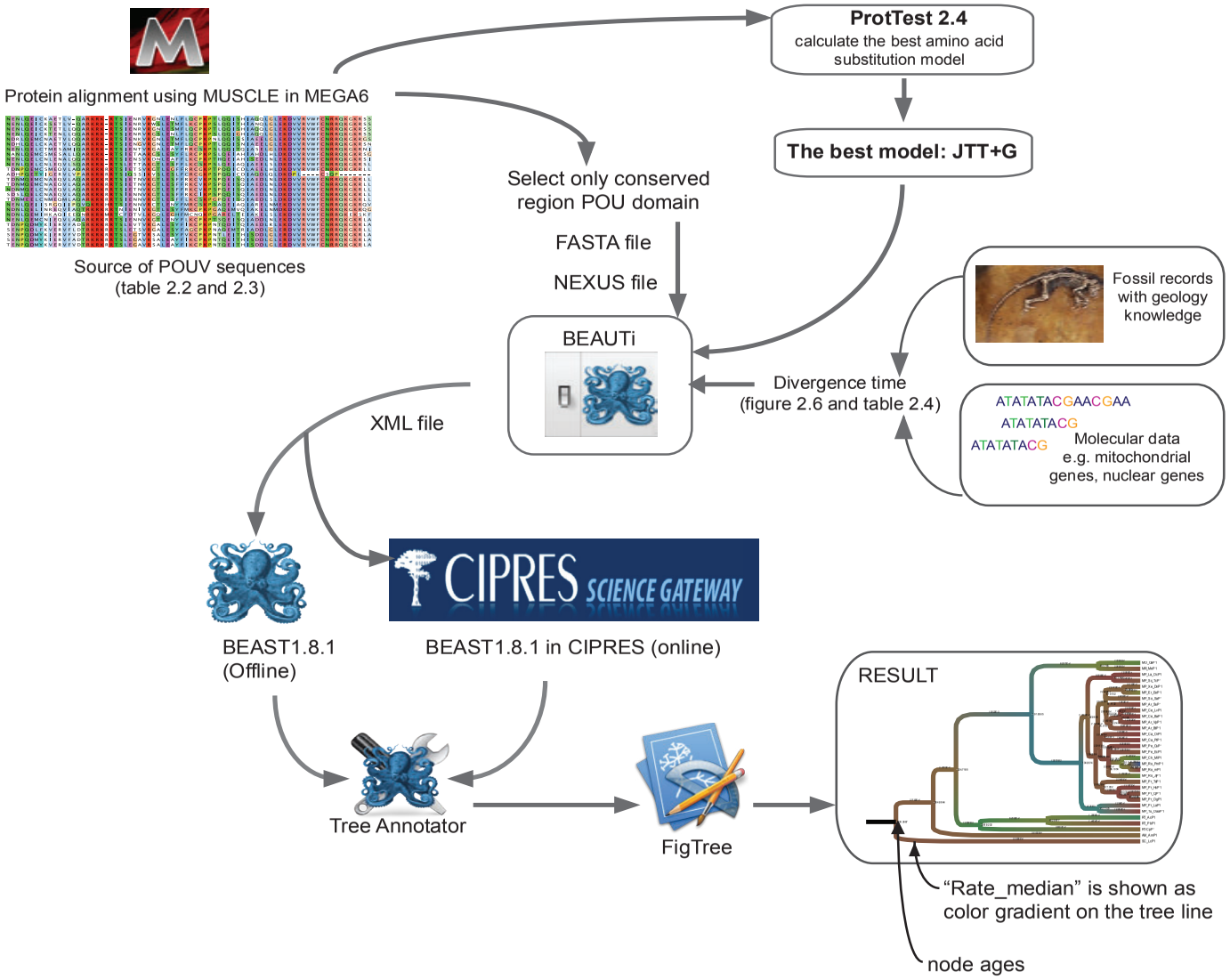


Figure 2.5 Procedures for phylogenetic analysis and estimation of rate of protein evolution
see the description in section 2.9

Table 2.2 Sources for POU5F1 protein sequences

Species	Genbank ID
Latimeria chalumnae (Coelacanth)	AFYH01090867 (17555-18199), BAHO01146254 (698-809, 2975-3105, 3982-4143, 4560-4850)
Ambystoma mexicanum (Axolotl or Mexican salamander)	AY542376
Python bivittatus (Burmese python)	XP_007432810.1
Anolis carolinensis (Green anole)	XP_008120168.1
Chrysemys picta bellii (Painted turtle)	AHGY01526669 (432-995), AHGY01303135 (1193-1082, 623-493), AHGY01303134 (178-17), AHGY01303133 (2081-1811)
Equus caballus (Horse)	XP_001490158
Ceratotherium simum simum (White rhinoceros)	XP_004424441
Odobenus rosmarus divergens (Walrus)	XP_004409093
Panthera tigris altaica (Siberian tiger)	XP_007090678
Myotis lucifugus (Little brown bat)	XP_006104717
Vicugna pacos (Alpaca)	XP_006215371
Sus scrofa (Wild pig)	NP_001106531
Bos Taurus (Cow)	NP_777005
Balaenoptera acutorostrata scammoni (Minke whale)	XP_007194049
Lipotes vexillifer (Baiji-Chinese river dolphin)	XP_007449241
Sorex araneus (Common shrew)	XP_004619648
Erinaceus europaeus (European hedgehog)	XP_007532379
Homo sapiens (Human) isoform1, OCT4A	NP_002692.2
Callithrix jacchus (Common marmoset)	NP_001252513
Tarsius syrichta (Philippine tarsier)	XP_008069928
Otolemur garnettii (Northern greater galago)	XP_003789042
Tupaia chinensis (Treeshrew)	XP_006172307
Mus musculus (Mouse)	ENSMUSG00000024406
Peromyscus maniculatus bairdii (Prairie Deer Mouse)	XP_006997372
Jaculus jaculus (Lesser Egyptian Jerboa)	XP_004671946
Oryctolagus cuniculus (European rabbit)	NP_001093427
Orycteropus afer afer (Aardvark)	XP_007949571
Loxodonta Africana (African bush elephant)	XP_003422494
Dasypus novemcinctus (Nine-banded armadillo)	XP_004457909
Macropus eugenii (Tammar wallaby)	FJ998419
Ornithorhynchus anatinus (Platypus)	NP_001229656

Table 2.3 Sources for POU5F3 protein sequences

Species	Genbank ID
Latimeria chalumnae (Coelacanth)	coelacanth POU5F3, JH126767 (739970-739293, 735489-735378, 725094-724964, 718795-718631, 713786-713511)
Acipenser sinensis (Sturgeon)	JN099311
Lepisosteus oculatus (Spotted gar)	XP_006640642.1
Labeo rohita (Rohu)	ADC96616.1
Danio rerio (Zebrafish)	NM_131112
Astyanax mexicanus (Blind cave tetra)	XP_007230899.1
Cynoglossus semilaevis (Tongue sole)	XP_008322749.1
Oryzias latipes (Medaka fish; Japanese rice fish)	NP_001098339.1
Poecilla reticulate (Rainbow fish)	XP_008422195.1
Xiphophorus maculatus (Platyfish)	XP_005799711.1
Oreochromis niloticus (Nile tilapia)	XP_005451421.1
Neolamprologus brichardi (Cichlid)	XP_006788794.1
Larimichthys crocea (Yellow croaker)	NP_001290294.1
Stegastes partitus (Damselfish)	XP_008302560.1
Takifugu rubripes (Japanese puffer)	XP_003965650.1
Xenopus laevis (African clawed frog), Xlpou91	NP_001081342.1
Xenopus laevis (African clawed frog), Xlpou25	ABH07383
Ambystoma mexicanum (Axolotl or Mexican salamander)	KF020689
Chrysemys picta bellii (Painted turtle)	AHGY01090879 (23030-23629, 30326-30440, 31140-31270), AHGY01090878 (4146-4310), AHGY01090877 (649-903)
Alligator sinensis (Alligator)	XP_006022783
Gallus gallus (Chicken)	NP_001103648.1
Melopsittacus undulates (Budgie)	XP_005143614.1
Macropus eugenii (tammar wallaby)	FJ998420

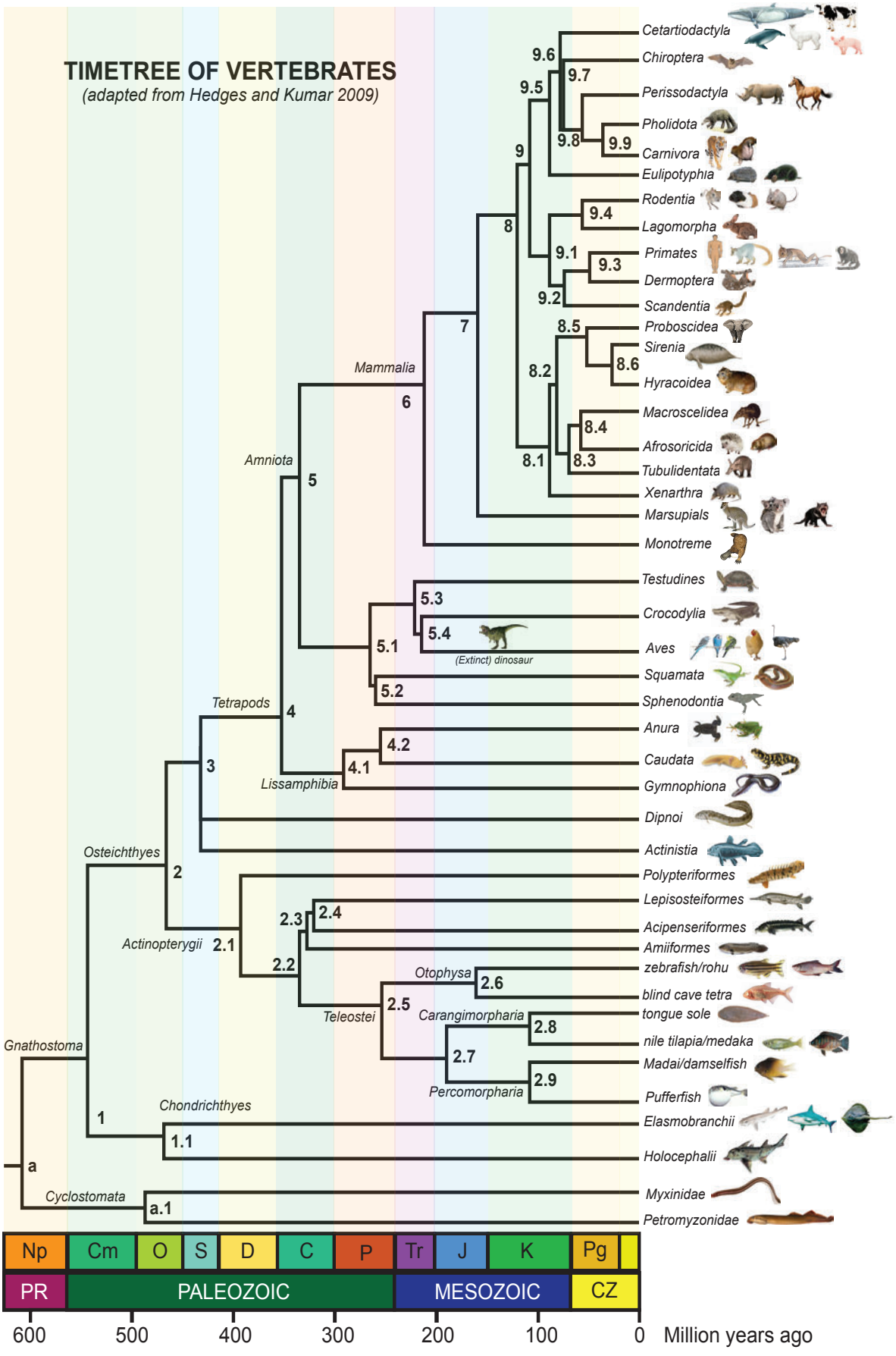
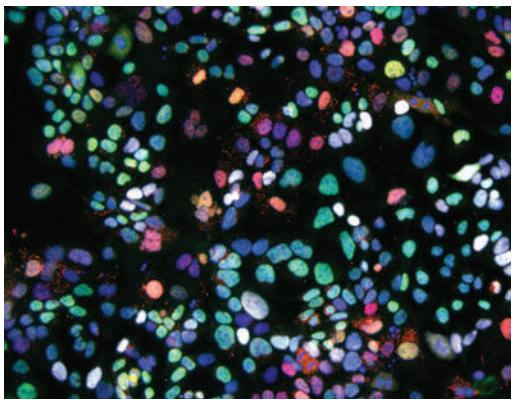


Figure 2.6 Timetree of vertebrates The figure is adapted from Hedges and Kamar 2009. The numbers in each node are refered to divergence time in table 2.4 Abbreviations: C (Carboniferous), Cm (Cambrian), CZ (Cenozoic), D (Devonian), J (Jurassic), K (Cretaceous), Np (Neoproterozoic), O (Ordovician), P (Permian), Pg (Paleogene), PR (Proterozoic), S (Silurian), and Tr (Triassic).

NODE	Divergence of	Time/CL *Reference 1*		Time/CL *Reference 2*		Time/CL *Reference 3*		Time/CL *used in BEAST*
a	Gnathostome-Cyclostome	652/742-605	*(1)*	564/710-418	*(8)*	N/A		not included
a.1	Hagfish-lamprey	520/596-461	*(1)*	432/473-391	*(9)*	478/497-459	*(9)*	not included
1	Chondrichthyes-Osteichthyes	525/580-494	*(1)*	528/639-417	*(8)*	N/A		525/580-494
2	Actinopterygii-Sarcopterygii	476/494-442	*(1)*	450/520-380	*(8)*	N/A		476/494-442
2.1	Polypteriformes-Percomorpharia	407/376-446	*(5)*	N/A		N/A		not included
2.2	Acipenseriformes-Percomorpharia	343/310-381	*(5)*	372/347-391	*(4)*	N/A		343/310-381
2.3	Acipenseriformes-Amiiformes	327/295-366	*(5)*	N/A		N/A		not included
2.4	Acipenseriformes-Lepisosteiformes	312/279-351	*(5)*	N/A		N/A		312/279-351
2.5	Otophysa-Percomorpharia	307/371-215	*(5)*	230/264-200	*(10)*	278/314-241	*(10)*	278/314-241
2.6	Zebrafish-Blind cave tetra	170.5/153.1-187.9	*(13)*	N/A		N/A		170.5/153.1-187.9
2.7	Carangiomorpharia-Percomorpharia	114.5/105.7-125.1	*(13)*	N/A		N/A		114.5/105.7-125.1
2.8	Tongue sole-Nile tilapia	N/A		N/A		N/A		*Prior default*
2.9	Madai-Pufferfish	N/A		N/A		N/A		*Prior default*
3	Actinistia/Dipnoi-Tetrapods	430/438-421	*(1)*	N/A		N/A		430/438-421
4	Lissamphibia-Amniota	360/389-331	*(8)*	360/373-346	*(20)*	N/A		360/373-346
4.1	Caecilian-frog/axolotl	282/356-250	*(11)*	337/353-321	*(20)*	294/319-271	*(19)*	not included
4.2	frog-axolotl	308/328-289	*(20)*	254/257-246	*(11)*	264/276-255	*(19)*	264/276-255
5	Aves/Reptiles-Mammalia	323/343-305	*(16)*	326/354-311	*(1)*	N/A		326/354-311
5.1	Testudine-Squamata	276/383-169	*(8)*	285/296-274	*(3)*	289/302-276	*(3)*	276/383-169
5.2	Squamata-Sphenodontia	268/278-256	*(3)*	275/292-258	*(3)*	N/A		not included
5.3	Testudine-Aves	225/238-205	*(14)*	265/278-252	*(3)*	273/291-255	*(3)*	225/238-205
5.4	Crocodylia-Aves	222/325-119	*(8)*	214/259-169	*(10)*	259/282-236	*(20)*	222/325-119
6	Protherian-therian	213/262-164	*(10)*	204/218-190	*(3)*	227/251-203	*(3)*	213/262-164
7	Marsupial-Eutherian	192/238-146	*(17)*	178.5/192-166	*(3)*	173.0/197-149	*(5)*	178.5/192-166
8	Atlantogenata-Boreoeutheria	104.7/115-96	*(12)*	105.0/118-92	*(8)*	104.5/122-90	*(18)*	104.7/115-96
8.1	Xenathra-Proboscidea	103.3/114-95	*(12)*	N/A		N/A		103.3/114-95
8.2	Tubulidentata-Proboscidea	77.8/86-70	*(12)*	89.0/104-75	*(18)*	N/A		77.8/86-70
8.3	Tubulidentata-Afrosoricida	75.1/83-67	*(12)*	80/96-67	*(18)*	N/A		75.1/83-67
8.4	Macroscelidea-Afrosoricida	73.0/81-65	*(12)*	85/100-71	*(18)*	N/A		73.0/81-65
8.5	Proboscidea-Sirenia	64.2/73-56	*(12)*	79.0/94-65	*(18)*	N/A		64.2/73-56
8.6	Sirenia-Hyraxcoidea	61.1/70-53	*(12)*	N/A		N/A		61.1/70-53
9	Euarchontoglires-Laurasiatheria	97.4/106-90	*(12)*	92.0/95-89	*(8)*	92.5/105-81	*(18)*	97.4/106-90
9.1	Rodentia-Primate	91.0/99-84	*(12)*	90.8/95-87	*(8)*	85.5/98-76	*(18)*	91.0/99-84
9.2	Primate-Scandentia	89.1/97-82	*(12)*	85.9/109-63	*(8)*	78.0/88-70	*(18)*	89.1/97-82
9.3	Primate-Dermoptera	86.2/71-98	*(12)*	N/A		N/A		86.2/71-98
9.4	Rodentia-Lagomorpha	86.4/94-80	*(12)*	74.0/89-62	*(18)*	85.7/71-98	*(7)*	86.4/94-80
9.5	Eulipotyphra-Chiroptera	87.2/93-82	*(12)*	81.5/91-74	*(18)*	N/A		87.2/93-82
9.6	Cetartiodactyla-Chiroptera	84.6/80-90	*(12)*	83.0/91-75	*(8)*	78.0/85-72	*(18)*	84.6/80-90
9.7	Carnivora-Chiroptera	84.2/89-79	*(12)*	74.0/85-63	*(8)*	N/A		84.2/89-79
9.8	Perrisodactyla-Carnivora	82.5/87-78	*(12)*	N/A		N/A		82.5/87-78
9.9	Pholidota-Carnivora	79.8/85-75	*(12)*	74.0/81-67	*(18)*	N/A		not included

Time = time estimate of divergence event/ CL = 95% confidential Interval

Table 2.4 Divergence times of vertebrate lineages in million of years for BEAST analysis to estimate the rate of POUV protein evolution The nodes are referred to figure 2.6 timetree of vertebrates (adapted from Hedges & Kumar 2009). The second column is the description of lineage divergence, for example "Gnathostome-Cyclostome" means the divergence of Cyclostome lineage from Gnathostome ancestor occured around 652 million years ago. The third-fifth columns are time estimates of divergence event in million year unit and CL is the 95% confidential interval. The last column contains time estimates I used for BEAST analysis of POUV protein evolution. N/A stands for "not available age estimates". "Not included" means that there is no POUV protein/DNA sequence available for that specific lineages. References in brackets (1) Blair and Hedges 2005 (2) Hedges and Kumar 2009 (3) Hugall et al., 2007 (4) Hurley et al., 2007 (5) Ionue et al., 2003 (6) Ionue et al., 2005 (7) Janecka et al., 2007 (8) Kumar and Hedges 1998 (9) Kuraku and Kuratani 2006 (10) Mannen and Li 1999 (11) Marjanovic and Laurin 2007 (12) Murphy et al., 2007 (13) Near et al., 2012 (14) Paton et al., 2002 (15) Peng et al., 2006 (16) Pereira 2006 (17) van Rheede 2005 (18) Woodburne et al., 2003 (19) Zhang et al., 2008 (20) Zhang et al., 2005



CHAPTER 3

THE EVOLUTION OF CLASS V POU PROTEINS IN REGULATING PLURIPOTENCY

CHAPTER 3

THE EVOLUTION OF CLASS V POU PROTEINS IN REGULATING PLURIPOTENCY

Introduction

Pluripotency is the ability of cells to give rise to all somatic lineages including germ cells (the origin of sperm and oocyte). In mouse embryo, inner cell mass (ICM), epiblast, and primodial germ cells are pluripotent cells, they can be isolated and maintained *in vitro* as pluripotent stem cells called embryonic stem cells (ESCs), epiblast stem cells (EpiSCs), and embryonic germ cells (EGCs) respectively. Pluripotent stem cells can also be generated via the reprogramming of adult somatic cells through the exogenous expression of transcription factors normally expressed in ESCs, to make induced pluripotent cells (iPSCs).

Pluripotency is regulated through a series of extrinsic signals that input into a complex network of transcription factors. A class V (POUV) transcription factor called Oct3/4 (here referred to Oct4) is an essential component of this network. The removal of Oct4 from the mouse blastocyst or the ESCs resulted in a failure to maintain pluripotent cell types and differentiation toward an extraembryonic tissue called trophectoderm, which is the origin of placenta (Nichols et al., 1998). Oct4 also regulates differentiation in dose-dependent manner such that either up- or down-regulation of Oct4 can cause differentiation and Oct4 is therefore both a mediator of self-renewal and differentiation (Niwa et al., 2000). ESC-expression levels of Oct4 are required for both the maintenance of ESC self-renewal and efficient differentiation into all lineages (Radzisheuskaya et al., 2013). Oct4 is also a central player in reprogramming to generate iPSCs (Takahashi and Yamanaka, 2006). Moreover, while different cocktails of transcription factors can reprogram somatic cells to

iPSCs, Oct4 is required for most combinations (Kim et al., 2009; Li et al., 2011; Radzisheuskaya and Silva, 2013; Yuan et al., 2011).

Recent advances in genome assembly technologies have enabled the completion of a vertebrate genome sequence database that includes almost all branches of vertebrate taxa. So far, more than 100 potential *POUV* gene sequences have become available in a range of vertebrate genomes, providing important information to help understand how this class of POU proteins originated and evolved. What are the origins of the regulatory program driven by this master stem cell factor? To what extent are different POUV activities conserved functionally? Based on several pieces of evidence including conserved protein region called POU domains, exon-intron structure, and the conserved synteny, jawed vertebrate *POUV* genes likely share a common ancestor; hence it is logical to call all *POUV* genes in several jawed vertebrate taxa together as *POUV*/*Oct4* homologs**.**

The conserved synteny of *POUV*, which is preserved co-localization of *POUV*-flanking genes in different species, leads to further classification of *POUV* genes into two subfamilies named *POU5F1* and *POU2*/*POU5F3* (In my thesis I refer to only POU5F3). It is believed that both classes of *POUV* originated from a single *POUV* gene through genome duplication, which occurred before the divergence of extant cartilaginous fish and bony fish. Hence, all vertebrate *POU5F1* genes can be called technically together as *POU5F1* orthologs while all *POU5F3* genes as *POU5F3* orthologs. Based on the assumption that these two genes originated after duplication from a single gene, *POU5F1* can be called as a paralog of *POU5F3* (or *POU5F3* paralog) and vice versa (Frankenberg and Renfree, 2013).

A significant number of extant vertebrate species have lost either *POU5F1* or *POU5F3* gene. The vertebrate taxa retaining only *POU5F3* include cartilaginous fish, actinopterygians (e.g. sturgeon, zebrafish, medaka), birds, and crocodiles. The vertebrate

taxa retaining only *POU5F1* include lizards, snakes, eutherians (e.g. mouse, human, elephant). Based on available expression profiles of POUV in some species, the only remaining POUV protein (either POU5F1 or POU5F3) is generally found and play roles in both epiblast and germ cells. In eutherian mammals, POU5F1 is additionally found in ICM, an evolutionary novel structure of that enables the early formation of both embryonic and extra-embryonic lineages (Nichols et al., 1998). It is noteworthy that germ cells/EGCs and ICM/ESCs share lot of gene expression and epigenetic status in common (Leitch et al., 2013). Hence, this suggests that POUV protein activities can basically be categorised into two subfunctions: epiblast like activity and germ cell/ICM like activity. Interestingly, some species including some cartilaginous fish, coelacanths, axolotls, turtles, monotremes (e.g. platypus) and marsupials (e.g. tammar wallaby) retain both classes of *POUV* gene. Following the trend of some other duplicated genes, POUV subfunctions in these species seem to segregate into different POUV proteins, in that POU5F1 regulates pluripotency related to germ cells, while POU5F3 plays roles in the epiblast during gastrulation, as suggested by expression profile of both POUV proteins in tammar wallaby (Frankenberg et al., 2013). How can this functional distinction between POU5F1 and POU5F3 be verified? How is this distinction related to cell potency, early development and germ cell specification? Do those species carrying both POUV forms show this functional segregation? These are unanswered questions I investigated in this chapter.

At the beginning (section 3.1), I performed *in silico* analysis of vertebrate POUV proteins to understand the evolutionary relationships between POUV orthologs. Integrative knowledge of fossil records and molecular analysis enables many research groups to identify the divergence time among vertebrate taxa. Based on these divergence times and our combined POUV protein sequences from available online database, I illustrate how POUV proteins have evolved in each vertebrate lineage. Next, I investigated their biological functions to address the hypothesis that POU5F1 maintains pluripotency in early

embryo/ESC and germ lines while POU5F3 play roles during gastrulation. I replaced the endogenous Oct4 in mouse ESCs with either *POU5F1* or *POU5F3* genes from different species and analyzed the ability of these POUVs to regulate pluripotency and differentiation. I found that POU5F1 proteins have a greater capacity to maintain ESC self-renewal than POU5F3 proteins. Moreover, germ cell/ESC markers were consistently expressed at higher levels in cells rescued by POU5F1 proteins, while POU5F3-rescued ESC lines showed a higher degree of differentiation. The most striking difference in the ESC rescue activity of different POUV paralogs was observed between coelacanth POU5F1 and POU5F3 (section 3.2-3.3). Global gene expression profiling of coelacanth POUV-rescued lines provided evidences that POU5F1 has conserved roles in regulating germ line level pluripotency while POU5F3 regulates more epiblast-related developmental programs including cell differentiation and cell adhesion (section 3.4-3.7).

Lastly, I aimed to find the origin of *POU5F1* and *POU5F3* genes. As it is believed that genome duplication took place before gnathostome (jawed vertebrate) ancestor originated (Donoghue and Keating, 2014), the single *POUV* gene might present in extant vertebrates whose ancestor diverged before jawed vertebrates. The only closest extant vertebrates to gnathostomes belong to Cyclostomata including lamprey and hagfish. In the last section 3.8, I describe a *POUV* gene in the arctic/Japanese lamprey and show that also in this species the *POUV* gene is structured into 5 protein-coding exons. Interestingly, there is only one *POUV* gene in lamprey, which support that this might represent a *bona fide* ancestral gene of both *POU5F1* and *POU5F3.*

Section 3.1 Evolutionary Model of POU5F1 and POU5F3 proteins

Several studies have conveyed an in-depth phylogenetic analysis of a conserved protein region called POU domain. Based on these analyses, several proteins containing POU domains were identified across animal kingdom and categorised into the class I to class VI (Gold et al., 2014). Oct4 and its homologs (both POU5F1 and POU5F3) are classified into the same major group of class V POU proteins. In POUV protein, there are very divergent N- and C- terminal protein domains (NTD, CTD respectively) that flank the conserved domain. Inside the POU domains, there are two distinct conserved sequence regions: the first part joined to the NTD is the POU specific domain (POU_S), and the second part joined to the CTD is the POU homeodomain (POU_{HD}). These two POU subdomains are joined by a highly diverged protein linker sequence (Figure 3.1-3.4).

Based on the already characterised distinction between POU5F1 and POU5F3, all gnathostome POUV proteins were assumed to be either POU5F1 or POU5F3-like protein. However, in the NCBI database, most *POU5F3* genes were misidentified as *POU5F1* or *POU5F1*-like genes. Here I was be able to gather all vertebrate POUV genes/proteins available online and categorised them into the right subfamilies for further analysis. Originally, global phylogenetic analysis of both POU5F1 and POU5F3 was not enough to distinguish all POUV proteins into two distinct subclasses because some POUV proteins extremely diverged, in particular Xenopus POUV proteins and this led to the misconception that all vertebrate POUV genes were orthologous. Frankenberg and Renfree, 2013 has gathered both syntenic analysis and protein alignment sto identify unique amino acid composition of vertebrate POUV protein sequences, enabling the easier way to classify them into two subclasses and confirming that POU5F1 is a paralog of POU5F3. These unique sequences identified to distinguish POU5F1 from POU5F3 protein, include the region

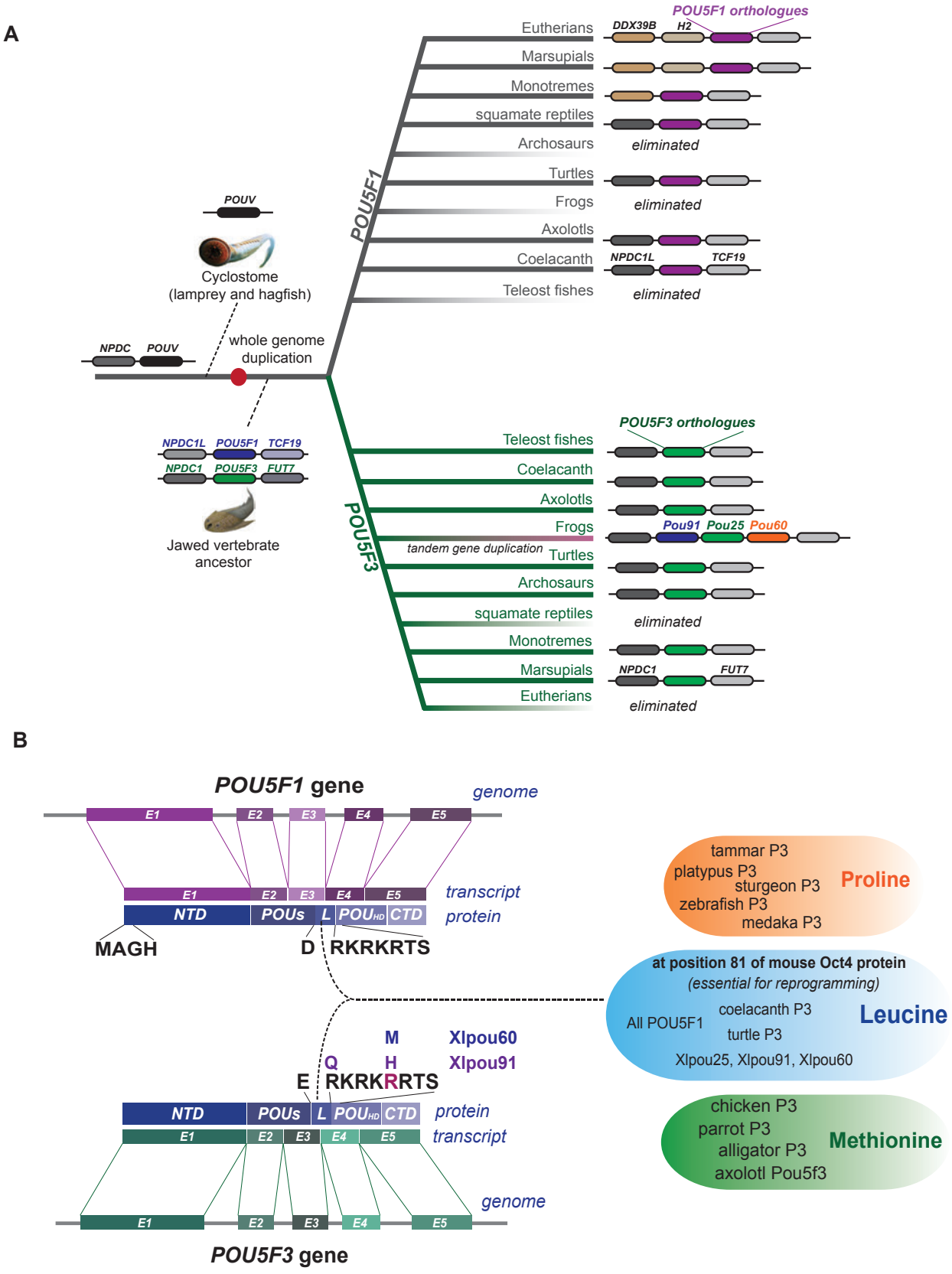


Figure 3.1 Details of POUV homologues and unique protein sequences of POU5F1 and POU5F3. A) The simplified vertebrate tree depicts the evolution of duplicated *POUV* genes (*POU5F1* and *POU5F3*) from a single *POUV* gene. Transparent lines in some branches of vertebrates illustrate that the elimination of *POUV* gene during evolution and marked as "eliminated". The cartoons on the right side of the tree indicate the synteny of *POUV* genes marked in purple color for *POU5F1* and green color for *POU5F3*. In frog lineage, tandem duplicated gene *Pou91* (*Pou5f3.1*) and *Pou60* (*Pou5f3.3*) are marked with purple and orange boxes indicating their predicted activities as POU5F1-like activity and diverged activity, respectively. The color boxes next to *POUV* genes mark the flanking genes which are used to confirm the identity of POUV homologues. B) Schematic illustration shows exon-intron structure of *POUV* genes and the corresponded protein domains. The letters of POUV protein cartoon indicate amino acids that are unique to either POU5F1 or POU5F3 protein or both. There are more consensus sequences common in both POUV proteins and those are shown in the alignments of POUV proteins in figure 3.2-3.4. Abbreviations; P1, POU5F1; P3, POU5F3; E1-E5, exon1-exon5; P, Proline; M, Methionine; R, Arginine; T, Threonine; S, Serine; Q, Glutamine; H, Histidine; E,Glutamic acid; D, Aspartic acid; G, Glycine; K, Lysine.

around the nuclear localization signal (NLS), which is located on N-terminal site of the POU_{HD} and an amino acid adjacent to the N-terminal site of the linker region. The consensus sequence around the NLS for POU5F1 proteins is RKRKRTS, while the consensus sequence for POU5F3 proteins has one additional arginine residue (R) (RKRKR**R**TS). They also feature an aspartic acid (D) before the linker in POU5F1 that is always a glutamic acid (E) in POU5F3 (both amino acids are negatively charged), see figure 3.4. In addition, there is a small unique consensus sequence MAGH that can be found only in POU5F1 proteins (Figure 3.2).

I was able to find POUV proteins in almost all branches of vertebrate taxa, which enabled several phylogenetic analysis and evolutionary modeling to elucidate how these POUV proteins are related. Here I built a specific model for the change in both POU5F1 and POU5F3 proteins during the vertebrate evolution. I combined the knowledge of time-constructed vertebrate phylogeny and POUV protein sequences to estimate when in vertebrate evolution specific changes occurred and which branches of vertebrate taxa contained the highest changes in amino acid composition for either POU5F1 or POU5F3 protein. The changes in amino acid composition relative to the estimated time of each vertebrate divergence can infer to the rate of protein evolution.

I used the BEAST software package, which implements a family of Markov Chain Monte Carlo (MCMC) algorithms for Bayesian phylogenetic inference, divergence time dating and other related molecular evolutionary analyses (Drummond et al., 2012). The full methods and parameters used for BEAST analysis are described in chapter 2 section 2.9. Briefly, we used BEAST to estimate the rate of protein evolution by gathering two parameters: conserved sequences of POU5F1 or POU5F3 proteins (POU domains) sampled from several vertebrate lineages (NCBI database, GENBANK detail is listed in chapter 2) and known calibration dates based on molecular data and fossil records through literature

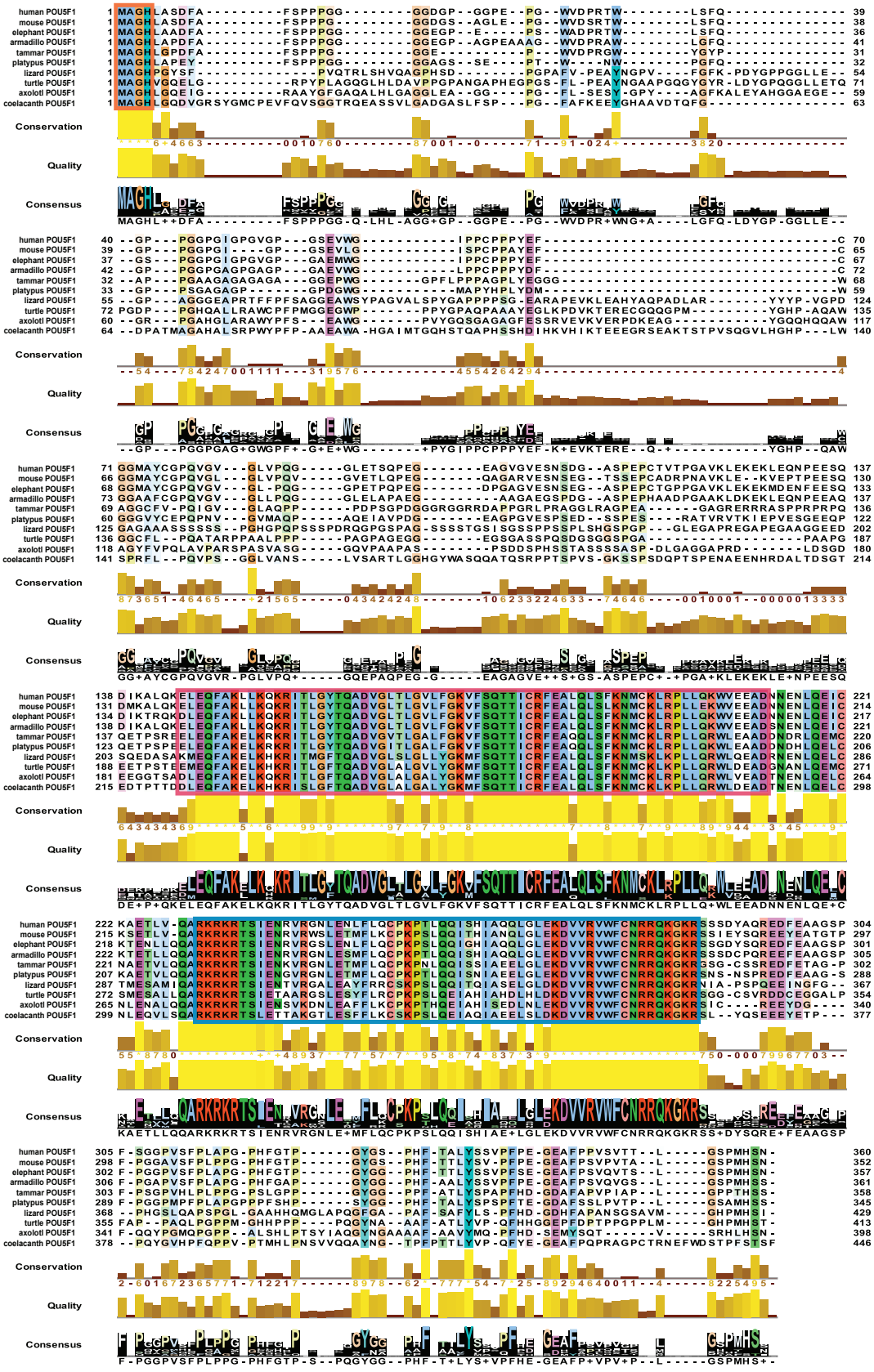


Figure 3.2 Protein sequence alignments of selected vertebrate POU5F1 orthologues
The protein sequences were aligned using MUSCLE in Jalview. The colorscheme at the sequences is based on CLUSTAL X. The intensity of color is scaled by the degree of amino acid property conservation. Histograms under the sequence alignments represent amount of conserved amino acid composition with consensus sequence logo. The red box indicates conserved POU specific domain, the blue box indicates conserved POU homeodomain and the orange box indicates MAGH consensus sequence at the N-terminal domain.

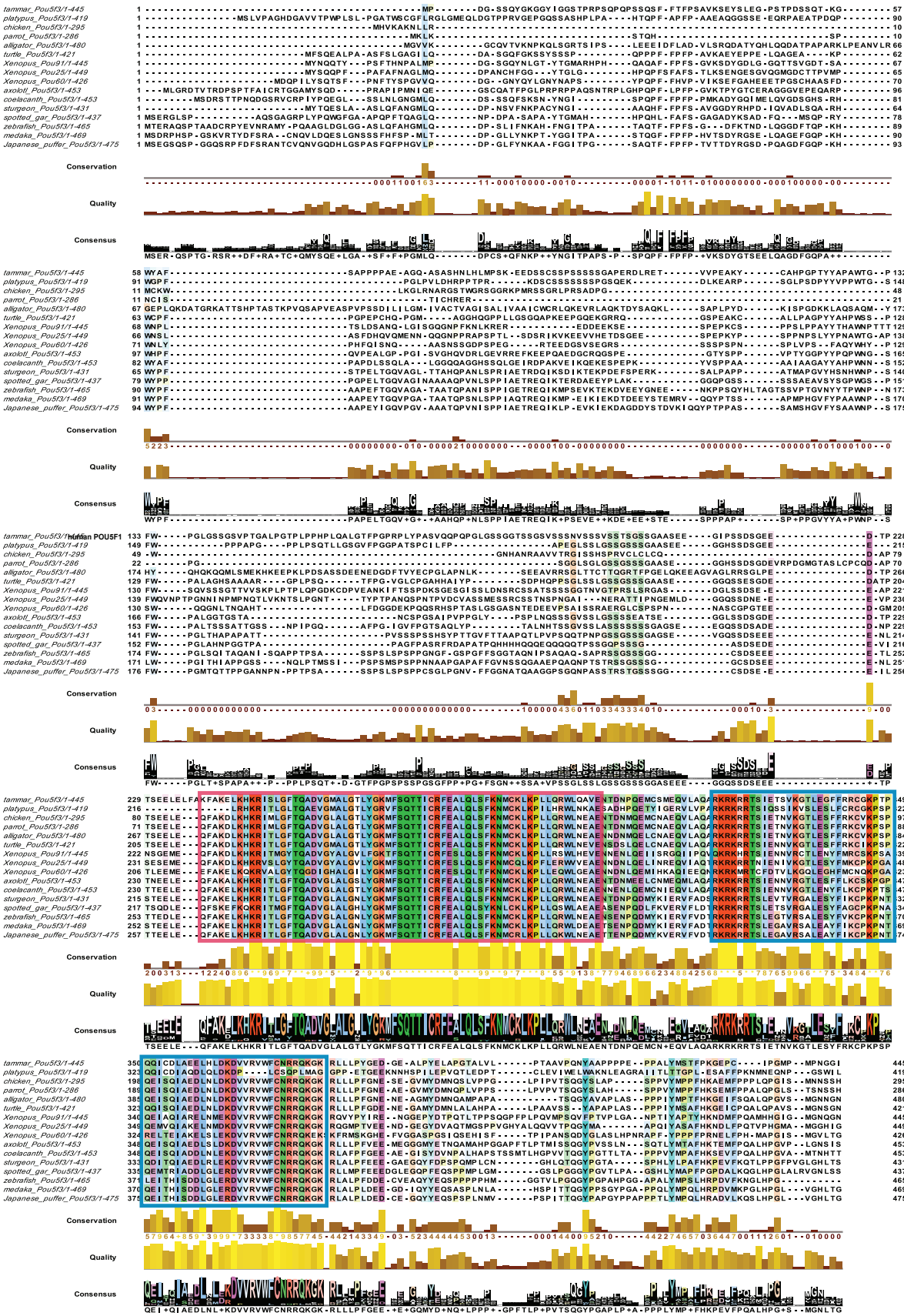


Figure 3.3 Protein sequence alignments of selected vertebrate POU5F3 orthologues

The protein sequences were aligned using MUSCLE in Jalview. The colorscheme at the protein sequences is based on CLUSTAL X. The intensity of color is scaled by the degree of amino acid property conservation. Histograms under the sequence alignments represent amount of conserved amino acid composition with consensus sequence logo. The red box indicates conserved POU specific domain and the blue box indicates conserved POU homeodomain.

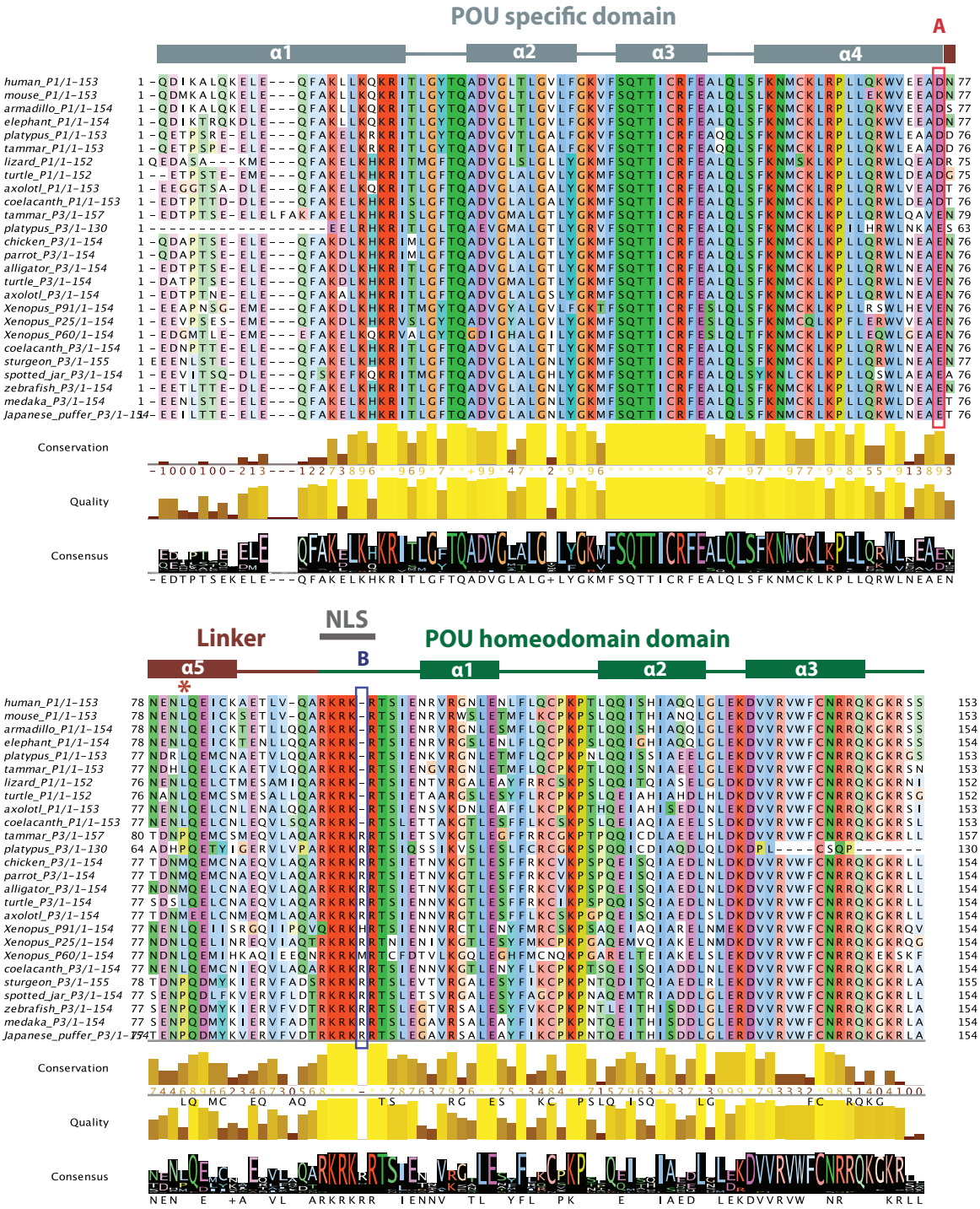


Figure 3.4 Protein sequence alignments between POU domains of selected vertebrate POU5F1 and POU5F3 orthologues

The protein sequences were aligned using MUSCLE in Jalview. The colorscheme at the sequences is based on CLUSTAL X. Histograms under the sequence alignments represent amount of conserved amino acid composition with consensus sequence logo. α helices define the protein tertiary structure corresponding to the underlying sequences. A and B indicate the unique signatures between POU5F1 and POU5F3. NLS indicates nuclear localization signal (RKRKR). Asterisk (*) indicates Leusine which is the key amino acid on linker required for murine reprogramming, as reported by Esch D et al 2013.

review (summarised in figure 3.5 and table 3.1). Calibration dates indicate the estimated timing of an evolutionary divergence event. BEAST analysis sets these calibration dates and DNA/protein sequences input as parameters to infer rates of protein changes or vice versa.

Figure 3.6 shows the result of BEAST analysis on POU5F1 protein evolution. Based on this analysis there are three major changes of the amino acid composition in POU domains over time: at the divergence of testudines (turtles) from squamata (snake and lizard) around 280 million years ago (MYA), the second one at the divergence of monotreme (e.g. platypus) from therians (marsupial and eutherians) around 200 MYA, and the third one at the divergence of afrotherians (e.g. elephant) from boreoeutherians (e.g. mouse and human) around 110 MYA. The latter shows the highest rate amongst the examined POU5F1 proteins. Both afrotherian and boreoeutherian belong to eutherians, which developed a unique developmental program during early embryonic development, like an inner cell mass (ICM) and an enhanced implantation machinery (Hedges and Kumar, 2009; Warren et al., 2008). The eutherian ICM (as shown mostly in mouse) is maintained by a complex regulatory pluripotency circuit, in which POU5F1 is one of the essential players (Chen et al., 2008b; Chew et al., 2005; Ng and Surani, 2011; Nichols et al., 1998). Highly diverged POU5F1 around 110 MYA might be a part of this key evolutionary novelty. After 110 MYA timepoint, POUV proteins among eutherians stay highly conserved, despite the observation that mammals have undergone extreme species radiation and exhibit morphological changes specific to the environment conditions found at different geographical locations. This preservation indicates that gene regulatory networks involved in the early developmental programs of mammals were unaffected by those environment adaptations.

There are only two eutherian branches showing a high rate of POU5F1 protein evolution, the Rodentia (mouse, rat, hamster) and Chiroptera (bats). All examined rodents including deermouse, Chinese hamster, golden hamster, southern vole and mouse exhibit a

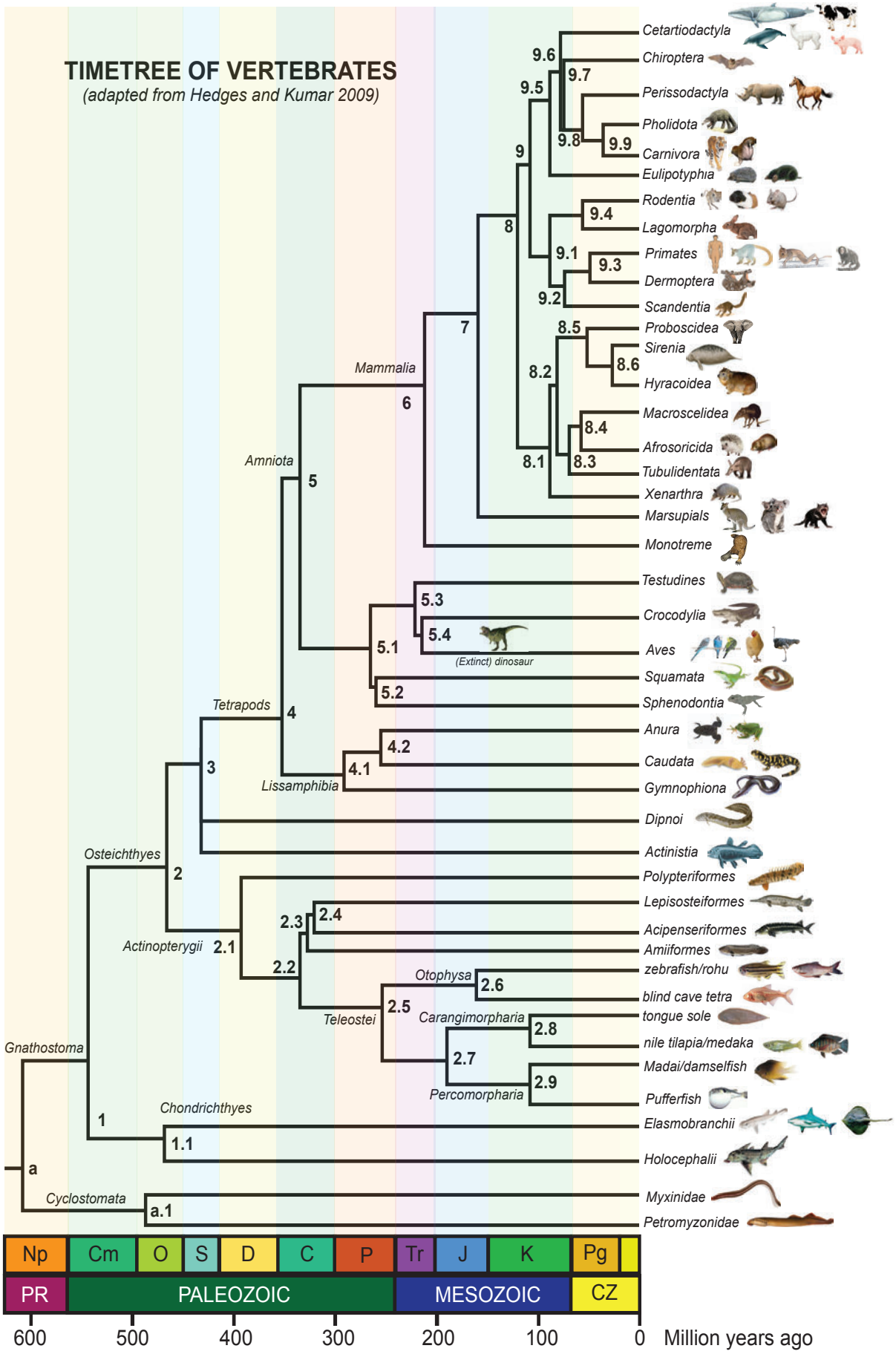


Figure 3.5 Timetree of vertebrates used as a reference for BEAST analysis The figure is adapted from Hedges and Kamar 2009. The numbers in each node are refered to divergence time in table 3.1 Abbreviations: C (Carboniferous), Cm (Cambrian), CZ (Cenozoic), D (Devonian), J (Jurassic), K (Cretaceous), Np (Neoproterozoic), O (Ordovician), P (Permian), Pg (Paleogene), PR (Proterozoic), S (Silurian), and Tr (Triassic).

NODE	Divergence of	Time/CL *Reference 1*		Time/CL *Reference 2*		Time/CL *Reference 3*		Time/CL *used in BEAST*
a	Gnathostome-Cyclostome	652/742-605	*(1)*	564/710-418	*(8)*	N/A		not included
a.1	Hagfish-lamprey	520/596-461	*(1)*	432/473-391	*(9)*	478/497-459	*(9)*	not included
1	Chondrichthyes-Osteichthyes	525/580-494	*(1)*	528/639-417	*(8)*	N/A		525/580-494
2	Actinopterygii-Sarcopterygii	476/494-442	*(1)*	450/520-380	*(8)*	N/A		476/494-442
2.1	Polypteriformes-Percomorpharia	407/376-446	*(5)*	N/A		N/A		not included
2.2	Acipenseriformes-Percomorpharia	343/310-381	*(5)*	372/347-391	*(4)*	N/A		343/310-381
2.3	Acipenseriformes-Amiiformes	327/295-366	*(5)*	N/A		N/A		not included
2.4	Acipenseriformes-Lepisosteiformes	312/279-351	*(5)*	N/A		N/A		312/279-351
2.5	Otophysa-Percomorpharia	307/371-215	*(5)*	230/264-200	*(10)*	278/314-241	*(10)*	278/314-241
2.6	Zebrafish-Blind cave tetra	170.5/153.1-187.9	*(13)*	N/A		N/A		170.5/153.1-187.9
2.7	Carangiomorpharia-Percomorpharia	114.5/105.7-125.1	*(13)*	N/A		N/A		114.5/105.7-125.1
2.8	Tongue sole-Nile tilapia	N/A		N/A		N/A		*Prior default*
2.9	Madai-Pufferfish	N/A		N/A		N/A		*Prior default*
3	Actinistia/Dipnoi-Tetrapods	430/438-421	*(1)*	N/A		N/A		430/438-421
4	Lissamphibia-Amniota	360/389-331	*(8)*	360/373-346	*(20)*	N/A		360/373-346
4.1	Caecilian-frog/axolotl	282/356-250	*(11)*	337/353-321	*(20)*	294/319-271	*(19)*	not included
4.2	frog-axolotl	308/328-289	*(20)*	254/257-246	*(11)*	264/276-255	*(19)*	264/276-255
5	Aves/Reptiles-Mammalia	323/343-305	*(16)*	326/354-311	*(1)*	N/A		326/354-311
5.1	Testudine-Squamata	276/383-169	*(8)*	285/296-274	*(3)*	289/302-276	*(3)*	276/383-169
5.2	Squamata-Sphenodontia	268/278-256	*(3)*	275/292-258	*(3)*	N/A		not included
5.3	Testudine-Aves	225/238-205	*(14)*	265/278-252	*(3)*	273/291-255	*(3)*	225/238-205
5.4	Crocodylia-Aves	222/325-119	*(8)*	214/259-169	*(10)*	259/282-236	*(20)*	222/325-119
6	Protherian-therian	213/262-164	*(10)*	204/218-190	*(3)*	227/251-203	*(3)*	213/262-164
7	Marsupial-Eutherian	192/238-146	*(17)*	178.5/192-166	*(3)*	173.0/197-149	*(5)*	178.5/192-166
8	Atlantogenata-Boreoeutheria	104.7/115-96	*(12)*	105.0/118-92	*(8)*	104.5/122-90	*(18)*	104.7/115-96
8.1	Xenathra-Proboscidea	103.3/114-95	*(12)*	N/A		N/A		103.3/114-95
8.2	Tubulidentata-Proboscidea	77.8/86-70	*(12)*	89.0/104-75	*(18)*	N/A		77.8/86-70
8.3	Tubulidentata-Afrosoricida	75.1/83-67	*(12)*	80/96-67	*(18)*	N/A		75.1/83-67
8.4	Macroscelidea-Afrosoricida	73.0/81-65	*(12)*	85/100-71	*(18)*	N/A		73.0/81-65
8.5	Proboscidea-Sirenia	64.2/73-56	*(12)*	79.0/94-65	*(18)*	N/A		64.2/73-56
8.6	Sirenia-Hyraxcoidea	61.1/70-53	*(12)*	N/A		N/A		61.1/70-53
9	Euarchontoglires-Laurasiatheria	97.4/106-90	*(12)*	92.0/95-89	*(8)*	92.5/105-81	*(18)*	97.4/106-90
9.1	Rodentia-Primate	91.0/99-84	*(12)*	90.8/95-87	*(8)*	85.5/98-76	*(18)*	91.0/99-84
9.2	Primate-Scandentia	89.1/97-82	*(12)*	85.9/109-63	*(8)*	78.0/88-70	*(18)*	89.1/97-82
9.3	Primate-Dermoptera	86.2/71-98	*(12)*	N/A		N/A		86.2/71-98
9.4	Rodentia-Lagomorpha	86.4/94-80	*(12)*	74.0/89-62	*(18)*	85.7/71-98	*(7)*	86.4/94-80
9.5	Eulipotyphra-Chiroptera	87.2/93-82	*(12)*	81.5/91-74	*(18)*	N/A		87.2/93-82
9.6	Cetartiodactyla-Chiroptera	84.6/80-90	*(12)*	83.0/91-75	*(8)*	78.0/85-72	*(18)*	84.6/80-90
9.7	Carnivora-Chiroptera	84.2/89-79	*(12)*	74.0/85-63	*(8)*	N/A		84.2/89-79
9.8	Perrisodactyla-Carnivora	82.5/87-78	*(12)*	N/A		N/A		82.5/87-78
9.9	Pholidota-Carnivora	79.8/85-75	*(12)*	74.0/81-67	*(18)*	N/A		not included

Time = time estimate of divergence event/ CL = 95% confidential Interval

Table 3.1 Divergence times of vertebrate lineages in million of years for BEAST analysis to estimate the rate of POUV protein evolution The nodes are referred to figure 3.5 timetree of vertebrates (adapted from Hedges & Kumar 2009). The second column is the description of lineage divergence, for example "Gnathostome-Cyclostome" means the divergence of Cyclostome lineage from Gnathostome ancestor occured around 652 million years ago. The third-fifth columns are time estimates of divergence event in million year unit and CL is the 95% confidential interval. The last column contains time estimates I used for BEAST analysis of POUV protein evolution. N/A stands for "not available age estimates". "Not included" means that there is no POUV protein/DNA sequence available for that specific lineages. References in brackets (1) Blair and Hedges 2005 (2) Hedges and Kumar 2009 (3) Hugall et al., 2007 (4) Hurley et al., 2007 (5) Ionue et al., 2003 (6) Ionue et al., 2005 (7) Janecka et al., 2007 (8) Kumar and Hedges 1998 (9) Kuraku and Kuratani 2006 (10) Mannen and Li 1999 (11) Marjanovic and Laurin 2007 (12) Murphy et al., 2007 (13) Near et al., 2012 (14) Paton et al., 2002 (15) Peng et al., 2006 (16) Pereira 2006 (17) van Rheede 2005 (18) Woodburne et al., 2003 (19) Zhang et al., 2008 (20) Zhang et al., 2005

high rate of POU5F1 protein changes (Figure 3.6). What were the evolutionary forces leading to this dramatic divergence of rodent POU5F1 proteins? It is well known that rodent gastrulation produces unique cylinder-like shape epiblast while other eutherians have a disc-like epiblast (Sheng, 2014). In addition, it has been shown that Oct4 regulates cell adhesion in ESCs (Livigni et al., 2013). The high changes of Oct4 during Rodentia evolution might affect to cell adhesion property and be responsible for distinctive epiblast morphology.

We also examined the rate of vertebrate POU5F3 protein evolution (Figure 3.7). From this analysis there appear two major changes of POU5F3 proteins: (1) at the origin of actinopterygians including the divergence of chondrostei (sturgeon) and holostei (spotted gar) from other teleost fish in Clupeocephala (e.g. zebrafish and medaka) around 380-280 MYA and (2) along the evolution of the frog lineage around 250-130 MYA. The high rate of fish POU5F3 protein evolution fits well with previous findings that these proteins have both distinct and varied expression patterns among fish and very little potential to sustain and establish eutherian pluripotency (Morrison and Brickman, 2006; Tapia et al., 2012). Although there was a teleost specific genome duplication and rearrangement that occurred around 300 MYA (M et al., 2007), this event may not be responsible for these extreme changes in the POUV family. My observations suggest that the rapid evolution of POUV protein sequence occurred prior to the split between the teleosts and Chondrostei/Holostei around 380 MYA. Future work on testing the biological function of sturgeon Pou5f3 may provide an indication of whether the functional divergence of Pou5f3 observed in the zebrafish is specific to all actinopterygian Pou5f3 or whether only Pou5f3 in Cupleocephala have diverged in their functions.

The second biggest divergence in the POU5F3 subfamily happened during frog evolution. In *Xenopus*, there are three paralogs of *Xenopus* Pou5f3: Xlpou91 (Pou5f3.1), Xlpou25 (Pou5f3.2) and Xlpou60 (Pou5f3.3). We included only Xlpou25 and Xlpou91 in

this analysis, because both these proteins have been shown to have distinct functions and are required for normal development. The function of the third, Xlpou60 is less clear. Based on functional studies, Xlpou25 is an epiblast specific protein, like POU5F3 proteins from other species (discussed in section 3.2-3.7); Xlpou91 is an early marker of primordial germ cells (Hinkley et al., 1992) and has some of the functional attributes of POU5F1 (shown in section 3.2 and (Morrison and Brickman, 2006)). Although Xlpou91 belongs to the POU5F3 subfamily, it has re-acquired some activities normally specific to POU5F1. Thus this completely functional segregation of these two homologs might relate to the high changes in their POU domains sequences.

The software also produced two additional informative data: the Mean Rate of POUV protein evolution and a Coefficient of Variation. The Mean Rate represents the rate of protein evolution averaged over the whole POUV phylogenetic tree. The distribution of POU5F1 and POU5F3 rates around the mean are shown in figures 3.6B, and 3.7B respectively. We found that the POU5F1 subfamily has a slightly higher mean rate of protein evolution than the POU5F3 subfamily. The Coefficient of Variation represents the extent to which the rate of protein evolution varies from lineage to lineage. The Coefficient of Variation is also higher in POU5F1 than in POU5F3, meaning that POU5F1 coding sequence is more heterogeneous amongst different vertebrate lineages. Moreover, the lowest rate in both POU5F1 and POU5F3 protein evolution are those derived from the coelacanth. This result is consistent with the findings that protein-coding genes of coelacanth have not changed much in over 450 million years and its contemporary morphological characters resemble those found in fossils (Amemiya et al., 2013).

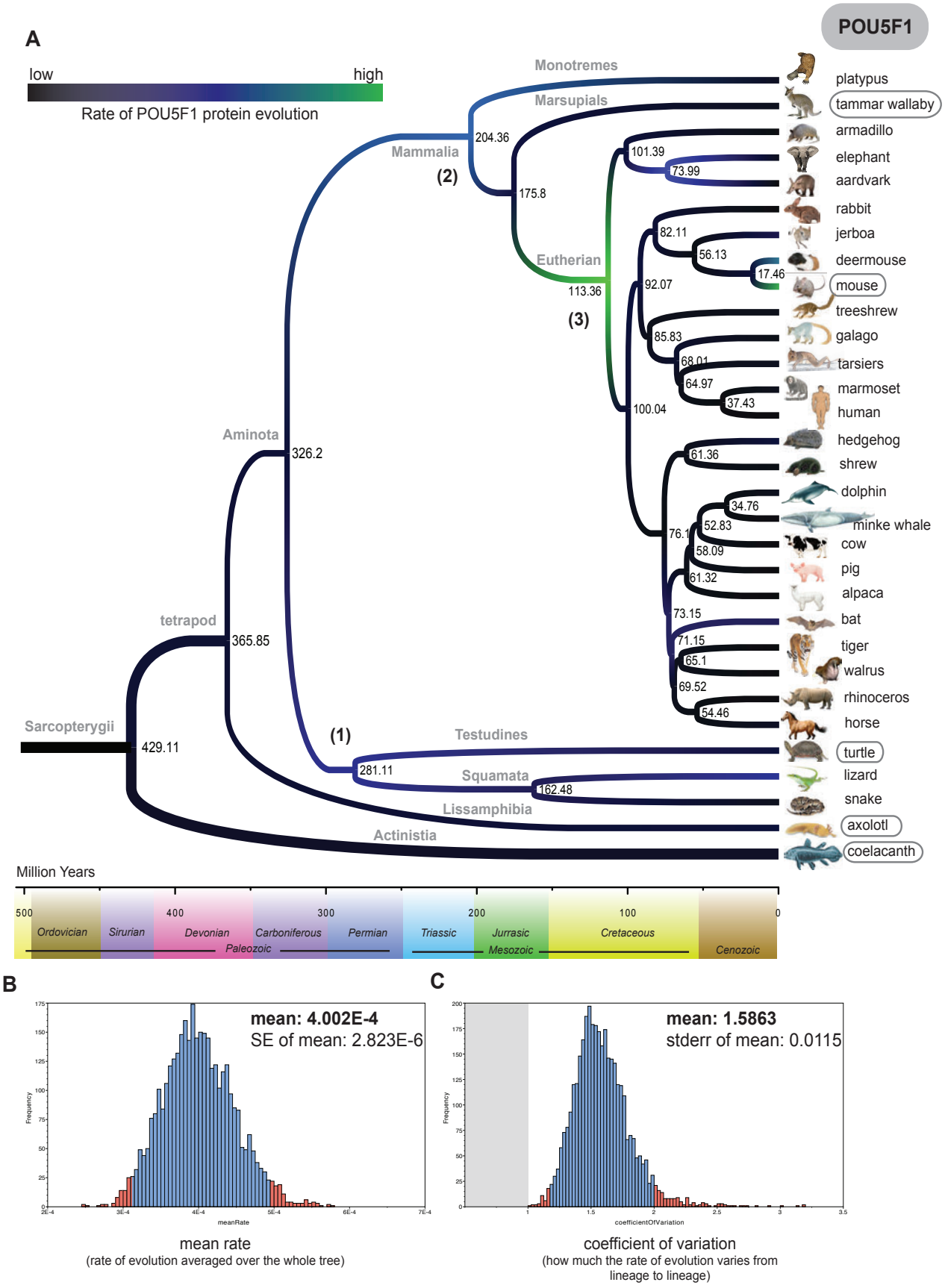


Figure 3.6 Molecular rate of POU5F1 protein evolution The full length POU5F1 protein sequences were firstly MUSCLE-aligned in MEGA6 and only POU specific and POU homeodomain including linker were selected for further analysis. The selected POU sequences were analyzed in BEAUTi/BEAST platform to estimate the rate of protein evolution. The prior setting and calibration node (C)/ divergence times for BEAUTi set up are described in full detail in chapter 2. The result from BEAST were analyzed by TRACER to provide mean rate (B) and coefficient of variation (C). The BEAST result file were then annotated by TREE ANNOTATOR and used to construct the tree as shown in A. The number at each node represents divergence time in miilion years. Boxes at the names of vertebrates indicate the species I investigated their POU5F1 activities in ESC rescue experiment in this study.

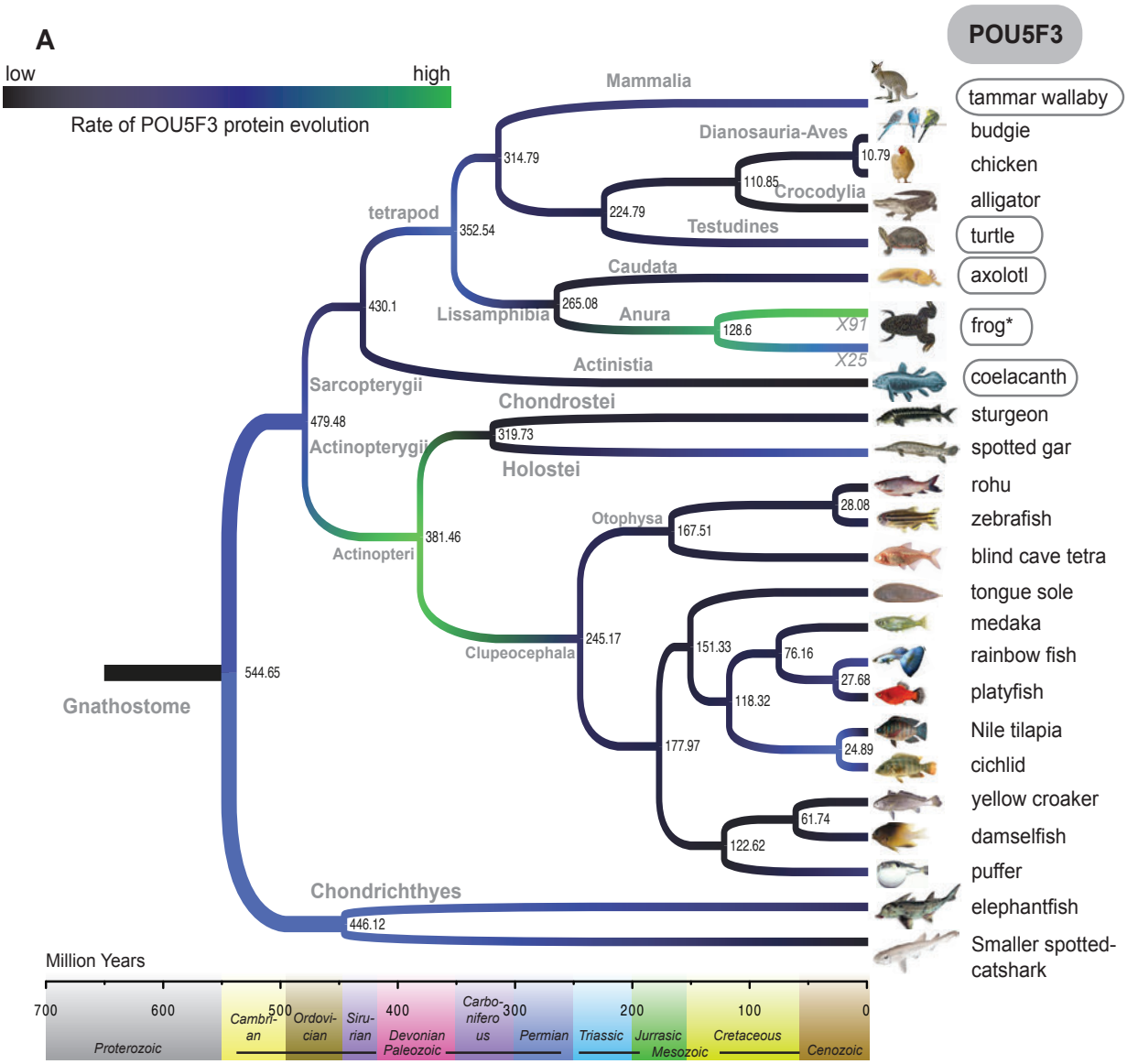


Figure 3.7 Molecular rate of POU5F3 protein evolution The full length POU5F3 protein sequences were firstly MUSCLE-aligned in MEGA6 and only POU specific and POU homeodomain including linker were selected for further analysis. The selected POU sequences were analyzed in BEAUTi/BEAST platform to estimate the rate of protein evolution. The prior setting and calibration node (C)/ divergence times for BEAUTi set up are described in full detail in chapter 2. The result from BEAST were analyzed by TRACER to provide mean rate (B) and coefficient of variation (C). The BEAST result file were then annotated by TREE ANNOTATOR and used to construct the tree as shown in A. The number at each node represents divergence time in miilion year. Asterisk (*) indicates the *Xenopus laevis* protein Xlpou91 (X91) and Xlpou25 (X25). Boxes at the names of vertebrate indicate the species I investigated their POU5F3 activities in ESC rescue experiment in this study.

Section 3.2 The Functional Capacity of POU5F1 and POU5F3 proteins to rescue Oct4 mutant ESCs

Oct4 rescue experiments were performed by using Oct4-suppressable ESC cell line (ZHBTc4). This is an excellent tool to address the conservation of biological functions of POU proteins, especially their abilities to support stem cell self-renewal. In the ZHBTc4 ES cell line, both alleles of endogenous Oct4 were disrupted by homologous recombination and a tetracycline-suppressable Oct4 transgene was stably integrated. In the absence of tetracycline (Tc), Oct4 is expressed from the transgene and the cell line is able to self-renewal and to maintain its ESC phenotype. In contrast, adding Tc blocks Oct4 expression from the transgene, leading to reduction of Oct4 protein within 12 hours and ESC differentiation. Addition of Tc to ZHBTc4 ESCs leads to rapid differentiates toward trophectodermal lineage within a few days. The ability of other POUV proteins to support self-renewal in Oct4 null ESC can therefore be tested by transfecting ZHBTc4 ESCs with a vector expressing different POUV proteins at the same time of adding Tc to remove mouse Oct4 expression.

Our laboratory has previously used this cell line to test the functional conservation of *Xenopus* Oct4 homologs, Xlpou91 (Pou5f3.1), Xlpou25 (Pou5f3.2) and Xlpou60 (Pou5f3.3) (Morrison and Brickman, 2006). Three *Xenopus* POUV genes are believed to originate as a result of tandem gene duplication of single ancestral *pou5f3* gene. Based on the syntenic analysis, three *Xenopus* POUV paralogs are all orthologous to *POU5F3* of other species (Frankenberg and Renfree, 2013). Interestingly, these three paralogs exhibit differences in expression during early *Xenopus* embryogenesis. Xlpou91 is the only POUV protein expressed in the primordial germ cells of *Xenopus* embryo (Hinkley et al., 1992; Venkatarama et al., 2010) while Xlpou25 exhibits the highest expression during gastrulation

(Cao et al., 2008) and Xlpou60 has the highest level of maternal expression in the oocyte. We discovered that these expression profiles are well correlated with their abilities to rescue eutherian pluripotency network supported by Oct4. In particular, Xlpou91 is equivalent to mouse Oct4 itself in its capacity to rescue the null ESCs, whereas Xlpou25 is unable to support large numbers of ESC colonies or maintain clear undifferentiated phenotypes (Morrison and Brickman, 2006). Based on the capacity of Xlpou91 to support ESC self-renewal, we hypothesised that POUV proteins expressed in germ cells might be more effective at supporting ESC self-renewal/pluripotency. This is in part based on the similarities of the germ cell network and ICM/ESC pluripotency network (Leitch et al., 2013).

The discovery of Frankenberg and colleagues also highlights a distinction between germ cell and epiblast expression in different POUV proteins. In this case they characterised the expression of POU5F1 and POU5F3 in tammar wallaby embryos. POU5F1 is expressed specifically in the germ cells while POU5F3 protein is expressed in the nucleus of gastrulation-stage epiblast (Frankenberg et al., 2013). This suggests, that perhaps in species with both paralogs, POU5F1 is a central player in germ cell potency, while POU5F3 protein mainly supports epiblast during gastrulation. Thus, I hypothesised that germ cell-related POU5F1 can support naïve pluripotency while POU5F3 cannot. However, in the species that lost either POU5F1 or POU5F3 protein in their lineage-specific evolution, the remaining POUV protein has to perform both germ cell and epiblast like functions. For example, eutherian mammals lost *POU5F3* gene and therefore Oct4 protein encoded by *Pou5f1* gene is expressed in both germ cells (also ICM) and epiblast. In case of *Xenopus*, its ancestor lost *pou5f1* and remaining ancestral *pou5f3* gene was then duplicated to enable subfunctionalization to be re-established, so that Xlpou91 underwent convergent evolution to specialise in germ cells, and performs Oct4/Pou5f1-like function.

To test this hypothesis and determine the extent to which POU5F1 and POU5F3 protein function in both naïve and primed pluripotency, I examined the activities of POUV in species that carry both paralogs, including coelacanth (*Latimeria chalumnae*), axolotl (*Ambystoma mexicanum*), turtle (*Chrysemys picta bellii*), and tammar wallaby (*Macropus eugenii*). We excluded species that carry only single POUV from these rescue experiments because these proteins might already adapt novel functions to compensate the loss of their paralogs. I included *Xenopus* Oct4 homologs in this functional study, as they have been characterised extensively and serve as an excellent control for the putative segregation of germ cell activity (Xlpou91) and epiblast-like activity (Xlpou25). The coding sequences of *POU5F1* and *POU5F3* from tammar wallaby (Stephen Frankenberg) and axolotl (Tapia et al 2013) were originally amplified from the embryos. The predicted *POU5F1* and *POU5F3* coding sequences of coelacanth and turtle was based on the available genome data and generated by gene synthesis (Invitrogen). All *POUV* coding sequences were introduced into ESC expression vectors employing the CAG promoter and a triple flag-tag sequence (to assess protein expression) and a selection cassette within the *POUV* cistron that employs an Internal Ribosomal Entry Site (*IRES*) to produce the puromycin resistance gene (*PAC*), see figure 2.3B and 3.8A. The details of GENBANK accession numbers for *POUV* coding sequences are described in chapter 2 section 2.1.2.

Rescue vectors were introduced into ZHBTc4 ESCs by electroporation and colonies expanded in the presence and absence of Tc (endogenous Oct4). Colonies were stained for the ESC Alkaline Phosphatase (AP) activity (ESC colonies with undifferentiated cells show red-violet appearance). The procedure for rescue experiment is described in chapter 2 section 2.8 and figure 2.3B. The results of AP staining of POUV-rescued ZHBTc4 colonies are shown in figure 3.8B. In the absence of any POUV protein, ESCs differentiate and do not expand as AP positive colonies. When mouse Oct4 was used to rescue itself, normal AP positive colonies were obtained. When the activity of Oct4 was compared to other POUV

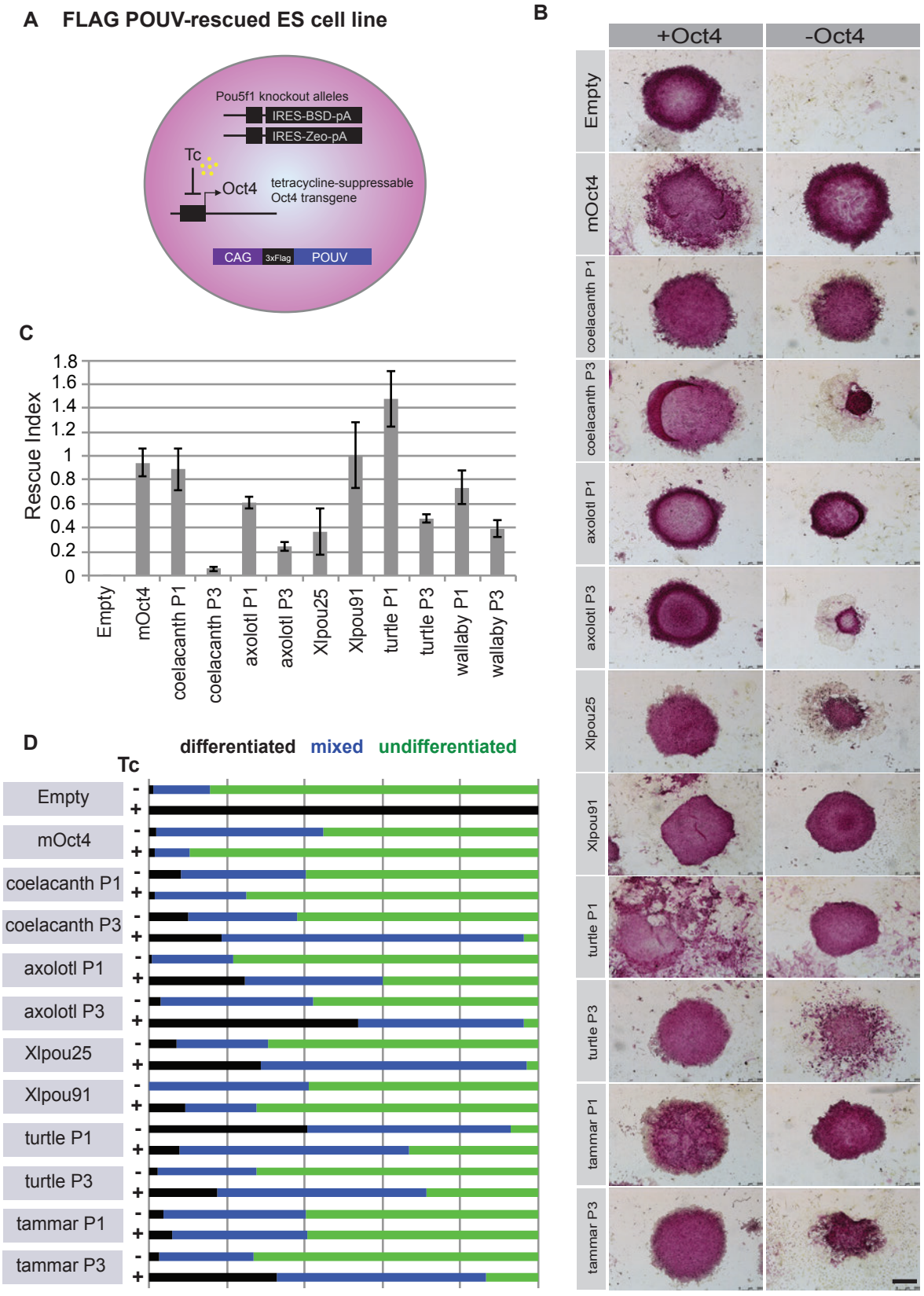


Figure 3.8 Oct4-null ESC rescued by different POU5F1 and POU5F3 proteins

a) Diagram shows rescue of ZHBTc4 Oct4-null ESCs with different 3xflag-tagged POUV proteins. POUV proteins were introduced into these cells by electroporation of plasmid containing POUV cDNA under CAG promoter. The cells were culture under treatment of tetracycline to suppress Oct4 expression from the transgene. B) POUV-transfected Oct4-null ESC cells were at clonal density in the presence of puromycin and presence/absence of tetracycline, and the colonies were stained for alkaline phosphatase (AP; red) after 9 days of growth. Scale bar, 500 μm C) Rescue index indicating the capacity of different Oct4 homologs in support ESC self-renewal. The rescue index is calculated by dividing the number of AP positive ES cell colonies in the absence of Oct4 (presence of tetracycline) by the number of AP positive ESC colonies present in the presence of Oct4 (absence of tetracycline). D) Quantification of clonal growth in ESCs supported by different Oct4 homologs in place of mouse Oct4. Colonies were scored as undifferentiated (green), mixed (blue), and differentiated (black). C)-D) Data represents the means values obtained from three independent experiments. Abbreviation: tc, tetracycline; P1, POU5f1 protein; P3, POU5f3 protein; empty, empty vector.

proteins, I found that all POU5F1 proteins are able to rescue Oct4 null ESCs and the colonies rescued by them show undifferentiated ESC colony morphology. In contrast, all POU5F3 proteins, except Xlpou91, produce morphologically abnormal, although to varying extents AP positive colonies. They all contain some degree of undifferentiated centre surrounded by mixed AP positive/negative differentiated cells. As Oct4 overexpression can induce differentiation, some POU5F1 proteins induce differentiation when expressed in the presence of Oct4 (absence of Tc). In particular the turtle POU5F1 and mOct4-rescued colonies exhibit the most extreme over-expression phenotypes. The differentiated cells from this POUV overexpression are mostly Gata6 positive, defining primitive endoderm lineages. Thus, it suggests that POU5F1 might also share common role in inducing primitive endoderm at higher dose. However, this is out of the scope of this thesis and future careful examination is required to resolve this interesting question.

The extent of ESC rescue can also be quantitated. The rescue index is the ratio of AP positive colonies obtained in the absence of Oct4 to the total colonies obtained in the absence of Tc (presence of Oct4). Oct4 has a rescue index of 1, and all the POU5F1 proteins show a higher rescue index than POU5F3 proteins (Figure 3.8C). The coelacanth POU5F1 has a rescue index indistinguishable from Oct4's, while its paralog was particularly ineffective at supporting self-renewal in the absence of Oct4. To quantitate Oct4 rescue in another way, I categorised the AP positive colonies into three classes: undifferentiated, mixed and differentiated colonies (figure 3.8D). Undifferentiated colonies are AP positive containing less than 5% differentiated cells around periphery of the colony. Mixed colonies exhibit AP positive cells in the centre of the colony, but also contain from 10 to 80% of differentiated cells around periphery. Differentiated colonies show only faint AP staining and consist of at least 90% differentiated cells (criteria for the classification of AP stained ESC colonies is shown in chapter 2 figure 2.2). Consistent with the rescue index, all POU5F1 proteins, except turtle POU5F1, produced high percentages of undifferentiated

colonies, while POU5F3 proteins produced the highest numbers of differentiated colonies. Using this double system of quantitation supports the notion that capacity to support naïve pluripotency and ESC self-renewal is conserved in POU5F1 proteins.

Section 3.3 Phenotypic distinctions of ESCs supported by POU5F1 or POU5F3

The rescue assay showed that all POU5F1 supported naïve pluripotency phenotype while all POU5F3 from different species produced various aberrant ESC colony morphologies. To further explore these differences, I aimed to better characterise the Oct4 null ESC cells supported by different POUV proteins. Firstly, I generated stable ESC lines from these POU5F1 and POU5F3-rescued colonies. At day 9 after electroporation, 24 colonies from each condition were picked and expanded in ESC culture medium containing puromycin and tetracycline (described in chapter 2, section 2.4A2 and 2.7). After several passages, almost all clones of POU5F1-rescued lines self-renewed and expanded better than those supported by POU5F3. In addition, the cell morphology of POU5F1 lines was identical to other ESC lines (e.g. E14Tg2a ESCs, ZHBTc4 ESCs without Tc), while that of POU5F3 lines showed strong differentiation. To better elucidate the phenotypes of stable POUV-rescued lines, three clones of each cell lines where characterised at passage 6. These cell lines were plated at clonal density onto gelatin-coated culture dishes for further characterization by immunofluorescent staining (figure 3.10), qRT-PCR (figure 3.9 and 3.10) and flow cytometry (Figure 3.12).

Before further in-depth phenotypic analysis, I aimed to verify that those rescued lines were solely maintained by transfected *POUV* genes and not influenced by endogenous Oct4 expression. The expression of different POUV proteins in the established cell lines was initially confirmed by immunofluorescence (figure 3.9A). I used flag antibody to detect the presence and localization of 3Xflag POUV proteins. I observed clear nuclear localization from all tested POUV-rescued lines. Remarkably, coelacanth POU5F3 protein was expressed strongest. I further analyzed the expression of *POUV* transcript level by qRT-PCR.

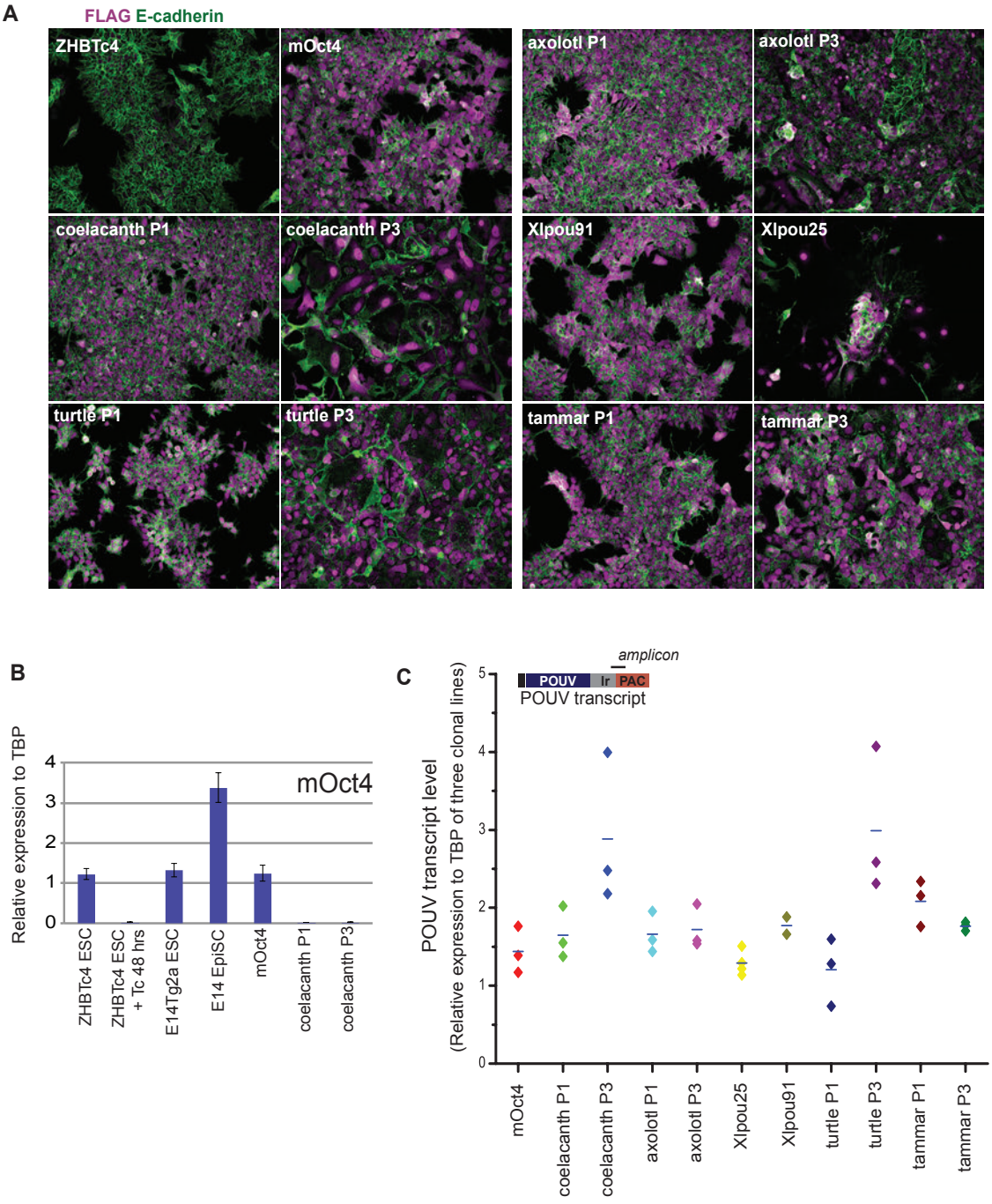


Figure 3.9 The detection of POUV proteins and the measurement of POUV transcipt levels

A) Immunoflorescent staining of POUV-rescued ESC clonal lines with Flag antibody was done to confirms the POUV protein expression and its nuclear localization, and I used E-cadherin antibody to mark individual cells. B) qRT-PCR on the expression of mouse Oct4 was performed to confirm that POUV-rescued ESC clonal lines were maintained solely by transfected POUV constructs. C) qRT-PCR on the expression of puromycin-resistant gene was performed to indirectly measure the POUV transcript levels between different POUV-rescued clonal lines. Data represents in boxplot showing average values from three independent clonal lines. Abbreviations: P1, POU5F1; P3, POU5F3; Tc, tetracycline; Ir, IRES: PAC, puromycin resistance gene.

qRT-PCR employed Oct4 primers that specifically recognised mouse Oct4, but not other *POUV* transcripts. This result confirmed that all established rescued lines were only maintained by transfected *POUV* constructs (figure 3.9B). Because *POUV* gene and *PAC* are expressed as one transcript separated by *IRES*, I could indirectly measure *POUV* transcript level by using primer recognizing *PAC*. I found that majority of clonal lines had similar levels of *POUV* transcripts, except higher level in those rescued by coelacanth POU5F3 and turtle POU5F3 (figure 3.9C). This suggests that POU5F3 from coelacanth and turtle probably require higher dose to partially support ESC potency.

The immunofluorescent staining of some ESC/differentiation markers on the POUV-rescued clonal cell lines is shown in figure 3.10. E-cadherin and p120 catenin were chosen as membrane-associated markers to observe cell morphology. Klf4 marks an undifferentiated naïve ESC population, Cdx2 marks trophectodermal lineage and Gata6 marks primitive endoderm. I found that all POU5F1-rescued Oct4 null ESC lines exhibited ESC cell morphology similar to mOct4-rescued lines, and showed no expression of the differentiation markers Gata6 and Cdx2. However, I did observe heterogeneity in Klf4 expression in coelacanth POU5F1 and turtle POU5F1-rescued lines while those rescued by mOct4 showed uniform expression. This indicates that there may be aspects of the naïve gene regulatory network that are not perfectly conserved. In contrast, all POU5F3-rescued lines exhibited different patterns of cell morphology. Coelacanth and axolotl POU5F3 produced $Cdx2^{+}$ trophectoderm-like cells containing small $Klf4^{+}$ clusters of cells growing on the top of them. Turtle and tammar POU5F3-lines exhibited a mix of Klf4, Cdx2 and Gata6 expressing cells. In ESC supported by Xlpou25, most of clones showed big clumps of cells (ball-shaped) expressing E-cadherin and Klf4 on the top of primitive endoderm-like cell layer expressing Gata6. In contrast, ESCs supported by Xlpou91 were almost identical to POU5F1-rescued lines exhibiting no differentiation. Taken together these experiments suggest that POU5F1-rescues ESC lines are largely undifferentiated, naïve ESCs, while POU5F3-rescued cells

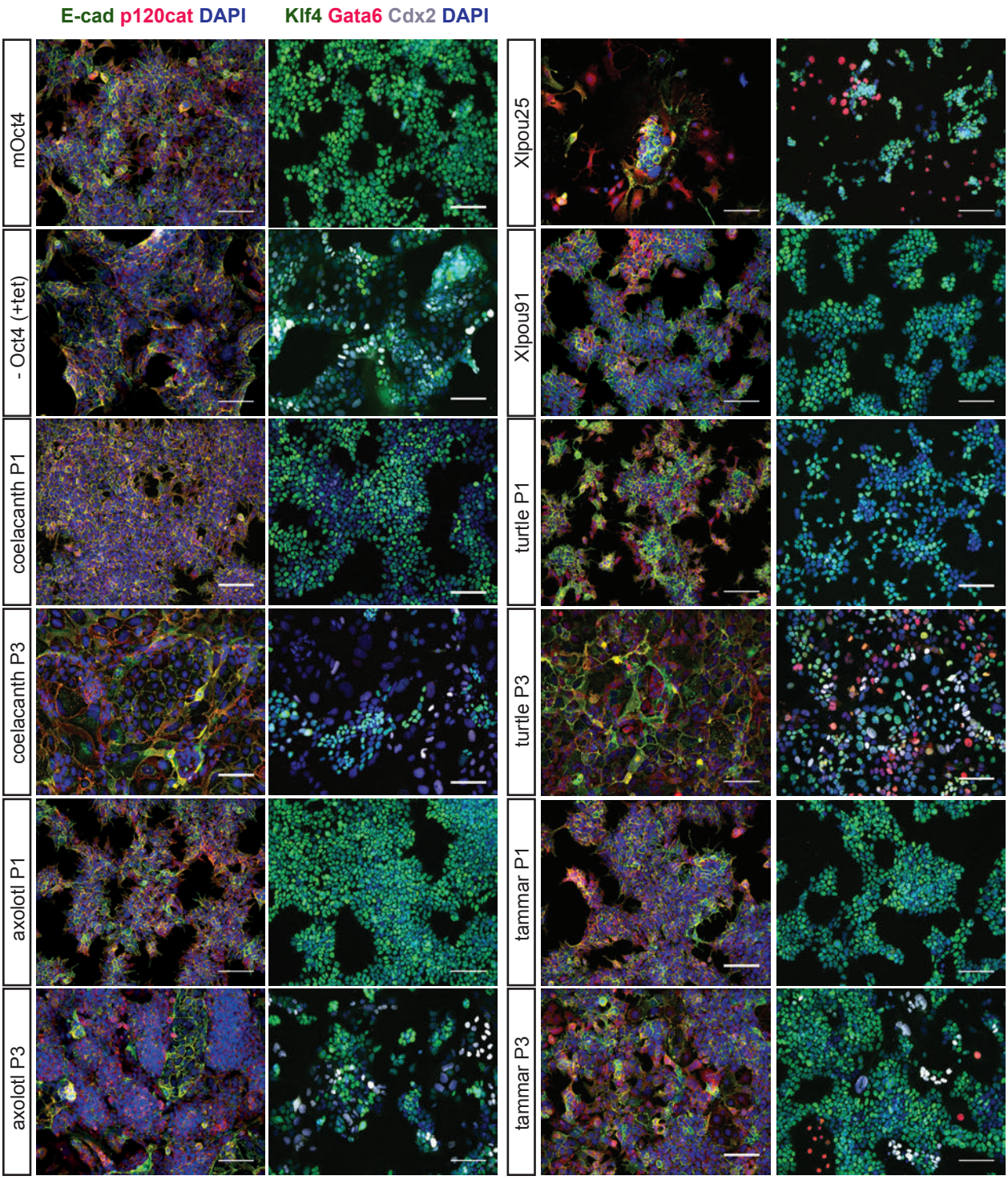


Figure 3.10 Characterization of POUV-rescued ESC clonal lines by immunofluorescence
The scale bars represent 100 µm. The images are the overlay of different fluorescence channels.
Abbreviations: P1, POU5F1; P3, POU5F3

contain mixed populations of potentially primed pluripotent and differentiated cells. Thus in species where both POUV proteins were retained, POU5F3 proteins diverged from the support of naïve pluripotency and potentially become different lineage specific regulators of potency and differentiation. On the other hand, POU5F1 proteins retained the capacity to support the naïve pluripotent network.

To better quantify degree to which different POUV proteins support naïve pluripotency or allow for differentiation, I performed qRT-PCR for the expression of Esrrb (ESC marker), Prdm14 (ESC/PGC marker), Gata6 and Cdx2 (differentiation markers). The results are shown in figure 3.11. Consistent with the immunofluorescence results, some clonal lines rescued by Xlpou25 and turtle POU5F3 exhibited higher level of Gata6 expression. Similarly, Cdx2 expression level was higher in lines rescued by coelacanth POU5F3, axolotl POU5F3, turtle POU5F3, and tammar POU5F3. Coelacanth POU5F3 showed the greatest variation in Cdx2 expression levels amongst clonal lines, some exhibiting levels approaching those produced in the absence of any Oct4 or POUV protein ZHBTc4+Tc. In general, with species carrying two POUV forms, POU5F1s were more effective at supporting Esrrb and Prdm14 expression than its paralog, but the levels of these gene expressions were varied among lines. Amongst all examined POU5F3, axolotl POU5F3 was the best in rescuing Prdm14 expression. Prdm14 expression in these axolotl POU5F3 supported lines was similar to that in Xlpou91-rescued ESCs. This suggests that in the case of the axolotl, its POU5F3 contains some capacity to activate the germ cell gene regulatory network, and this fits well with the expression pattern of the axolotl POU5F1 and POU5F3, as both are expressed in its germ cells.

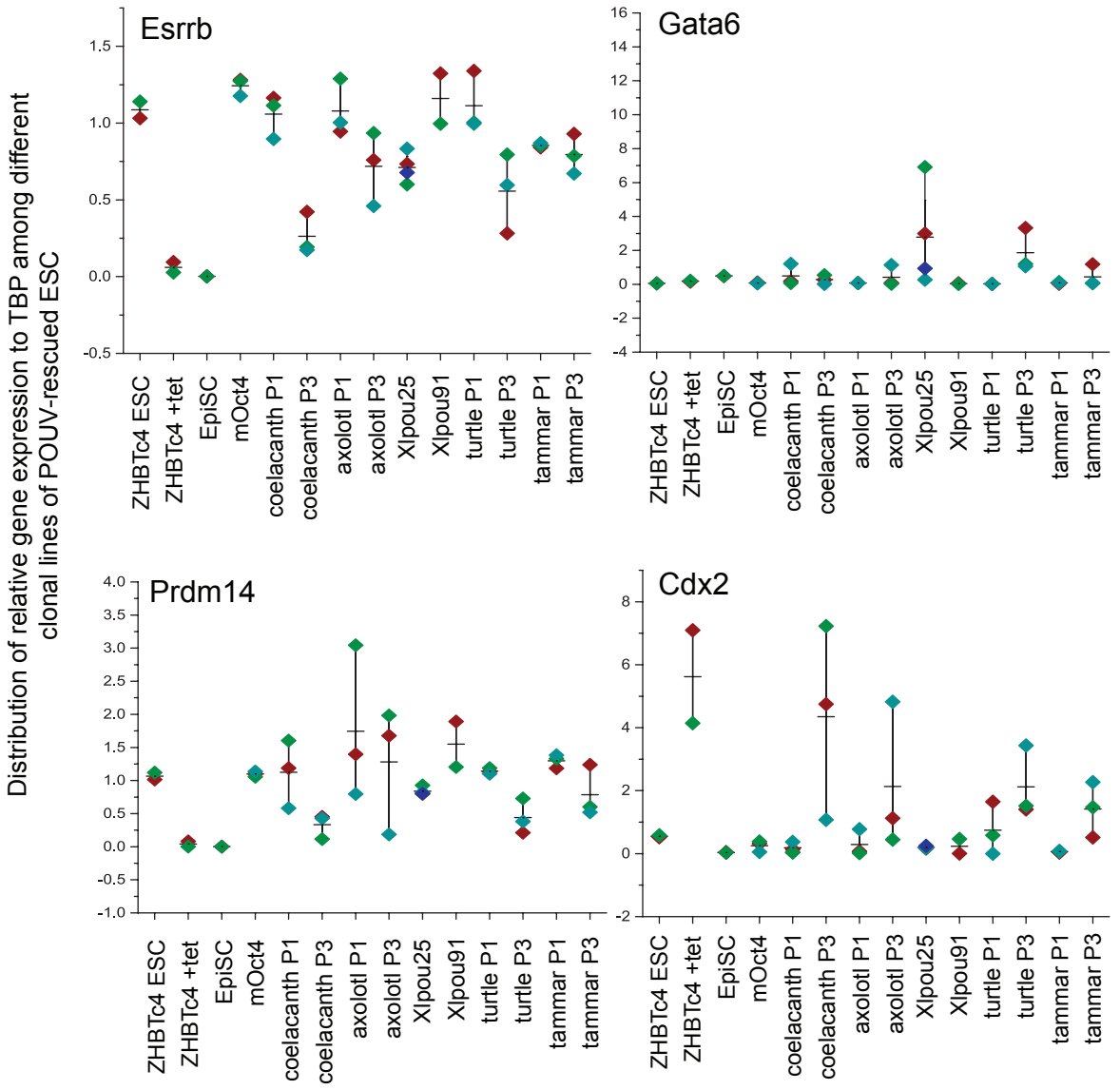


Figure 3.11 The analysis of gene expression among different POUV-rescued ESC clonal lines The relative expression of pluripotency markers (Esrrb and Prdm14) and differentiation markers (Gata6 and Cdx2) was quantified by qRT-PCR and calculated by normalization with a house keeping gene TBP. Each diamond symbol represents the average value of each individual clone. The horizontal bars between diamond symbols are the mean of average values of the transcript level in each POUV-rescued clones.

Next I aimed to examine ESC phenotype of rescued lines by flow cytometry using previously identified surface markers (figure 3.12). Livigni et al 2013 previously found that Xlpou25, a key player during gastrulation of *Xenopus* embryogenesis, drove higher E-cadherin level than Xlpou91 in mouse Oct4 null ESCs. This leads to the hypothesis that epiblast specific POUV proteins might be *bona fide* regulator of gastrulation stage adhesion. In tammar wallaby POU5F3 protein is expressed specifically in nucleus of epiblast cells during gastrulation while the POU5F1 is retained in the cytoplasm. This suggests that probably the conserved roles in the regulation of adhesion by POUV protein belong to POU5F3 subfamily. To address this, I assessed the E-cadherin levels of some POU5F1 and POU5F3-rescued lines, alongside with Xlpou91 and Xlpou25-rescued lines as the controls. In addition, the expression of Pecam-1, which is an undifferentiated ESC and germ cell surface marker, was also analyzed together with E-cadherin

Figure 3.12 shows flow cytometry exploring the expression of E-cadherin and Pecam-1 in rescued cell lines. I found that mOct4-rescued line exhibited similar E-cadherin and Pecam-1 level to ZHBTc4 (without Tc). Supporting the notion that POU5F1 has germ cell activity, tammar POU5F1 and Xlpou91 could drive higher levels of Pecam-1 expression than mouse Oct4, while E-cadherin levels of these cell lines remained similar to those rescued by mouse Oct4 and the ZHBTc4 ESC. Similar to what we found previously, Xlpou25 drove higher E-cadherin expression than Xlpou91. Moreover, tammar POU5F3-rescued lines contained a population of cells expressing E-cadherin level similar to Xlpou25-rescued line. I also investigated the E-cadherin level in the lines rescued by coelacanth POU5F1 and POU5F3. Similar to the observations made for the tammar wallaby, coelacanth POU5F3 also induced higher level of E-cadherin than the POU5F1 and mOct4, although the level was not as high as in the lines rescued by Xlpou25 and tammar POU5F3. Taken together these observations suggest that POU5F3 activity has been specialised in efficiently regulating adhesion and differentiation in gastrulation.

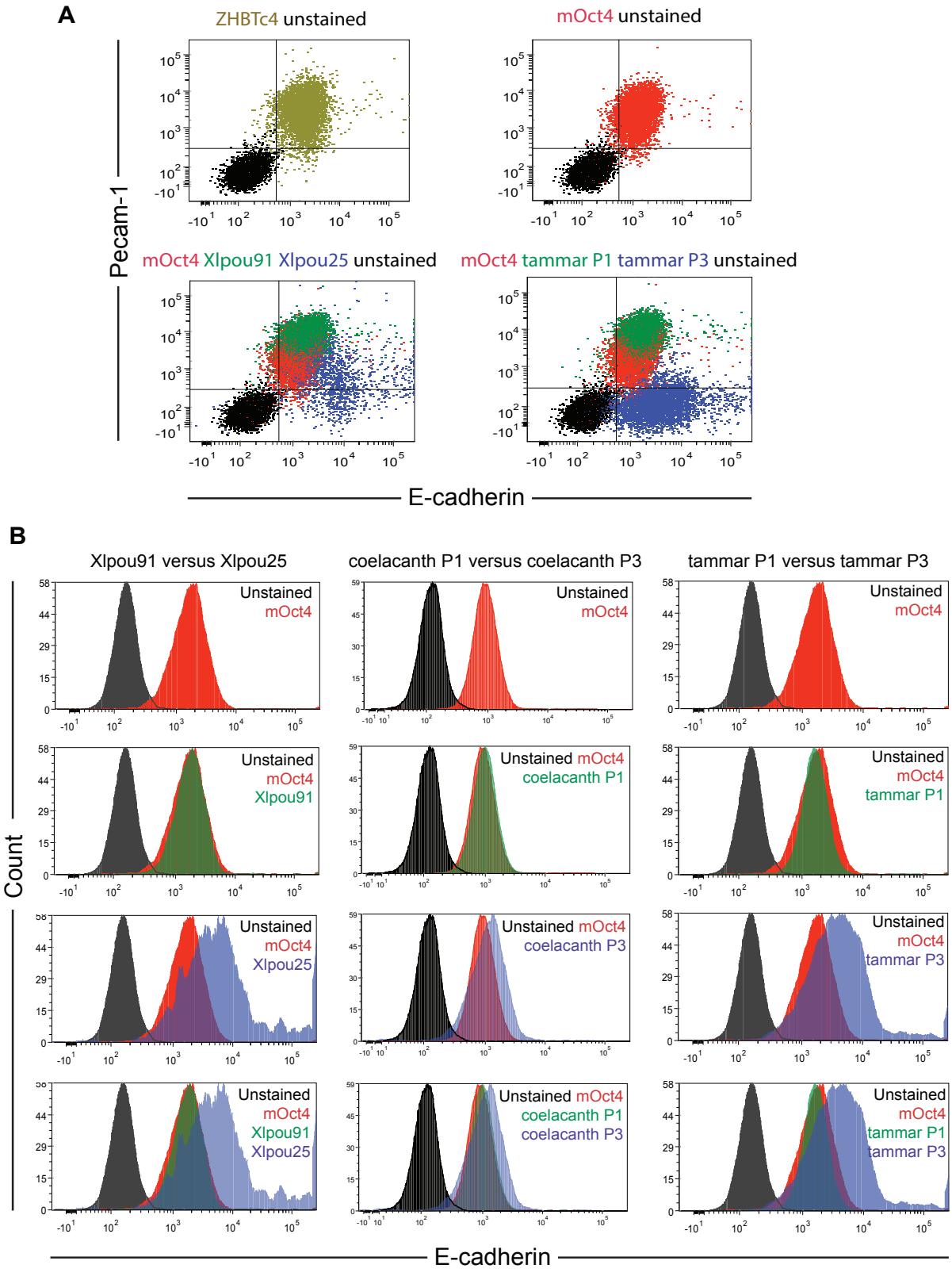


Figure 3.12 Flow cytometry analysis on Pecam-1 and E-cadherin of POUV-rescued ESC lines
A)-B) POUV-rescued ESC lines were plated on gelatinized culture dishes and cultured under ESC self-renewal condition for 5 passages after colony picking. The levels of pluripotency surface markers (Pecam-1 and E-cadherin) were assessed by flow cytometry and analysed by FCS EXPRESS software. A) Data represents the overlay on Pecam-1 against E-cadherin of different POUV-rescued ESC lines. B) Data represents the overlayed histograms on E-cadherin levels among different POUV-rescued lines. The counts were normalized by the peak value in FCS EXPRESS software.

Section 3.4 Global transcriptome analysis of coelacanth POUV-rescued ESC lines

To evaluate the expression states supported by POU5F1 and POU5F3 in an unbiased way, I assessed the transcriptomes of a pair of POUV paralogs. For this analysis I chose the coelacanth POU5F1 and POU5F3, as this pair is both well conserved and appears to have the largest difference in the rescue index for the two proteins. Stable coelacanth POU5F1 and coelacanth POU5F3-rescued Oct4-null ESC lines also exhibited the most distinctive phenotypes. The lines exhibited ESC phenotypes similar to those obtained from Oct4 rescue, while POU5F3 supported lines exhibited high levels of differentiation. Thus in ESC rescue activity, these two proteins appear to have distinct activities. As coelacanth protein coding genes exhibit very slow substitution rates (Amemiya et al., 2013), the segregated protein activities of coelacanth POUV proteins might represent ancestral division of POUV function. As a result, gaining insight into gene expression supported by these different proteins might provide insight into the ancestral pluripotency network and how its activity is effectively partitioned.

To explore the global gene network, we performed one-colour based microarray on two biological replicates of each coelacanth POU5F1-rescued line, coelacanth POU5F3-rescued line and mouse Oct4-rescued line. The full detail of microarray procedure is described in chapter 2 section 2.6.2. Briefly, I collected RNA from passage 6 of the stable-rescued lines. Only high quality RNA (RNA integration number (RIN) = 10) was used to generate Cyt3-labelled cRNAs, which were then hybridised to whole genome-oligonucleotide microarray slides (Agilent Technologies). The hybridised slides were scanned to obtain global gene expression profiles. To perform statistical analysis on microarray expression data, the raw datasets of samples were combined and uploaded onto

the web-based software NIA Array Analysis (http://lgsun.grc.nia.nih.gov/ANOVA/) (Sharov et al., 2005). All inputs were normalised by the platform to equalise multiple quantiles of the probability distribution of gene expression and analyzed based on the ANalysis Of VAriance (ANOVA). The statistical significance was determined using False Discovery Rate (FDR) method. This global gene expression analysis were based on false discovery rate threshold (FDR) = 0.05 and fold-change threshold = 2. From these criteria, I obtained 6374 significant genes different among 3 conditions: mOct4 versus coelacanth POU5F1, mOct4 versus coelacanth POU5F3, and coelancanth POU5F1 versus POU5F3-rescued ESC lines. I used GENE-E to hierarchically cluster gene expression based on mean Euclidean distance of most (FDR<0.05, totally 6374 genes). This clustering is visualised in the heat map shown in in Figure 3.13A, and show that the mOct4-rescued line clustered together with those rescued by coelacanth POU5F1, not coelacanth POU5F3.

I then investigated how global gene expression profiles are different between each rescued line by performing pair-wise comparisons (Figure 3.13B). The pairwise comparison of gene expression profiles between each pair of the rescued lines allowed the identification of genes that are differentially expressed at an significance level FDR < 0.05, in both overexpressed and underexpressed direction. Results of the pairwise comparisons of POUV-rescued lines are shown as following:

(a) Comparison of mouse Oct4 and coelacanth POU5F1 rescued line

(Figure 3.13B, left)

The log–ratio plot illustrates that both mouse Oct4 and coelacanth POU5F1 rescued lines are quite similar and show relatively few differences in gene expression. There were 597 genes upregulated in mouse Oct4 line and 102 genes upregulated in coelacanth POU5F1 line.

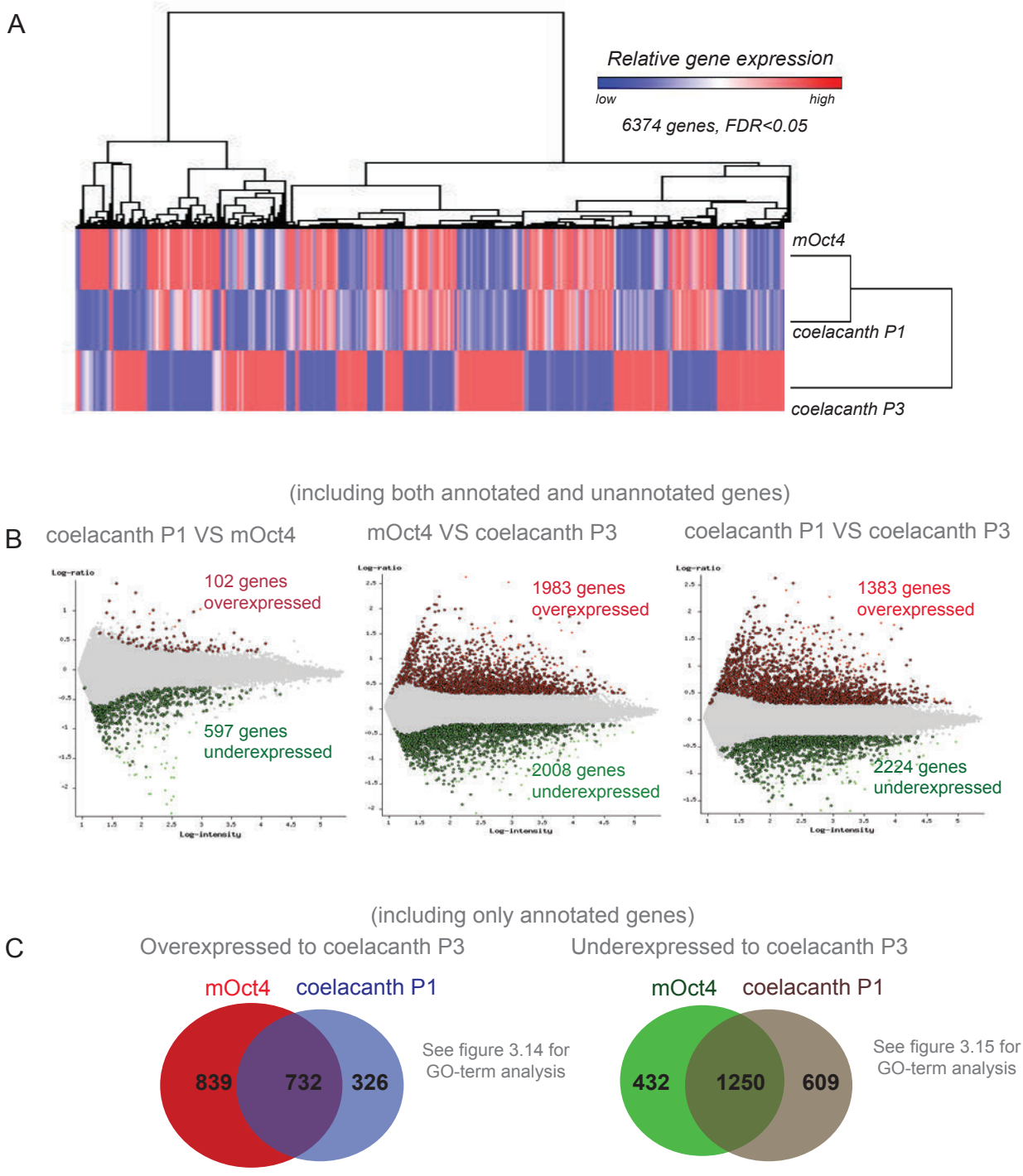


Figure 3.13 Microarray analysis of Oct4-null ESC rescued by coelacanth Pou5f1 and Pou5f3 proteins A) Hierarchical clustering of gene expression (FDR < 0.05) was constructed by GENE-E software. The relative expression level of each gene was shown in three-color format, in which red, white and blue mark high, medium and low level of that gene expression respectively. B) Log-ration plots showing genes overexpressed (in red) and underexpressed (in green) in pairwise comparison of coelacanth POU5F1 versus mOct4 (left), mOct4 versus coelacanth POU5F3 (middle), coelacanth POU5F1 versus coelacanth POU5F3 (right). C) Overexpressed/underexpressed gene lists of mOct4 and coelacanth POU5F1 against coelacanth POU5F3 (refer to middle and right log-ration plots in B) were used to produce Venn Diagrams to further identify common genes expressed in both mOct4 and coelacanth POU5F1 or specific to either mOct4 or coelacanth POU5F1 cluster. Abbreviations: P1, POU5F1; P3 POU5F3.

(b) Comparison of mouse Oct4 and coelacanth POU5F3 rescued line

(Figure 3.13B, middle)

The log-ratio plot illustrates that there were 1983 genes upregulated in mouse Oct4-rescued line and 2008 upregulated in coelacanth POU5F3-rescued line.

(c) Comparison of coelacanth POU5F1 and coelacanth POU5F3

(Figure 3.13B, right)

The log-ratio plot illustrates that there were 1383 genes upregulated in coelacanth POU5F1-rescued line and 2224 genes upregulated in coelacanth POU5F3-rescued line.

From these pairwise comparisons, it is apparent that the pattern of up- and downregulated genes of mouse Oct4 line and coelacanth POU5F1 lines are very much similar. Thus, to decipher which up- and down-regulated genes are shared by both mouse Oct4 and coelacanth POU5F1 lines, I extracted the list of these genes from NIA Array tool, and analyzed them by web-based software to produce Venn diagram, which visualise shared or unique genes between gene lists.

Venn Diagram in figure 3.13C illustrates the results as following:

(a) Overexpressed genes of mouse and coelacanth POU5F1 compared to coelacanth POU5F3

In the Venn diagram, there are 732 genes shared by both mouse and coelacanth POU5F1. 839 genes are specific to mOct4 and 326 genes are specific to coelacanth POU5F1.

(b) Underexpressed genes of mouse and coelacanth POU5F1 compared to coelacanth POU5F3

In the Venn diagram, there are 1250 genes shared by both mouse and coelacanth POU5F1. 432 genes are specific to mOct4 and 609 genes are specific to coelacanth POU5F1.

In section 3.5-3.6, I explored more on functional significance of these gene lists by gene annotation (GO) term enrichment analysis and gene network analysis by GENEMANIA.

Section 3.5 Gene ontology (GO)-term analysis and gene network comparison of coelacanth and mouse POUV-rescue ESC lines

To investigate the functional significance of genes regulated by POU5F1, POU5F3 or both of them, GO enrichment analysis was performed using ExATLAS (NIA) and BINGO in CYTOSPACE platform. As described in section 3.4 and Figure 3.13C, lists of significant gene for GO term analysis were obtained from the Venn Diagrams: 732 genes overexpressed and 1250 genes underexpressed in both mOct4 and coelacanth POU5F1 (compared to coelacanth POU5F3). These gene lists were then analyzed by those web-based GO analysis platform to obtain overview of clusters of the biological/ cellular/ molecular functions represented as GO term networks.

Firstly, as shown by both ExATLAS and BINGO analysis, the list of 732 genes overexpressed in both mOct4 and coelacanth POU5F1 was enriched in genes regulating (1) biological functions involving developmental process, stem cell maintenance and differentiation, and reproduction process, in particular spermatogenesis (2) cellular function involving cell proliferation and (3) metabolic process involving primary metabolic process and nucleic acid metabolism. Stem cell maintenance genes included in the list were *Sox2*, *Tcl1*, *Rif1*, *Esrrb*, *Kit*, *Klf4* and *Fgf4* (figure 3.14). In addition, it was found that among reproduction genes upregulated by both mouse and coelacanth POU5F1, most were are related to basic germ cell development (including *Prdm14*, *Nanos3* and *Zfp42*), spermatogenesis, while less genes were related to oogenesis, see appendix figure X. In addition, mouse Oct4 specific cluster (839 genes) upregulated only in mOct4 was also analyzed by the same analysis platform. Interestingly, there is another gene set of developmental process (including reproduction and cell differentiation) and nucleic metabolic process specific to mouse Oct4, which is not induced by coelacanth POU5F1.

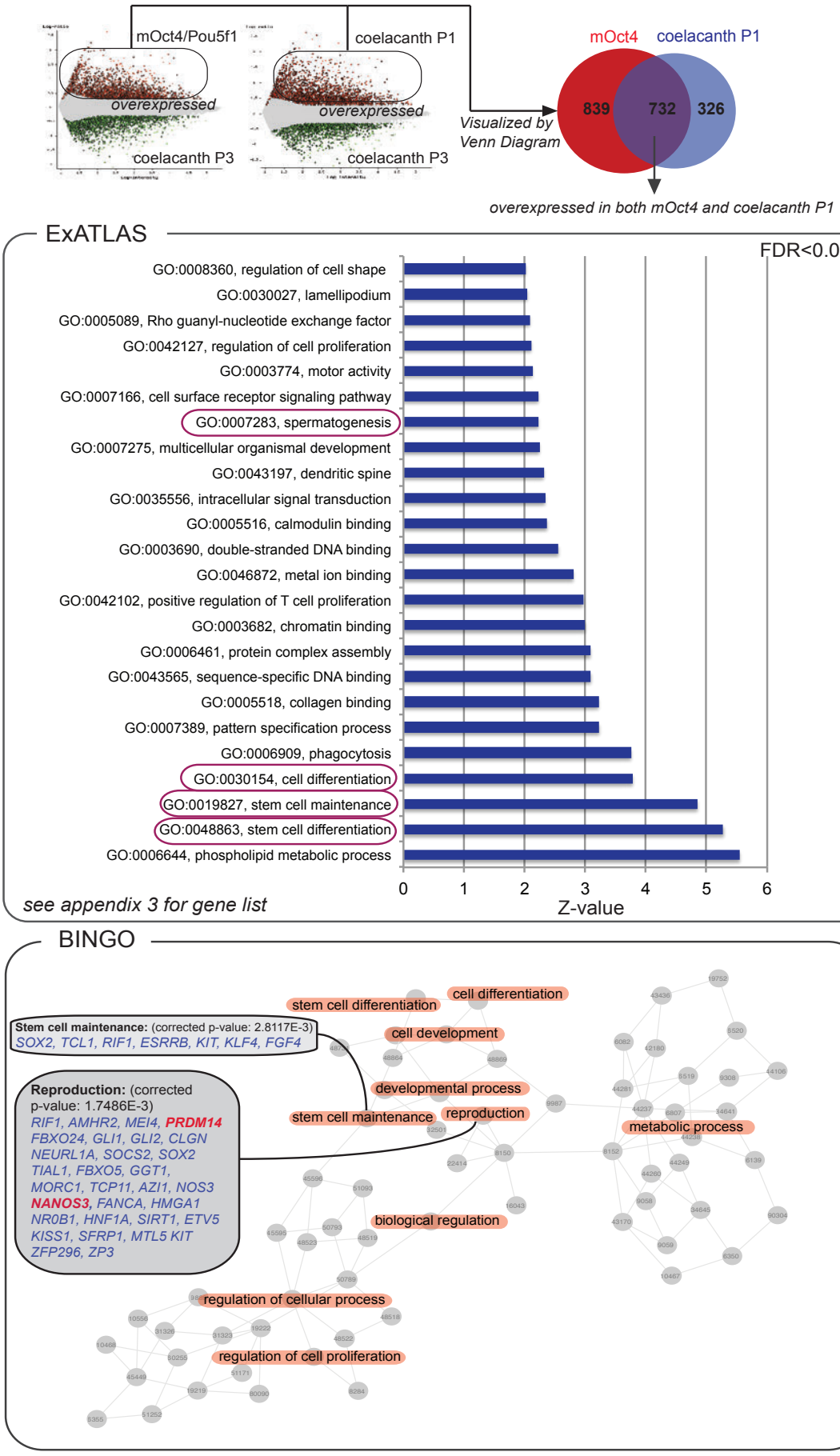
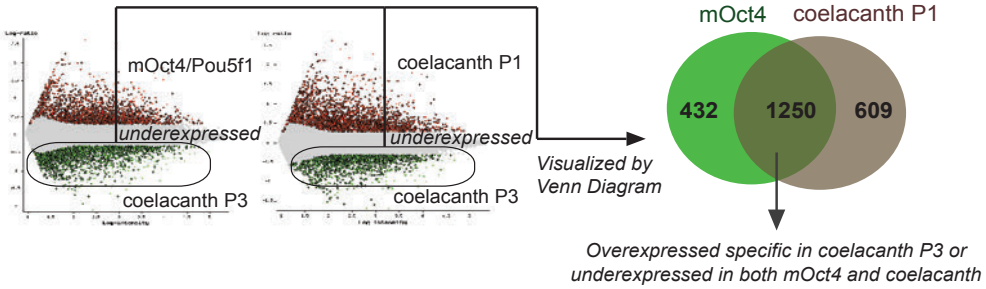
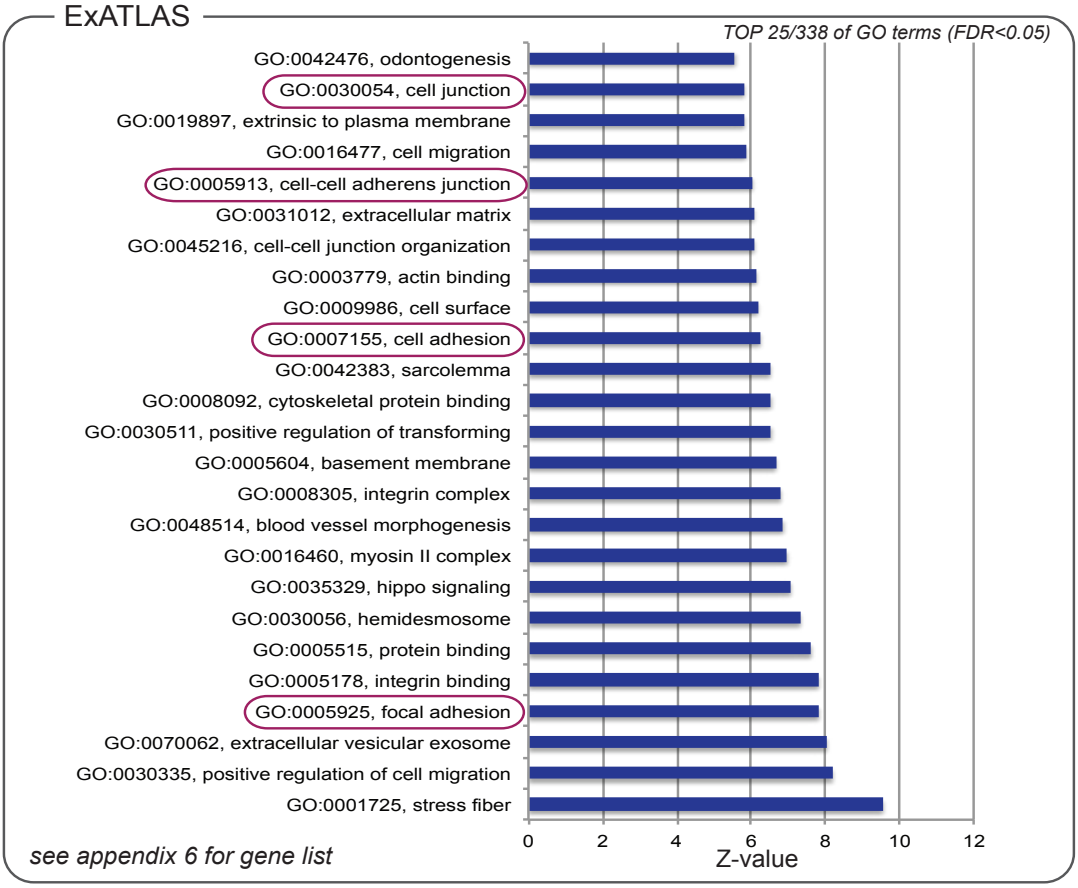


Figure 3.14 Gene annotation (GO)-term analysis of genes overexpressed in both mOct4 and coelacanth POU5F1 compared to coelacanth POU5F3. Overexpressed genes (732 genes) shared by both mOct4 and coelacanth POU5F1 were analyzed by two GO-term analysis platforms: ExATLAS and BINGO. A) GO term analyzed by ExATLAS. The parameter is FDR<0.05. Gene lists of each GO cluster are shown in appendix table X. B) GO term analyzed by BINGO in CYTOSPACE software. The parameters were including 1) Binomial statistical test 2) Bonferroni Family Wise Error Rate (FWER) correction 3) significant level *P*< 0.005 4) visualization by overrepresented categories after correction and 5) GO_Biological_process as ontology file. Gene lists of some GO cluster are shown in appendix 3.

Among these mOct4 specific genes, most of reproduction genes were involved mostly in spermatogenesis, see appendix 1 and 2. Gene lists in each GO term are shown in appendix 3-5.

Next, I also performed GO analysis with 1250 genes overexpressed specifically in coelacanth POU5F3. As shown in figure 3.15, highly significant GO term for this cluster was cell differentiation that included various differentiation programs including neurogenesis, cartilage development, bone development, blood vessel development, epithelial cell differentiation and striated muscle cell differentiation. Remarkably, GO term for cellular function was enriched in cell-adhesion, cellular component movement, cytoskeleton organization, cell migration, localization and transport. Surprisingly, both E-cadherin (*Cdh1*) and N-cadherin (*Cdh2*) were more significantly upregulated in POU5F3 than in POU5F1-rescued lines. We have previously shown that Xlpou25, which is the epiblast-specific POUV protein and adhesion-related regulator during *Xenopus* gastrulation, induces higher level of E-cadherin in the Oct4-null ESC cell compared to Xlpou91 and maintained high level of E-cadherin and p120 catenin during the forced induction of differentiation. The upregulation of *Cdh1* gene in coelacanth POU5F3 from the microarray data is consistent with higher level of E-cadherin observed in flow cytometry, implying that POU5F3 might be adhesion-related regulator during coelacanth gastrulation, similar to Xlpou25. In general, *Cdh2* encoding N-cadherin protein is downregulated once *Cdh1* is upregulated as shown by ESC-EpiSC conversion, and N-cadherin is specifically found in EpiSC. The upregulation of *Cdh2*, together with *Cdh1*, in the POU5F3 line is surprising while suggests that ESC phenotypes induced by Pou5f3 maybe biased toward epiblast/EpiSC-like. Gene lists in each GO term (specific to coelacanth POU5F3) are shown in appendix 6.





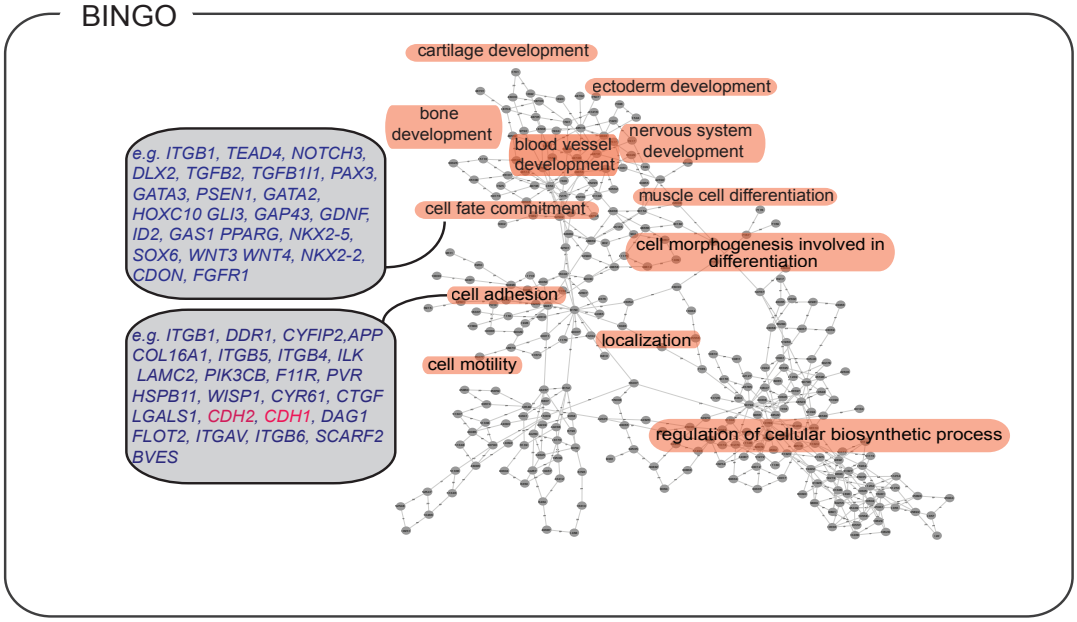


Figure 3.15 Gene annotation (GO)-term analysis of genes overexpressed in coelacanth POU5F3. Overexpressed genes (1250 genes) specific to coelacanth POU5F3 and downregulated in both mOct4 and coelacanth POU5F1 were analyzed by two GO-term analysis platforms: ExATLAS and BINGO. A) GO term analyzed by ExATLAS. The parameter is FDR<0.05. Gene lists of each GO cluster are shown in appendix table X. B) GO term analyzed by BINGO in CYTOSPACE software. The parameters were including 1) Binomial statistical test 2) Bonferroni Family Wise Error Rate (FWER) correction 3) significant level *P*< 0.005 4) visualization by overrepresented categories after correction and 5) GO_Biological_process as ontology file. Gene lists of some GO cluster are shown in appendix 6.

To better assess the relationship between POU5F3 supported ESC lines and EpiSC, I compared the list of genes upregulated in POU5F1 (732 genes) and POU5F3 (1240 genes) to the top 100 most significant genes upregulated and downregulated in EpiSC compared to ESC (Hayashi et al 2013). Here I identified,

(1) There are 31 genes shared by both EpiSC upregulated cluster and POU5F3-rescued line upregulated gene cluster (Figure 3.16, top panel).

(2) There are 25 genes shared by both EpiSC downregulated gene cluster and mOct4/coelacanth POU5F1-rescued lines upregulated gene cluster (Figure 3.16, bottom panel).

The significance of these comparisons is born out by the fact that only 1 gene in either the reciprocal POU5F1 or POU5F3 dependent gene lists is common to the EpiSC gene lists. To confirm that these genes were highly expressed in POU5F3 line or POU5F1 lines, I took these 31 and 25 genes to further analyze in NIA Array Analysis to obtain log-intensity representing the expression level from the microarray. The log-intensity values were plotted on the same histogram to illustrate the behavior of EpiSC genes in the POU5F1 and POU5F3 rescued cell lines (Figure 3.16). Genes highly expressed in POU5F1 lines, and reduced in POU5F3 lines are mostly related to naïve pluripotency (e.g. *Klf2*, *Klf4*, *Esrrb*, *Fbxo15*, *Fgf4*, *Kit*, *Nr0b1*, *Zfp42*). In summary, the combination of differentiation specific genes, and EpiSC markers, suggests that POU5F3 supports a primed pluripotent state, while POU5F1 has evolved an activity that enables it to support naïve pluripotency.

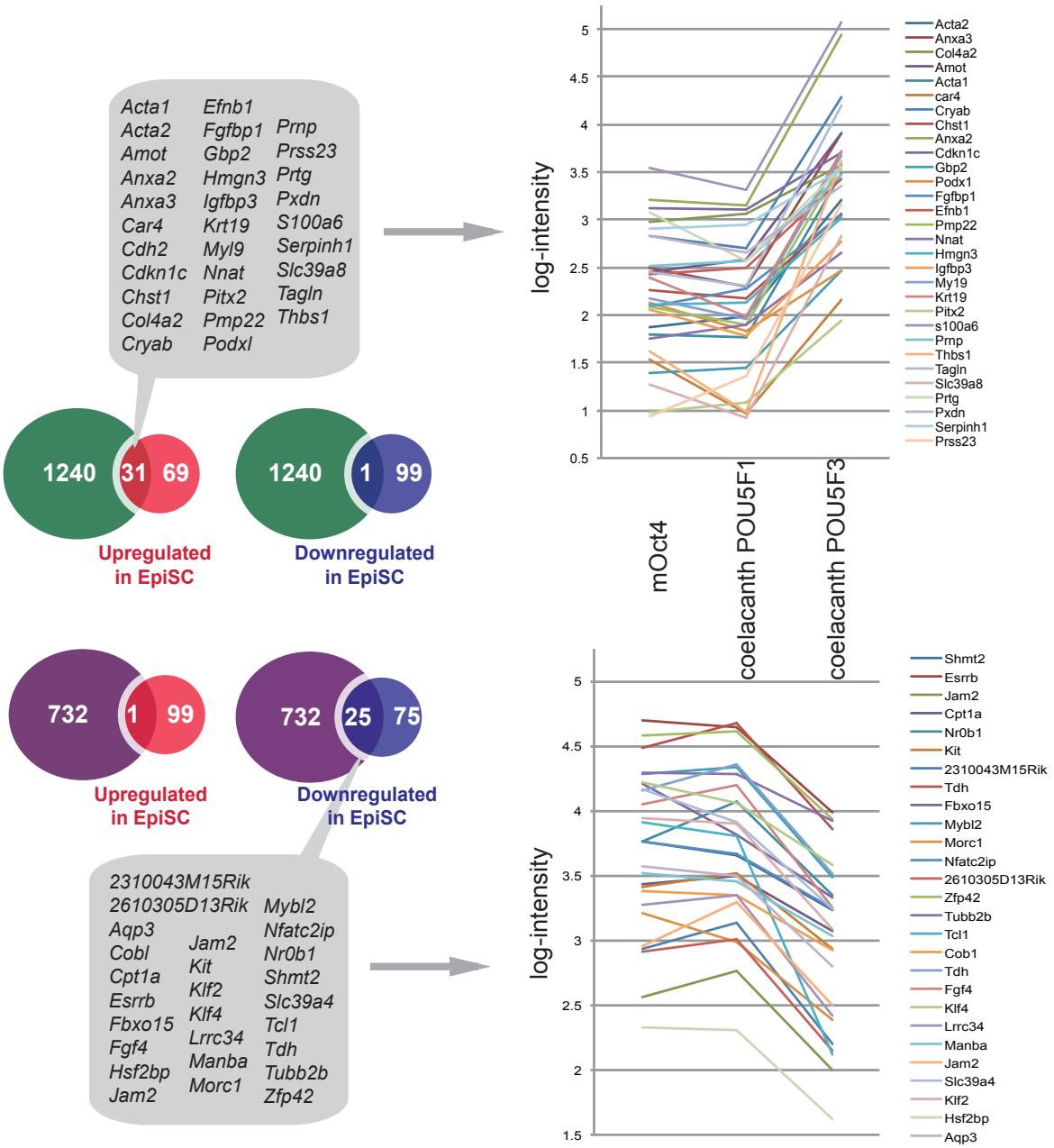


Figure 3.16 Coelacanth POU5F3-mediated rescue of Oct4 null ESCs induces EpiSC-like gene expression. We took (1) 732 genes upregulated in both mouse and coelacanth POU5F1 (compared to coelacanth POU5F3) and (2) 1240 genes upregulated specific to coelacanth POU5F3 (downregulated in both mOct4 and coelacanth POU5F1) and compared them to genes up- or downregulated in EpiSCs as compared to ESCs. The overlap of these lists with the top 100 most significant genes up- or downregulated in EpiSC (Hayashi et al., 2013) are shown in the Venn diagrams. 31 genes were shared by both coelacanth POU5F3 rescued line and EpiSC while 25 genes were shared by both mOct4/coelacanth POU5F1 and ESC (downregulated in EpiSC). The log intensity obtained from NIA Array Analysis of these 31 and 25 genes were plotted onto the same histogram to illustrate the relative expression level.

Section 3.6 Gene expression profiles of mouse and coelacanth POUV-rescued lines

In order to better understand biological nature of POUV-rescued lines, I investigated microarray-based expression pattern of some genes with known biological functions related to pluripotency and differentiation. I grouped genes with the same functions or from the same protein family to illustrate their expression levels, as shown in figure 3.17.

Group of genes based on gene functions

(1) Naïve pluripotency

Embryonic stem cells exhibit naïve pluripotency, in which Oct4 is a well-known crucial player in maintaining this level of cell potency. Deletion of Oct4 demolishes pluripotency circuity and the ESCs spontaneously differentiates toward trophectoderm (Nichols and Smith, 2009; Nichols et al., 1998). I found that coelacanth POU5F1 could potentially support naïve pluripotency-related gene expression (*Zfp42* (*Rex1*), *Nr0b1*, *Fgf4*, *Nanog* and *Esrrb*) as well as mouse Oct4. In contrast, coelacanth POU5F3-rescued lines expressed lower levels of these genes.

(2) Trophectoderm

In embryo and ESCs, the deletion of Oct4 leads to the upregulation of Cdx2 (one of key trophectoderm marker) and vice versa. The reciprocal transcriptional regulation of Oct4 and Cdx2 is important for inner cell mass/trophectoderm lineage segregation and maintaining ESC potency (Chew et al., 2005; Nichols et al., 1998; Wu et al., 2010). The microarray analysis revealed that *Cdx2* was one of the top significant genes upregulated in coelacanth POU5F3 lines. Other genes specific for trophectodermal lineage, in particular

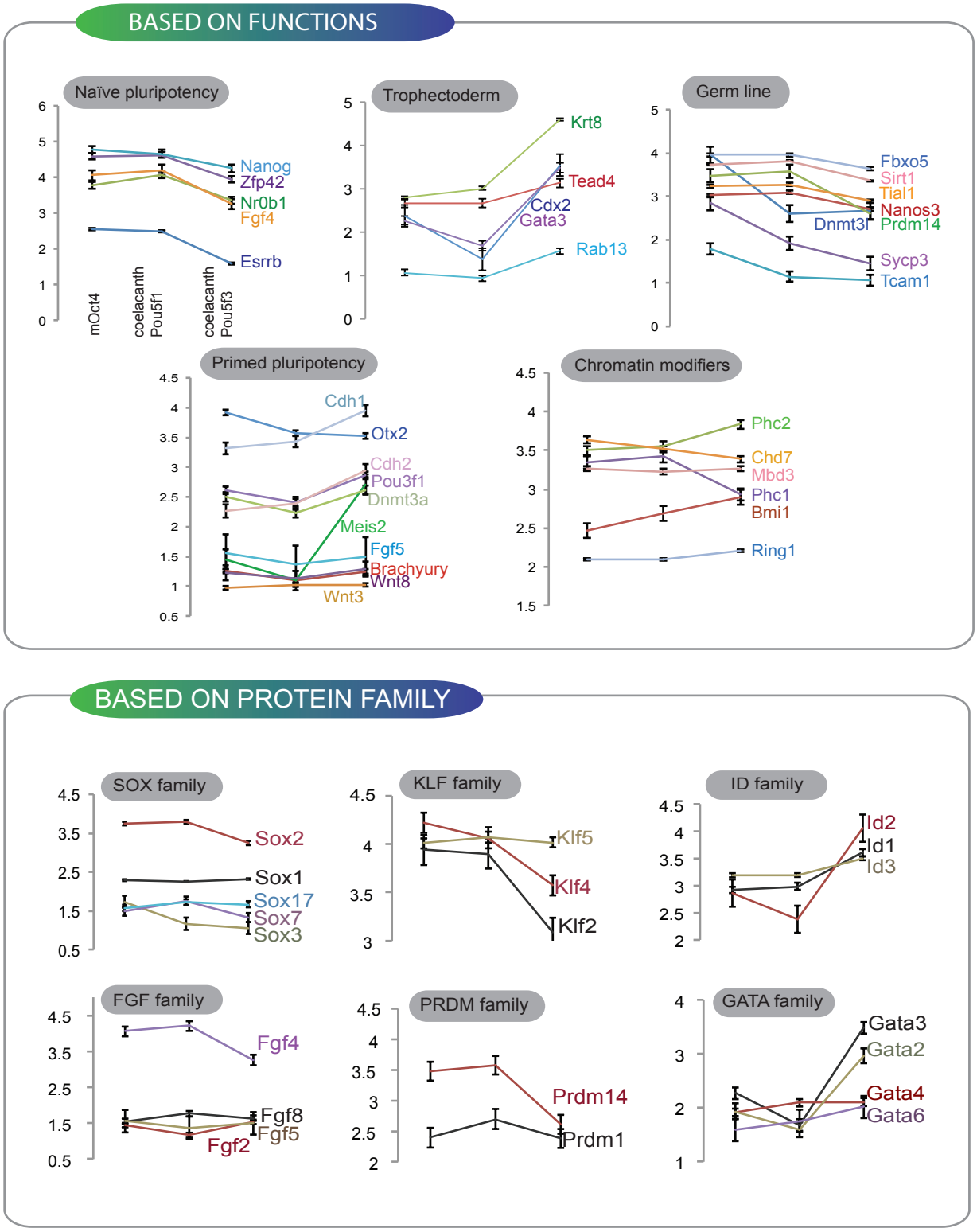


Figure 3.17 Rescue of POU5F1 mutant ESCs shows a high degree of functional conservation between mouse POU5F1 and coelocanth POU5F1, but not coelocanth POU5F3. Normalized log-intensity value plots for selections of differentially expressed genes in Oct4/Pou5f1 conditional null ESCs (ZHBTc4) rescued using mouse POU5F1, coelacanth POU5F1 or coelacanth POU5F3 are shown.

Krt8 and *Gata3, Tead4* and *Rab13* were also significantly higher in POU5F3. This indicates the conserved roles of POU5F1 in suppressing trophectoderm-related programs. Based on the fact that trophectoderm doesn't exist in non-mammalian vertebrates, downregulated *Cdx2* expression in coelacanth POU5F1 rescued line implies the existence of POUV-CDX network prior the emergence of sarcoptergians. As a number of trophoblast markers are also mesoderm and endoderm markers, this suggests that the evolution of trophoblast is based on an ancient gene regulatory network involved in germ layer specification.

(3) Germ cells

Several naïve pluripotency-related genes expressed in ESCs are shared with those from early specified PGC including PGC-derived EGCs (Hayashi et al., 2011; Leitch et al., 2013; Saitou and Yamaji, 2010). *Prdm14*, one of tripartite regulators (Prdm1/Blimp1, Prdm14, Tcfap2c/AP2) essential for germ cell specification (Magnúsdóttir et al., 2013), was expressed at similar level in both mouse and coelacanth POU5F1 line. However, *Prdm1* and *Tcfap2c* were not significantly different between three rescued lines. In addition, other germ line markers (*Fbxo5*, *Sirt1*, *Tial1*, *Nanos3*) were also expressed at similar level to *Prdm14*; whereas, some (*Dnmt3l*, *Sycp3*, and *Tcam1*) were specifically expressed at higher level in mOct4 rescued line. This indicates that mOct4 might have additional link to germ cell network not present in coelacanth POU5F1.

(4) Primed pluripotency

Primed state of pluripotency is found in post-implantation epiblast and epiblast stem cells, where genes related to various differentiation programs (e.g. *Sox9*, *T-Brachyury*, *Wnt3a*, *Wnt8*, *Sall3*, *Meis2*, *Pou3f1* (*Oct6*), *Otx2*, *Fgf5*, *Cer-1*, *Lefty-1*, *Cdh2*) are upregulated while those related to naïve state are downregulated (Buecker et al., 2014; Hayashi et al., 2011). I found that some of primed pluripotency–related genes (*Fgf5*, *T-Brachyury*, *Wnt3a*, *Wnt8*) were expressed at low level and not significantly different between

three rescued lines. Interestingly, these genes are primitive streak markers and other epiblast specific markers are upregulated such as *Cdh1*, *Cdh2*, *Pou3f1* and *Meis2* that are all expressed at higher levels in POU5F3 supported ESCs.

(5) Epigenetic regulators

Chromatin modifiers are known to be important for the maintenance of pluripotency and differentiation of ESCs (Boland et al., 2014; Kraushaar and Zhao, 2013). It has been shown by some studies that Oct4/POUV network is linked to these epigenetic regulators (Loh et al., 2006). From my microarray, most of the candidate epigenetic modifiers (e.g. *Ring1a*, *Ring1b*, *Prc1*, *Chd7*, *Mbd3*) were not significantly different between POUV-rescued lines.

Group of genes based on protein family

(1) SOX family

Transcription factors of SOX family regulate cell fates during development and are conserved among animals (Sarkar and Hochedlinger, 2013). Sox2, together with Oct4, are key players regulating self-renewal and pluripotency of embryonic stem cells (Rizzino, 2009). Sox1/Sox3 (also Sox2) and Sox7/Sox17 regulate neural development and endoderm/primitive endoderm differentiation, respectively (Artus et al., 2011; Niimi et al., 2004; Uchikawa et al., 2011). From the microarray, *Sox2* was expressed at higher level than other examined Sox genes (*Sox1*, *Sox3*, *Sox7* and *Sox17*). Coelacanth POU5F1 sustained ESC level of Sox2 expression, which could not be achieved by its paralogs. Other Sox genes were not significantly different among the rescued lines.

(2) Krüppel like factor (KLF) family

Klf2, Klf4 and Klf5 regulate ESC pluripotency and are functionally redundant. The deletion of all three KLF leads to ESC differentiation (Jiang et al., 2008). In ESCs, Oct4 primarily regulates Klf2 and Klf4 expression, with KLF4 expression also supported by LIF/STAT signaling. Both Klf2 and Klf4 are downregulated in epiblast stem cells and the induction of either of Klf2 or Klf4 expression can reactivate naïve pluripotency (Hall et al., 2009). The microarray analysis showed that both *Klf2* and *Klf4* were upregulated in both mouse and coelacanth POU5F1 lines and downregulated in the POU5F3 supported lines, in contrast to *Klf5* expression was unaffected by different POUV proteins. This result is in the line with previous findings and indicates the conserved POUV role is more linked to Klf2 and Klf4 than to Klf5.

(3) ID family

Id1, Id2, and Id3 play roles in ESC self-renewal, cell differentiation towards neural lineages, and in neural stem cell maintenance. In addition, Id2 is found in mouse chorionic trophoblast cells (Jen et al., 1997; Romero-Lanman et al., 2012). I found that *Id1* and *Id3* were expressed in all rescued cells and did not appear to significantly vary. However, *Id2*, like other trophectoderm markers such as *Cdx2*, *Krt8* and *Gata3* was upregulated in POU5F3 supported cells. This agrees well with the presence of trophoblast lineages in POU5F3 line.

(4) Fibroblast growth factor (FGF) family

FGFs regulated pluripotency in both the naïve and primed state. The expression of Fgf4 and Fgf2/5/8 is characteristic of ESC and epiblast state, respectively (Nichols and Smith, 2009; Sumi et al., 2013). However, Fgf4 is a marker specific to the naïve state and consistently exhibited higher expression in POU5F1 supported lines.

(5) GATA family

Zinc finger transcription factor GATAs play roles in early lineage differentiation. Gata2 and Gata3 are found in trophoblast cells and important for placenta functions (Ma et al., 1997), while Gata4 and Gata6 are upregulated during primitive endoderm differentiation and later are also found in parietal endoderm (Cai et al., 2008). From the microarray, *Gata2* and *Gata3* were expressed at the similar trend with *Cdx2*, upregulated in the POU5F3 line and more extensively downregulated in coelacanth POU5F1 supported cells. Other GATA genes were not significantly different among rescued lines, although we did observed Gata6 protein expression in immunofluorescent staining.

Section 3.7 The origin of POUV protein

Class V POU proteins are found across vertebrates in two forms POU5F1 and POU5F3. I have provided insights to evolutionary rate and conserved activities of those proteins in section 3.1-3.6. In this section I explore the history of these two paralogs and attempt to identify a single ancestral protein. The candidate ancestor of *POU5F1* and *POU5F3*, that I will refer to here as *POU5F1/3*, should be a five-exon gene encoding a protein similar to both POU5F1 and POU5F3. It has been shown recently that both POUV forms are also present in little skate (*Leucoraja erinacea,* one of chondrichthyes) (Frankenberg and Renfree, 2013). Thus, it seems that gene duplication leading to the emergence of *POU5F1* and *POU5F3* occurred at least as early as gnathostome ancestor. However, it is not clear how old this duplication is. The lamprey, a jawless vertebrate split from the ancestor of the gnathostomes around 550-650 MYA (Blair and Hedges, 2005; Kumar and Hedges, 1998). Did the *POUV* duplication take place prior to this split? To address this, I attempted to identify a *POUV*-related gene in jawless vertebrate (cyclostomes). There are two subclasses of cyclostomata: hagfish and lamprey. However, only the genome for the lamprey is publically available. For lamprey there are two genomes that have been sequences from both the arctic lamprey (*Lethenteron camtschaticum*) and sea lamprey (*Petromyzon marinus*). Noteworthy, arctic lamprey has several synonyms including *Lampetra camtschatica*, *Lampetra japonica*, and *Lethenteron japonicum*. The arctic lamprey and Japanese lamprey are indeed the same species (Potter et al., 2014).

In order to find the candidate *POUV* gene in jawless vertebrate, several gnathostome POU5F1 and POU5F3 DNA/protein sequences were used in BLAST searches against the available lamprey genome assembly database named LetJap7.0, which was derived from the testis of Japanese lamprey (*Lampetra japonica*). The BLAST search identified numerous

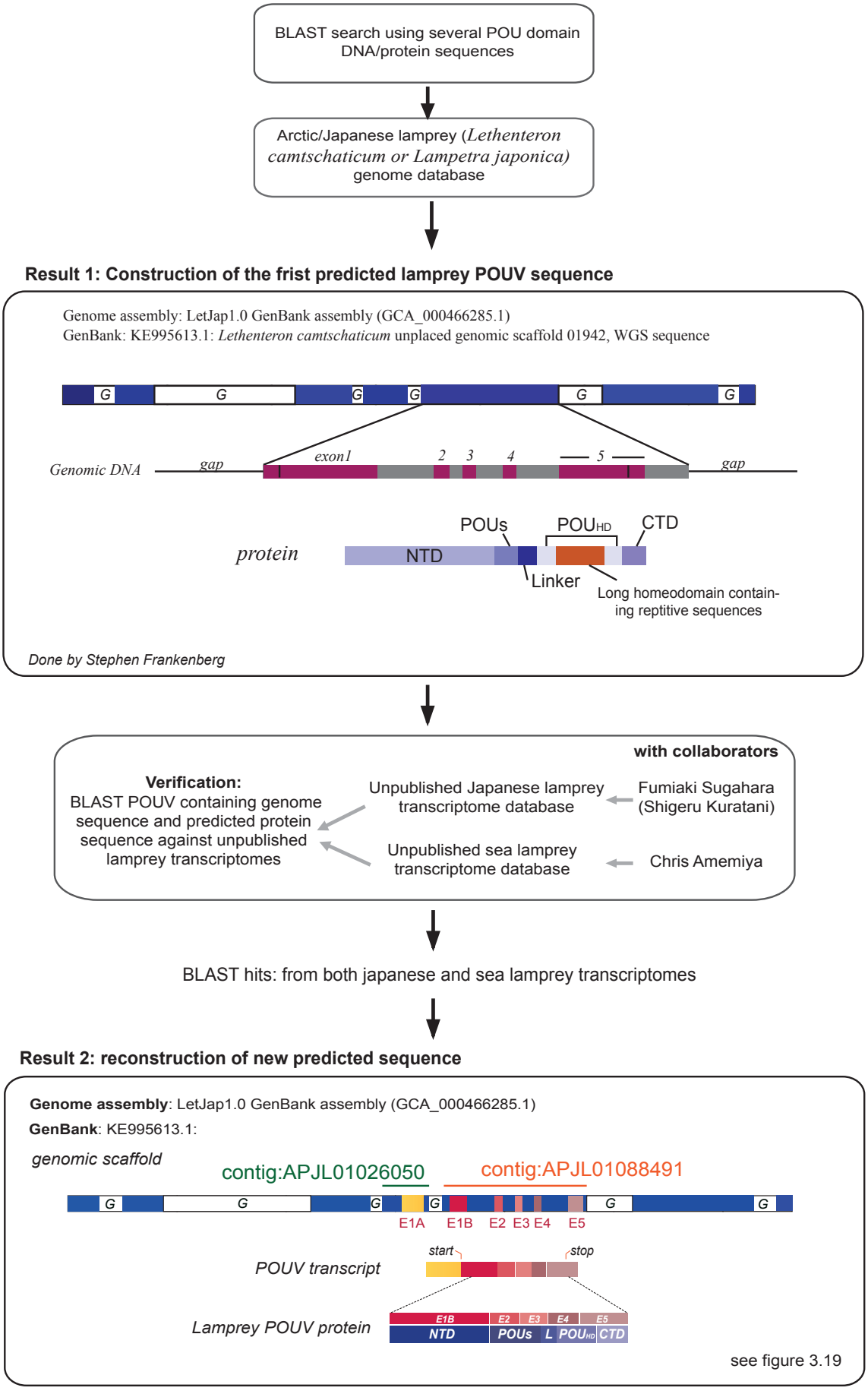


Figure 3.18 Work flow of lamprey *POUV* gene identification
Abbreviation: E, exon; NTD, N-terminal domain; POUs, POU specific domain; L, linker; POU$_{HD}$, POU homeodomain; CTD, C-terminal domain; aa, amino acids; G, Gap region (unknown DNA sequence)

sequences with some similarity to vertebrate POU domains, but most of them are other homeodomain containing proteins and some are POU3 (POU3 genes contain only one exon). However, there is only one genomic region or contig named APJL01088491 containing some sequences similar to that of POU specific and homeodomain and importantly with five-exon structure, suggesting a potential class V POU protein. These extensive searches were originally done by Stephen Frankenberg (University of Melbourne). To clarify the identity of this region, we initiated collaborations with Fumiaki Sugahara (RIKEN institute, Japan) and Chris Amemiya (Benaroya Reseach Institute, Seattle), and searched the predicted *POUV*-carried genomic region against unpublished transcriptome databases of Japanese lamprey and sea lamprey. Both transcriptomes lead to the identification of partial transcripts that were useful in constructing a complete full-length lamprey transcript sequence, which was then used to construct exon-intron structure on the contig. This strategy of lamprey POUV identification is illustrated in figure 3.18.

Indeed there are distinct five exons that encode the full-length protein. The alignment of these sequence tags to the lamprey contig is shown in Figure 3.19 and suggests the existence of an additional exon, which contain 5' untranslated region (5'UTR), in another contig named APJL01026050. This additional exon is referred to here as exon 1A and the second exon containing start site is referred as exon 1B (E1B). Similar to jawed vertebrate POU5F1 and POU5F3 genes, the first exon (E1B) containing the start site encodes for only the N-terminal domain. Exon 2 to 4 encode for the POU domain. Exon 5 contains coding sequence for C-terminal domain and 3' untranslated region (3'UTR). In contig APJL01026050, there are also possible promoters located upstream of exon1A. Thus, lamprey POUV protein is encoded by 6 exons and appears the only gene present. The exon-intron organization of *POU5F1/3* is similar to *POU5F1* and *POU5F3* genes (Figure 3.20)

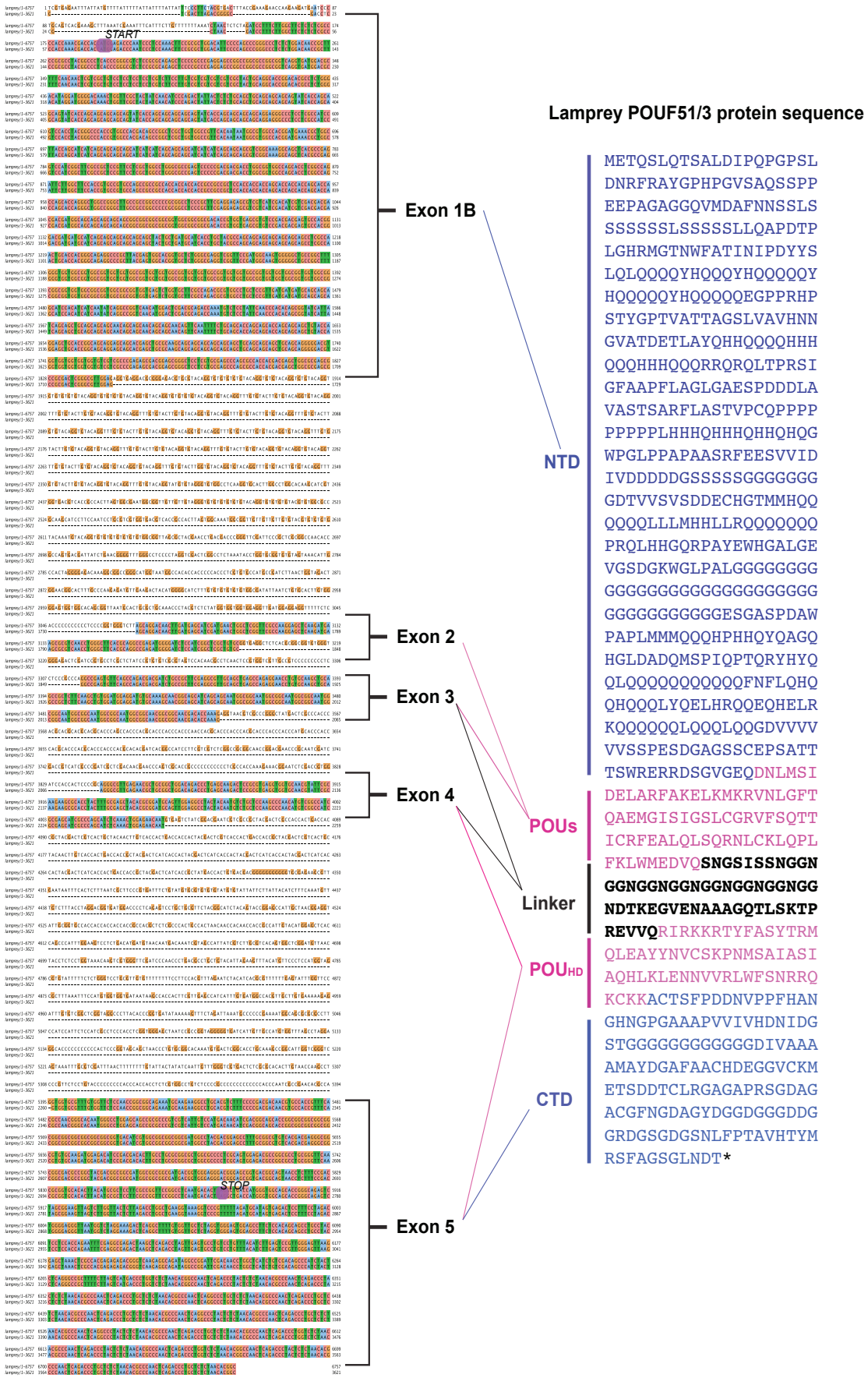
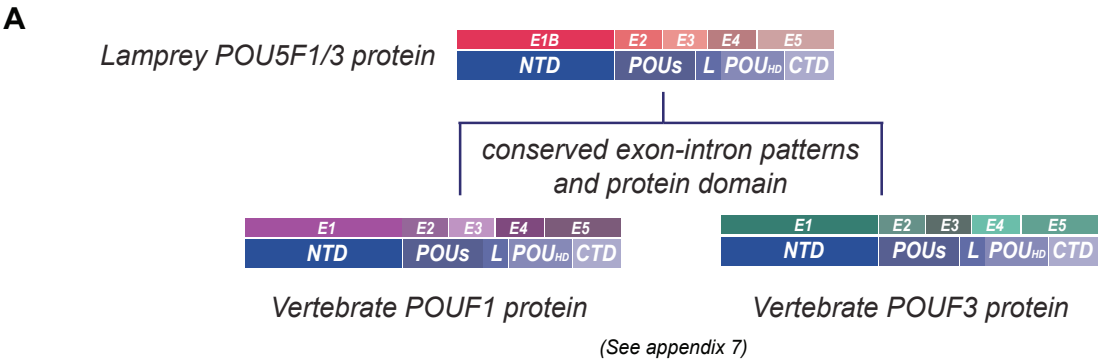


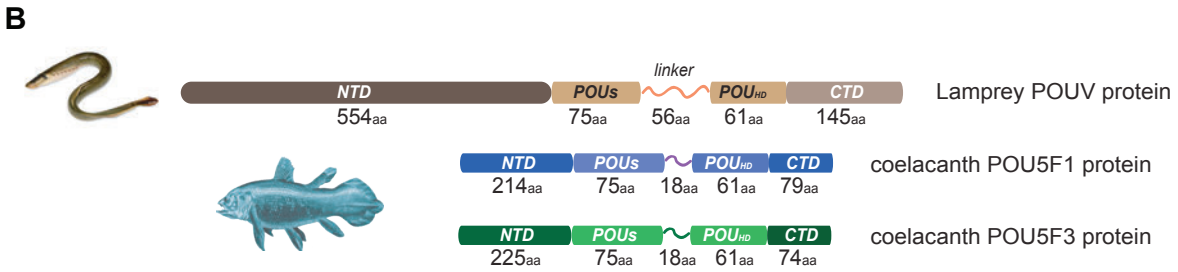
Figure 3.19 Exon-intron structure of *POUV* gene in lamprey

Alignment of Japanese lamprey genomic regions contig:APJL01088491 with predicted lamprey POUV sequnces obtained from the transcriptomes (on the left) and predicted protein sequence (on the right) is shown. The alignment is visualized by JALVIEW. Abbreviations: NTC, N-terminal domain; POUs, POU specific domain; POU_{HD}, POU homeodomain; CTD, C-terminal domain.

The complete full length of *POUV* transcript enables us to construct a predicted protein coding sequence. As in figure 3.20, coelacanth POU5F1 and POU5F3 are used as the representative of vertebrate POUVs. The length of POUs (75 amino acids) and POU_{HD} (61 amino acids) of lamprey POUV protein are exactly the same as those of vertebrate POUV proteins. Whereas, length of the linker, NTD and CTD in Lamprey are much longer than those observed in other vertebrates. Based on an alignment of the POU specific domain and POU homeodomain, the lamprey protein contains conserved sequence representative of neither POU5F1 or POU5F3, as the identifying sequences described in Figure 3.1-3.4 match neither protein. There is also a very large linker domain that does not exist in either POU5F1 or POU5F3 that contains a tripepetide repeat of NGG (Figure 3.20).



(See appendix 7)



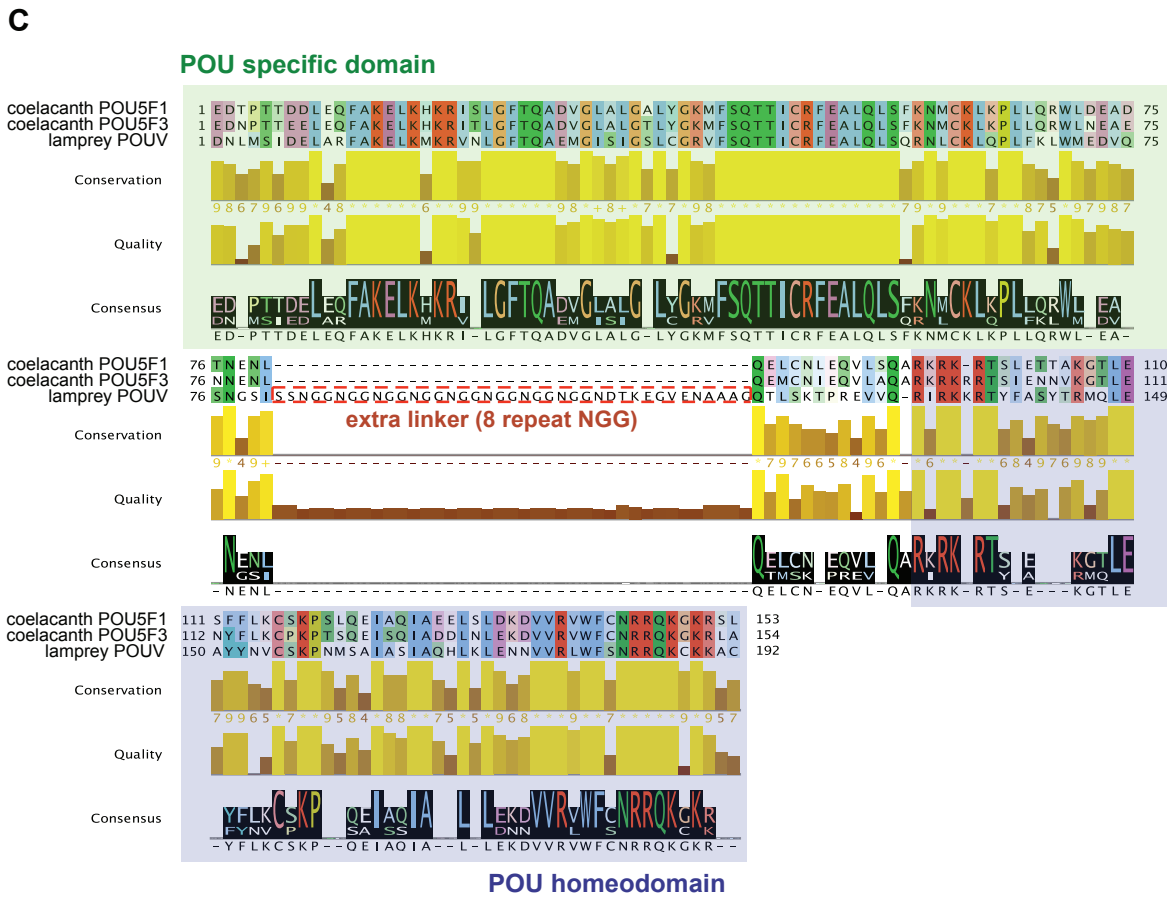


Figure 3.20 Characteristics of predicted lamprey POU5F1/3 protein

A) Predicted lamprey POUV exon-intron structure and corresponded protein region is similar to both gnathostomePOU5F1 and POU5F3 genes and proteins. B) Detail of lamprey POUV protein compared to coelacanth POU5F1 and POU5F3 proteins. C) An alignment of POU domains of POUV proteins was done by using JALVIEW. Green box marks POU specific domain, blue box marks POU homeodomain and unmark region represents linker. Red box indicates the linker part of lamprey POUV protein which is not found in coelacanth POU5F1 and POU5F3 protein. Abbreviation, E, exon, NTD, N-terminal domain, POUs, POU specific domain, L, linker, POUHD, POU homeodomain, CTD, C-terminal domain, aa, amino acids

Section 3.8 Discussion

In this chapter I have shown that the central pluripotency regulator Oct4 is conserved, both at the level of sequence and functionally. I have characterised the rate of POUV protein evolution for both POU5F1 and POU5F3 and shown that this rate is particularly enhanced in lineages that have lost one of these paralogs. In species retaining both POU5F1 and POU5F3, I found that the capacity of Oct4 to support the ESC gene regulatory network appeared particularly well conserved in the POU5F1 family and appeared to correlate with expression of this gene in the germ line. POU5F3 genes appeared to support something closer to primed pluripotency and this may correlate with expression in the epiblast at gastrulation. Finally, I present evidence for an ancestral POUV protein, a single gene in lamprey that does not fit in either class.

Previous work on the capacity of POUV proteins to support murine ESCs suggests a varied degree of POUV capacity to rescue Oct4 phenotypes. Based on data in *Xenopus*, it has been suggested that this activity might correlate with expression pattern of different POUV proteins in embryonic development. Thus, as shown in Morrison and Brickman, 2006) Xlpou91, a germ specific POUV protein, is better at rescuing Oct4-null ESC cells than Xlpou25, which is an epiblast specific POUV protein. These initial observations lead us to the suggestion that there were two independent functions of Oct4, one that is specific to germ/naïve ESCs and the other is specific to adhesion in the epiblast (Livigni et al., 2013). In my thesis I have examined the capacity of wide range of POUV proteins to rescue Oct4-null murine ESCs and found in species that possess two POUV proteins, this sort of segregation of function is observed. While POU5F1 appears specific to naïve ESC culture and germ cell related pluripotency, the loss of POU5F1 and further duplication of POU5F3 allowed rapid

diversification and the evolution of a specialised germ cell function, as seen in the case of POU5F1-like Pou5f3 protein (*Xenopus* Pou91).

Mouse Oct4 is expressed in the inner cell mass, the germ cells and epiblast during gastrulation. As a result, Oct4 appears to regulate different aspects of pluripotency. Interestingly, the Oct4 promoter actually contains distinct regulatory regions, one that controls expression in naïve/germ cell pluripotency, and the second, that regulates expression in the epiblast (Ovitt and Scholer, 1998). In a number of species that retain both paralogs that go back to the beginning of the jawed vertebrates, it is clear that this function remains segregated. This is particularly interesting as the mechanism that regulates naïve and primed (epiblast) pluripotency are different, although both involve Oct4. For example, naïve pluripotency depends on the inhibition of FGF/ERK signaling, while primed pluripotency requires it. Primed pluripotency is dependent on Nodal-related TGF-β signaling, while naïve cells rely on LIF (Nichols and Smith, 2009). Perhaps a closer inspection of both POU5F1 and POU5F3 paralogs will help resolve the aspects of the network downstream of Oct4 that is responsible for interacting with these signaling pathways.

The finding that POU domains have changed dramatically at the birth of eutherian radiation, I hypothesise further that POU5F1 in eutherian ancestor might have evolved to help specializing novel structure called inner cell mass as discussed above, where naïve ESCs can be derived. One interesting aspect of this study is the strong implication that POU5F1 proteins regulate germ cell specification and that this is related to naïve pluripotency. This suggests that the relatively new mechanism for pre-implantation development evolved from the gene regulatory network employed during primordial germ cell specification. This idea is consistent with the observation that ESCs and EGCs share gene expression profile and epigenetic status in common and any gene expression differences between them reflect the culture conditions (Leitch et al., 2013).

As there is evidence of POUV linked to germ cell network, how can we apply this POUV activity to the evolutionary model to the germ cell mode specification varied among vertebrate branches? First of all, there are two modes of germ cell specification: predetermined mode and or inductive mode. Predetermined modes of germ cell specification rely on the localization of maternally inherited determinants to drive germ cell specification. Inductive modes for germ cell specification employ integrative signaling from surrounding tissues for the specification of primordial germ cells during gastrulation (Extavour, 2003). In vertebrates, some cartilaginous fish, axolotls, turtles and mammals have inductive modes of germ cell specification. Other vertebrates like actinopterygian fish, frogs and archosaurs (birds and crocodiles) have predetermined mode of germ cell specification. While the inductive mode of germ cell specification is highly conserved and may have been the ancestral mechanism of germ cell specification, predetermination appears a secondary derived trait that exhibits a high level of variability (Bachvarova et al., 2009; Johnson et al., 2003; Saito et al., 2014). All vertebrate taxa with predetermined mode have lost *POU5F1* during its evolution. My study here showed that there is a link between POU5F1 and the inductive germ cell/reproduction network, in particular Prdm14 and other downstream germ cell networks. When *POU5F3* is lost, some level of primed pluripotency is presumable lost and gastrulation could become constrained, when *POU5F1* is lost, naïve/germ cell pluripotency is lost, necessitating the re-acquisition of germ plasm and free organisms from the structural constraints of making germ cells during embryonic development.

The expansion of genome by duplication clearly enables the diversification of protein function in evolution. There are several fates that the duplicated genes can experience including non-functionalization (loss of functionality), neofunctionalization (acquisition of new function) and subfunctionalization (partitioning of subfunctions into different duplicated genes) (Cañestro et al., 2007). From this study, I found that duplicated *POUV* genes (*POU5F1* and *POU5F3*) appear to have undergone partitioning of

subfunctions. Here I provide a model of what might happen during the evolution of vertebrate POUV activity. This model is summarised in figure 3.21. Based on the finding that there is only one *POUV* gene found in the lamprey genome with an exon-intron structure similar to that of jawed vertebrate genes *POU5F1* and *POU5F3*, this suggests that the origin of duplication event giving rise to these genes from a single ancestral *POUV* gene might be occurred after the divergence of jawless vertebrate lineage from the jawed vertebrates around 550-600 million years ago. Studies on chordate evolution suggest that it involves at least two rounds of whole genome duplication (called 2R model). The first round (1R) occurred in a lineage of protochordates and maybe linked to the emergence of jawless vertebrates, and the second round (2R) occurred in an ancestral cyclostome lineage and maybe linked to the emergence of gnathostomes (Dores, 2011). The nature, timing and significance of this second round is still controversial. Our finding that there is only single *POUV* in an extant jawless vertebrate give the impression that this second round of WGD might probably occurred around the emergence of gnathostomes and might be responsible for the birth of *POU5F1* and *POU5F3*.

In my model, these dual functions might already present in jawless vertebrate ancestor and later after duplication this dual functions has been segregated between POU5F1 and POU5F3. It has been shown that Oct4 dose is sensitive to development, as shown in mouse (Radzisheuskaya et al., 2013). The dose of other vertebrate POU5F1 and POU5F3 might also influence their development. This might be why most of vertebrates have lost either *POU5F1* or *POU5F3* gene or have undergone complete POUV functional segregation to remove redundancy, which might provide benefit for developmental programs. I would like to suggest that while gene duplication enables the diversification of regulatory networks, these networks are not so different that these POUV proteins can't regain the capacity to regulate different stages in development. More recently, POU5F3 in axolotl or tammar shows some capacity to support Prdm14 expression and therefore is perhaps showing a

plasticity of function that maybe able to eventually compensate for potential gene loss. This loss of one POUV paralog appears to occur frequently in evolution, while the re-duplication of *pou5f3* in frog, appears only once in to my knowledge. As a result, it would seem that there is selective advantage to losing one of the POUV paralogs.

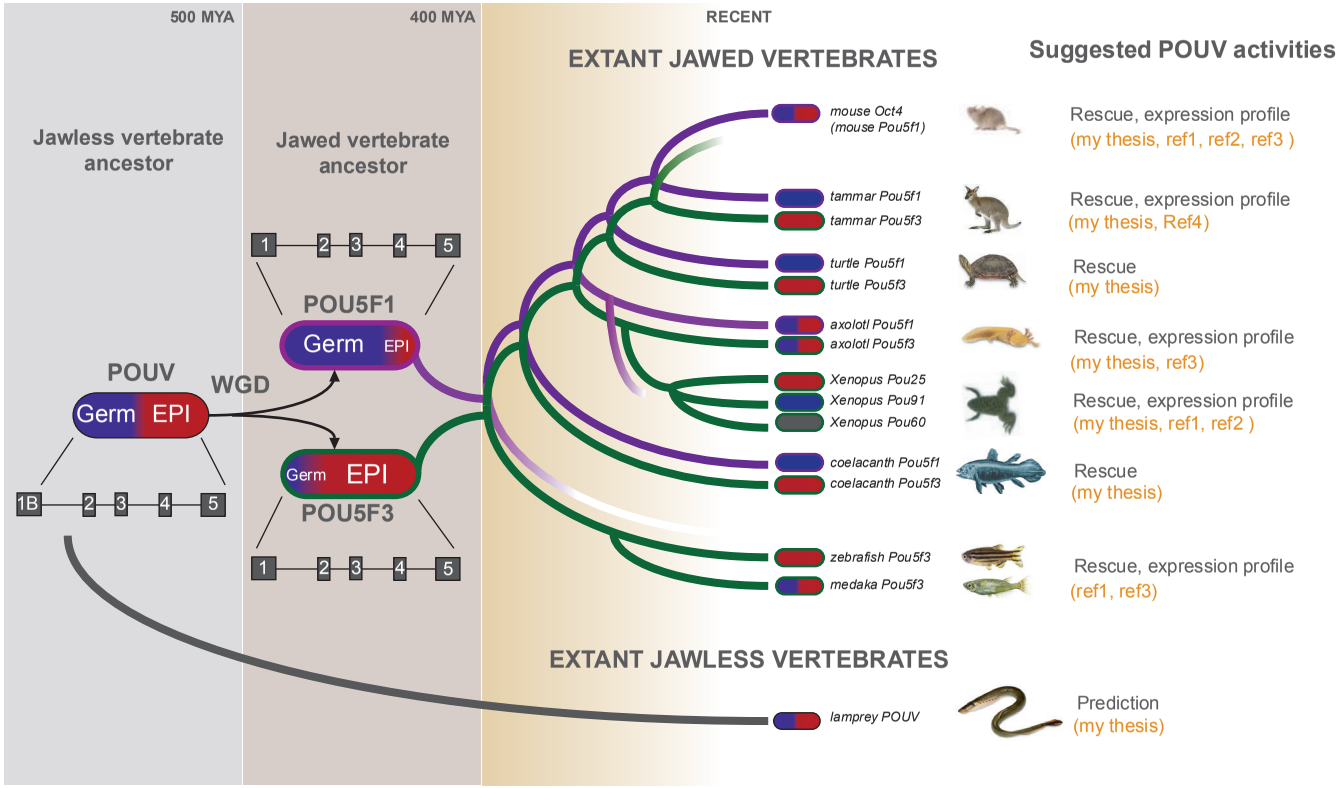
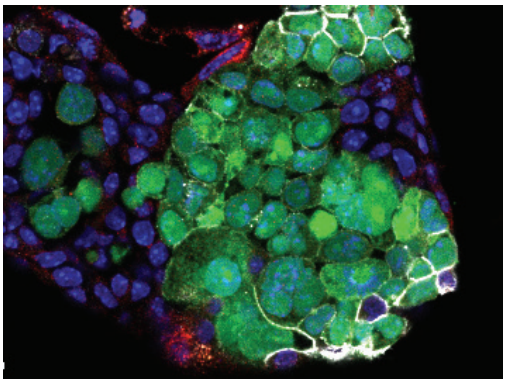


Figure 3.21 Model summarizing on the evolution of POUV homologs' activities.
The model is described in section 3.8 Discussion. Abbreviations: Germ, Germ cell like activities; EPI, epiblast like activity; WGD, whole genome duplication. The purple and green lines of vertebrate tree represent the fate of POU5F1 and POU5F3 genes, respectively. Ref1, Morrison and Brickman 2006; Ref2, Livigni et al 2013; Ref3, Tapia et al 2013; Ref4, Frankenberg et al 2013.



CHAPTER 4

CONSERVED ROLES OF GERM CELL SPECIFIC AND EPIBLAST SPECIFIC POUV ACTIVITIES IN INDUCING PLURIPOTENCY

CHAPTER 4

CONSERVED ROLES OF GERM CELL SPECIFIC AND EPIBLAST SPECIFIC POUV ACTIVITIES IN INDUCING PLURIPOTENCY

Introduction

Pluripotency is the ability of cells to give rise to all somatic lineages including germ cells (the origin of sperm and oocyte). The mouse embryo contains pluripotent cells including inner cell mass (ICM) cells and Oct4-positive gastrulation-stage epiblast cells. They can be isolated and maintained as pluripotent stem cells *in vitro*, called embryonic stem cell (ESCs) and epiblast stem cells (EpiSCs) respectively. In the last decade, there have been successful attempts to reprogram differentiated, mature somatic cells to these pluripotent, embryonic states. The initial efforts in reprogramming were inspired by the success in generating tadpoles from differentiated intestinal epithelial cells (Gurdon et al., 1958) and later the success in nuclear transfer in mammalian cloning (Wilmut et al., 1997). As a result several laboratories attempted to reprogram somatic cells by transferring somatic nuclei to an oocyte or ESC (reviewed in (Hochedlinger and Jaenisch, 2006)). Later in 2006, Shinya Yamanaka and colleagues made another breakthrough by converting somatic cells to an ESC state through the introduction of embryonic transcription factors. They obtained so called induced pluripotent stem cells (iPSCs), by overexpression of a specific set of transcription factors that are normally expressed in ESCs. They examined several combinations of ESC-related transcription factors, and found one combination with the best potential in iPSC generation, Oct4, Sox2, c-Myc and Klf4, which together are known as Yamanaka factors (here I refer to this combination as “Yamanaka factor” or “OSKM”) (Takahashi and Yamanaka, 2006). Although some combined pharmaceutical inhibitors can be used in

reprogramming instead of Sox2, Klf4 and c-Myc, Oct4 appears to be required and essential for reprogramming (reviewed in (Radzisheuskaya and Silva, 2013)).

A form of reprogramming is also used in the generation of germ cells during development. During the formation of the germ cells, pluripotency is re-established as the rest of the embryo is progressing on in embryonic differentiation (Saitou and Yamaji, 2010). In vertebrates, there are two basic modes of germ cell specification: inductive (epigenesis) and predetermined (preformation). The inductive modes employ integrative signaling from surrounding tissues for specifying primodial germ cells (PGCs) during gastrulation. Predetermined modes of germ cell specification depend on the localization of maternally inherited determinants (Extavour, 2003; Johnson et al., 2003). In mouse embryos, the inductive mode of specification takes place in a small cluster of epiblast cells located in the posterior proximal region of the embryo. Cells in this region experience the highest level of BMP signaling in the epiblast and as a result are induced to become PGCs (Arnold and Robertson, 2009). In contrast, the germ cell program of African clawed frog (*Xenopus laevis*) is specified by maternal determinants and is therefore classed as predetermined mode (Kobayashi et al., 1998). Despite the differences in the modes of germ cell induction, both mouse and *Xenopus* embryos express Oct4 and its homolog in their germ cells (Hinkley et al., 1992; Sabour et al., 2010). Does this mean that Oct4 plays a similar role in predetermined or inductive germ cell specification? Based on studies of cellular reprogramming, it appears that Oct4 is one of the key factors driving the acquisition of pluripotency and that a number of vertebrate homologs of Oct4 have the capacity to support pluripotency in mouse. In particular, the *Xenopus* Oct4 homolog expressed in germ cells, Pou5f3.1 (also referred to here as Xlpou91) is as effective as mouse Oct4 in ESC complementation assays and can also substitute for Oct4 in murine and human reprogramming (Morrison and Brickman, 2006; Tapia et al., 2012).

In chapter 3, I have shown that Xlpou91 has the highest rescue index of Oct4 null ESC cells among POU5F3 proteins, similar to all POU5F1, whereas all other POU5F3 lack this capacity. There is another Oct4 homolog in *Xenopus* embryos that has a rescue activity similar to other POU5F3 proteins, Pou5f3.2 (Oct25 or referred to here as Xlpou25). In the *Xenopus* embryo, Xlpou25 is highly expressed in gastrula stage embryos (Hinkley et al., 1992). (Livigni et al., 2013) has shown that Xlpou25 is a regulator of adhesion required for *Xenopus* gastrulation, and it has also been shown that Xlpou25 can induce enhanced expression of E-cadherin in Oct4-null lines. This suggests that one of the Oct4 roles in primed state pluripotent epiblast is the regulation of cell adhesion. As a result it appears that in *Xenopus*, where the *pou5f3* gene was duplicated, two of these genes were specialised to perform specific subsets of Oct4 function. Thus, the two Oct4 homologs have distinct activities in germ cells/ naïve pluripotency and epiblast/primed pluripotency. In mouse, both of these functionalities are encoded in a single factor, Oct4, while in *Xenopus* these functions has been distributed between Xlpou91 and Xlpou25 during evolution. Thus, Xlpou91 is a germ cell regulator, while Xlpou25 is an epiblast regulator. In this chapter, I ask how the epiblast and germ cell activities of Oct4 influence the cellular reprogramming or induction of pluripotency, by exploiting the specialization that occurred in *Xenopus*.

To address the relationship of developmental activity and reprogramming, I asked about the relative abilities of Xlpou91 or Xlpou25 to replace mouse Oct4 in the induction of pluripotency during reprogramming. Using Nanog-green fluorescent protein (GFP) mouse embryonic fibroblast cells (MEFs) in which GFP has been placed under the control of the *Nanog* promoter (Chambers et al., 2007), the acquisition of pluripotency and reprogramming dynamics could be monitored and scored. I found that similar doses of mOct4 and Xlpou91 could induce pluripotency, whereas higher doses of Xlpou25 were required to achieve cellular reprogramming. I then further characterised iPSCs derived from these experiments. Based on the examination of other ESC/germ cell surface markers together with Nanog-

GFP, I found that cells reprogrammed with Xlpou25 eventually reached an "iPSC-like," state, but that it was much less stable than that achieved with either Xlpou91 or Oct4, and contained large numbers of differentiated cells. Based on the transcriptomes of the sorted iPSC sub-fraction of these cultures, I could confirm that all derived iPSC lines contained cells that exhibited naïve pluripotency, although there were some differences. I found that iPSCs derived from Xlpou91 and mOct4 expressed genes related to reproduction that were reduced in Xlpou25 iPSCs. Consistent with these observations, I found that Xlpou91 and mOct4 induced pluripotency by similar routes, whereas Xlpou25 promoted reprogramming via the induction of developmental differentiation programs first. Taken together, my observations suggest that direct and efficient reprogramming to naïve pluripotency may be related to the induction of germ cell programs.

Section 4.1 Optimization of induced pluripotent stem cells (iPSCs) generation

There are several methodologies for the generation of iPSCs. These include the introduction of the four factors by retrovirus, transposon, episome, as well as the addition of small molecules that can improve the reprogramming process (Hou et al., 2013; Kaji et al., 2009; Shi et al., 2008; Takahashi and Yamanaka, 2006; Yu et al., 2011). The transition from MEFs to iPSCs takes time from several days to weeks. In general, reprogramming efficiencies are low, approximately 0.01-0.001% of cells exposed to the four factors. To achieve the best cellular reprogramming, we used several experimental strategies to increase the efficiency of induction before proceeding to examine different POUV homologs' activities.

(a) The choice of cell line to monitor the acquisition of reprogramming?

To monitor the acquisition of pluripotency, different pluripotency-related markers have been used, in particular fluorescence-tagged Oct4 or Nanog, cell surface staining for SSEA-1, and alkaline phosphatase staining. Single cell analysis of gene expression during reprogramming has revealed that endogenous Oct4 is also one of the early markers of the process, with its expression originating as soon as cells reach an early or partially reprogrammed state. Nanog comes up later in the reprogramming process and better marks fully reprogrammed cells (Buganim et al., 2012). Moreover, even within the Nanog positive population there is a range of Nanog expression (these can be sorted with the Nanog-GFP reporter) and not all Nanog positive cells are fully reprogrammed. However, with the addition of two further cell surface markers, a stable fully iPSC population can be identified based of the cell phenotyping as $Nanog^{+}ICAM^{+}CD44^{-}$ profile (O'Malley et al., 2013).

Hence, I used Nanog as the marker to monitor the reprogramming efficiency and did molecular analysis on the $Nanog^{+}ICAM^{+}CD44^{-}$ population. The detailed method for iPS generation I used is described in chapter 2 section 2.8. Briefly, (Figure 4.1A) I used MEFs derived from mice carrying a Nanog-GFP reporter allele (Chambers et al., 2007). Nanog-GFP MEFs at passage 1 were used for transfection with retrovirus. The derivation of MEF lines is described in chapter 2. The induced cells were seeded onto irradiated feeders and cultured in defined iPSC medium for 24-27 days before iPSC colonies were picked and expanded. The acquisition of pluripotency was monitored and scored through the appearance of Nanog-GFP colonies.

(b) The choice of retrovirus and their doses for reprogramming

I chose to use a retrovirus-based strategy for the induction of reprogramming because we could easily manipulate the identity and dose of transcription factors used in reprogramming. Each transcription factor was encoded by a cDNA subcloned under the retroviral promoter of pMXs-expression system. The retrovirus were produced by 293LTV cell lines, and the doses of the retrovirus production were quantified by qRT-PCR using primers specifically detecting the retroviral RNA backbone. As shown in figure 4.1B, the retrovirus produced from different plasmids expressed different levels of transcription factor encoding message. Following quantification of the amount of transcription factor expression, I determined the optimal infection efficiency. Retrovirus carrying the dsRED gene was used to monitor infection. I found that retrovirus doses of 10^9 copies and $5X10^9$ copies led to infection efficiency of >70% and >96% respectively, as judged by flow cytometry analysis of dsRED expression level (Figure 4.1C).

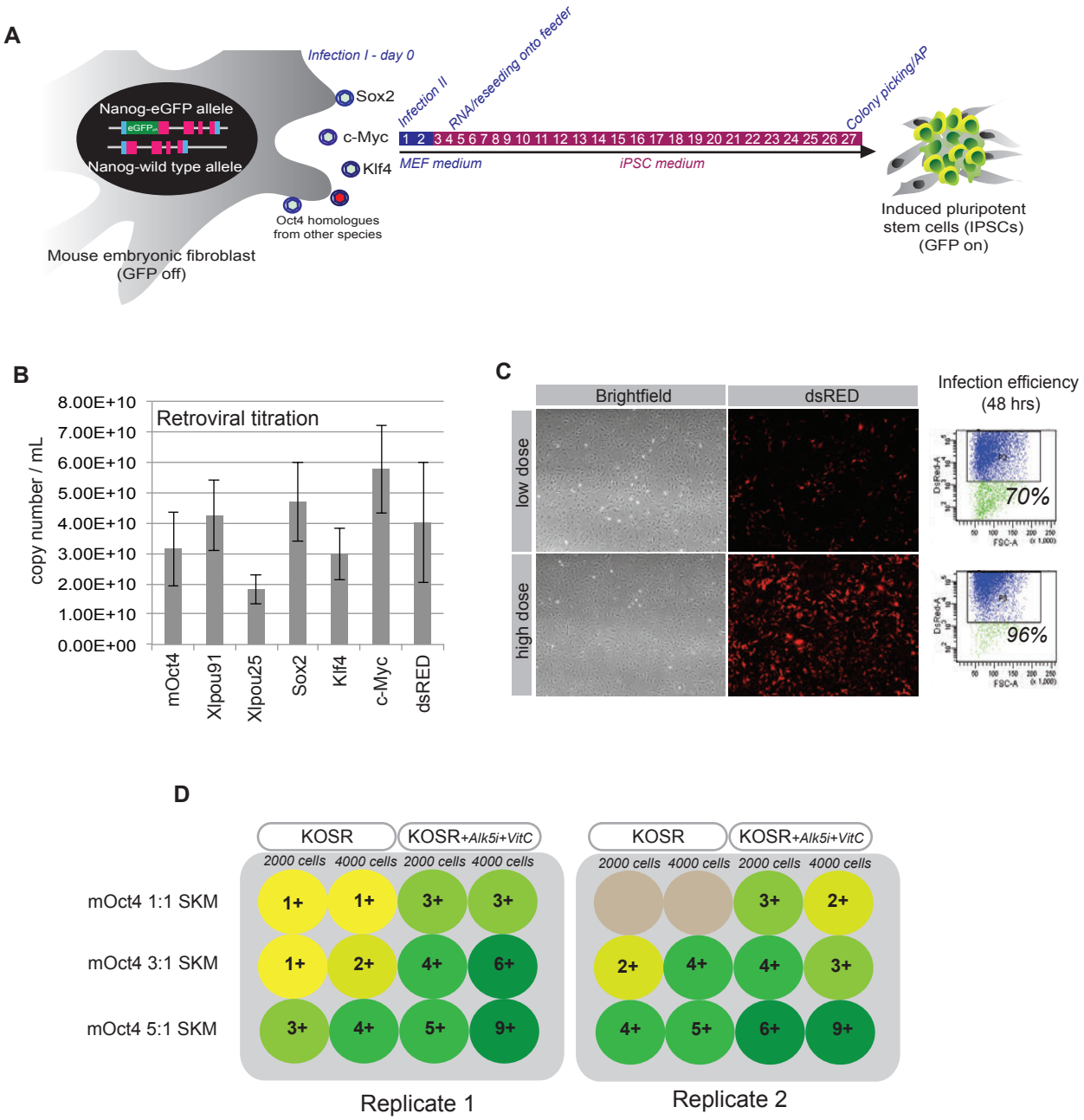


Figure 4.1 Optimization of induced pluripotent stem cells (iPSCs) generation A) Schematic illustration of strategy of iPSC generation. Nanog-GFP mouse fibroblast cells at passage 1 was used as a source of somatic cells for cellular reprogramming.The exogenous expression of Oct4, Sox2, Klf4 and c-Myc was transduced by ecotropic retrovirus. The numbers indicate number of days after the first infection. B) Before infection, retrovirus doses were measured by qPCR. Data represents two independent retrovirus productions. C) Infection efficiency was confirmed by fluorescence microscopy to detect dsRED expression. After 48 hours, the dsRED infected MEF cells were harvested and analysed by flow cytometry. Low dose (low) and high dose (high) indicates the amounts of dsRED/Oct4 homologues-carrying retrovirus approximately 1XE9 copies and 5XE9 copies, respectively. D) Nanog-GFP MEF cells were infected with different doses of Oct4 and constant dose of Sox2, Klf4 and c-Myc. After four days post the first infection, the induced cells were seeded onto feeders with two clonal densities: 2000 cells/well and 4000 cells/well, and the induced cells were maintained in different iPSC induction mediums: “KOSR” means GMEM+KOSR+LIF and “KOSR+Alk5i+VitC” means GMEM+KOSR supplement with Alk5i inhibitor and Vitamin C (see section 4.1c). The Nanog-GFP appearance can be detected from day 10 onward after the first infection. The numbers in color circles indicate the total number of Nanog-GFP appearance at day 15.

(c) The choice of culture medium for iPSC derivation

Yamanaka and colleagues originally used ESC medium containing fetal bovine serum (FBS) supplemented with LIF for their reprogramming experiments (Takahashi and Yamanaka, 2006). We also tried this ESC condition and could not achieve reprogramming. This could be due to variation between FBS batches. As a result I tried a number of different media conditions for iPSC generation including:

(a) ESC medium: GMEM + FBS + LIF

(b) ESC medium: GMEM + FBS + LIF + Alk5 inhibitor + Vitamin C

(c) DMEM high glucose + 20% KOSR + LIF

(d) DMEM high glucose + 20% KOSR + LIF + Alk5 inhibitor + Vitamin C

These conditions were originally examined by Kumiko A Iwabuchi, in Keisuke Kaji's lab, University of Edinburgh and were based on the published data that the Alk5 inhibitor and Vitamin C can improve the induction of pluripotency (Esteban et al., 2010; Li et al., 2009). I tested these conditions with my retroviruses and found that there were no Nanog positive colonies formed when the reprogramming was done in either (a) or (b) conditions, but with (c) and (d) I observed the appearance of Nanog-GFP expression at day 10-15 after the first infection (figure 4.1D). Supplementing media with Alk5 inhibitor and Vitamin C improved the number of total Nanog-GFP colonies in both replicated experiments. I also observed that the number of GFP positive colonies was increased when higher Oct4 doses were used. From these colonies I was also able to generate stable clones of iPSC lines, which could maintain their Nanog-GFP during expansion. I concluded that the best condition for iPSC generation with my viruses and cell lines was DMEM high glucose with 20% KOSR supplemented with Alk5 inhibitor and Vitamin C. This was used for all iPSC experiments.

Section 4.2 Generation of induced pluripotent stem cells by different doses of *Xenopus* POUV proteins

After I optimised the conditions for iPSC generation (section 4.1). We analyzed the ability of different POUV proteins to induce pluripotency. Reprogramming experiments were performed with different doses of Oct4 and its homologs, alongside the same doses of Sox2, Klf4 and c-Myc. Thus, I tested the capacity of different doses of POUV protein alongside constant SKM dose by using the following parameters:

1) 10^9 copies of Oct4/Xlpou91/Xlpou25 retrovirus as **Oct4/Xlpou91/Xlpou25 low dose**
2) $5X10^9$ copies of Oct4/Xlpou91/Xlpou25 retrovirus as **Oct4/Xlpou91/Xlpou25 high dose**
3) 10^9 copies of each Sox2, c-Myc, and Klf4 retrovirus as a constant dose of SKM
(The infection strategy is described in chapter 2, section 2.8)

To verify that infected MEF expressed OSKM in expected doses, I collected infected cells at day 4 post the first infection and quantified the exogenous expression of these factors by qRT-PCR. Figure 4.2A shows that while there was some variation in levels of c-Myc, Sox2 and Klf4, these levels of variation are all within a tolerable range, less than 2-fold. Oct4/Xlpou91/Xlpou25 expression levels were proportional to the amount of retrovirus used, although Oc4 expression was slightly lower (20%) than the two *Xenopus* genes.

During the induction of iPSCs, I recorded the first appearance and number of Nanog-GFP colonies daily. I found that in all reprogramming experiments mOct4 high SKM always produced GFP positive colonies first at day 10-11 following the infection. Both Oct4 low and Xlpou91 low SKM emerged around the same time, around day 11-13. This contrasts

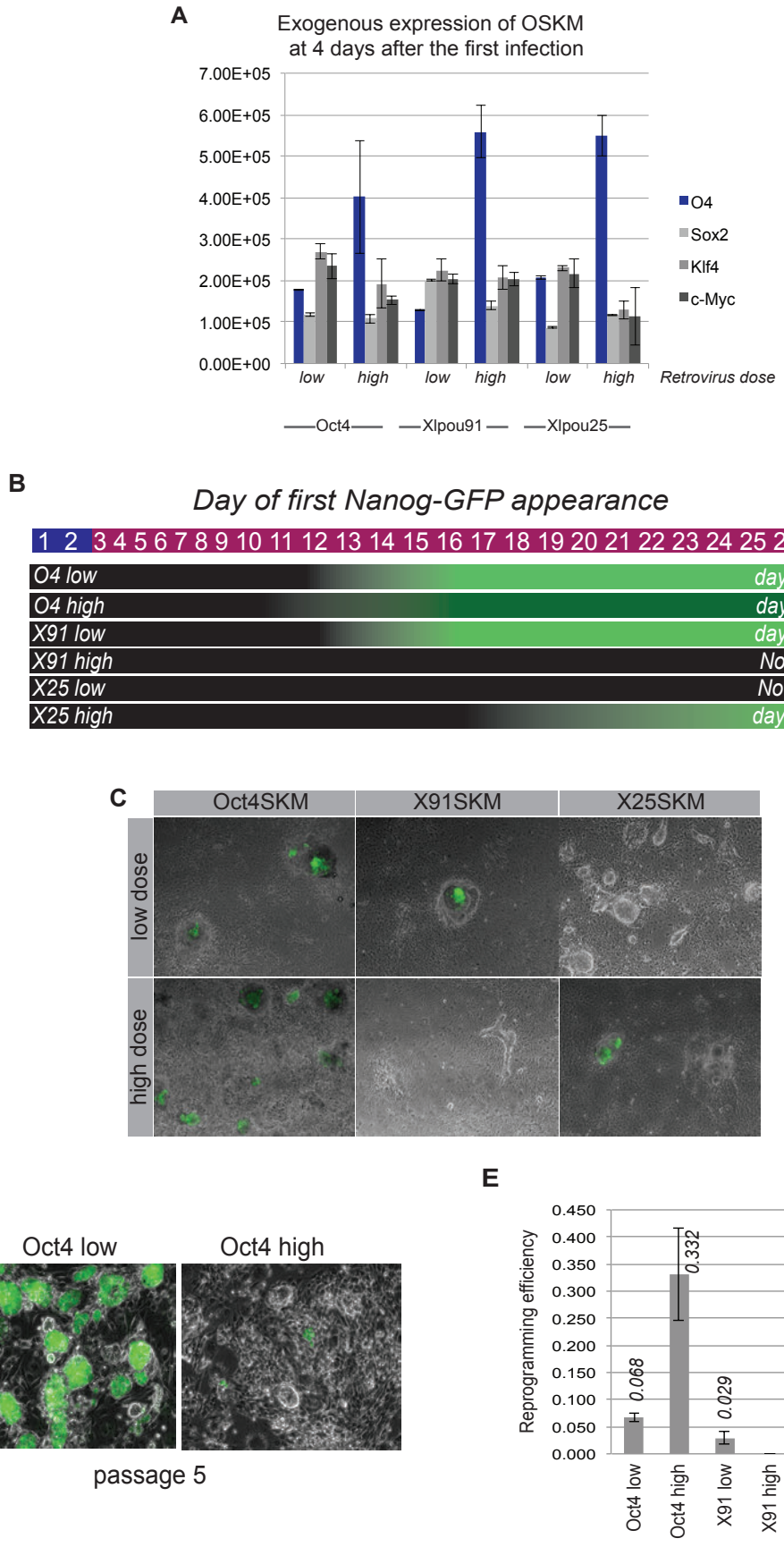


Figure 4.2 Generation of induced pluripotent stem cells (iPSCs) by different doses of *Xenopus* POUV proteins After infection for 4 days, exogenous gene expression was confirmed by qPCR. Data represents technical replicates of one experiment. B) The transduced cells were monitored daily and number of Nanog-GFP colonies were recorded. C) Merged images of brightfield and Nanog-GFP present iPSC colonies at day 24 post first infection. D) iPSC clones from Oct4 low and Oct4 high doses were picked and expanded for 5 passages on feeders. E) Reprogramming efficiency is calculated by dividing the number of Nanog-GFP colonies by the number of reseeded cells. Data represents three independent experiments.

with the activity of Xlpou25, which only produced Nanog positive colonies when infected at high levels and only at a later time point (day 16-21, see figure 4.2B). Xlpou25 low SKM and Xlpou91 high SKM both generated a lot of iPSC-like colonies, although none of them became Nanog positive (figure 4.2C). Moreover, while Oct4 high SKM was the most effective condition at inducing Nanog-GFP and could induce robustly GFP positive colonies (figure 4.2C), iPSCs formed with this dose of Oct4 were not stable, and could not maintain expression of the GFP over time in culture (Figure 4.2D). Consistent with the observations in ESCs that Oct4 overexpression can also induce differentiation, I also observed that while Oct4 high SKM induced Nanog expression at an early time (Figure 4C, D), this expression was unstable and Nanog expression was lost when these colonies were picked and expanded (Figure 4D). Taken together my observations show that there appears a correlation between POUV dose and reprogramming time, but establishing a stable reprogrammed line may be incompatible with high level POUV/Oct4 expression.

The reprogramming efficiency for each POUV protein was calculated by dividing the total number of Nanog-GFP colonies by the number of originally seeded cells (figure 4.2E). The most efficient was mOct4 high SKM (0.332%), although most Nanog-GFP colonies were not expandable as they did not maintain expression of the Nanog transgene. The reprogramming efficiency of mOct4 low SKM was about five fold lower (0.068%). Xlpou91 could also induce reprogramming at low dose, but the efficiency was significantly lower. The reprogramming efficiency of Xlpou91 was slightly higher than that of high dose Xlpou25, but like the Oct4 lines, these could efficiently maintain Nanog during expansion (see below).

Section 4.3 The derivation of clonal iPSC cell lines generated by different POUV proteins

The derivation of iPSC was performed using only the following conditions: Oct4 low dose, Xlpou91 low dose, and Xlpou25 high dose. In all text and figures, these will be referred as **mOct4 iPSC**, **Xlpou91 iPSC**, and **Xlpou25 iPSC**. These conditions were chosen based on the effective generation of Nanog positive colonies during reprogramming.

All Nanog-GFP colonies were picked and plated onto irradiated feeders and expanded until passage 6. Only clones retaining Nanog-GFP expression were kept for further analysis. At passage 6, I observed that colony morphology of most Xlpou91 clones showed less differentiation than the others. Most of Xlpou25 lines contained colonies with Nanog-GFP centres surrounded by differentiated cells. Interestingly, iPSC generated by mouse Oct4 low SKM produced a mix of both undifferentiated and differentiated clones in similar proportions (Figure 4.3A). I then assessed whether the established iPSC lines were maintained solely by endogenous Oct4. I screened for retroviral silencing by qRT-PCR on expanded clones at passage 6 and could not detect any residual expression of virally derived mOct4/Xlpou91/Xlpou25 in most clones, compared to the high level of those in day 4 transduced cells. Most Nanog-GFP clones derived from Oct4-mediated reprogramming could maintain GFP expression for several passages. Although exogenous Xlpou25 and 91 had been silenced, Xlpou91 induced clones had high proportions of Nanog positive, while Xlpou25 had both Nanog positive and Nanog negative (Figure 4.3, 4.9).

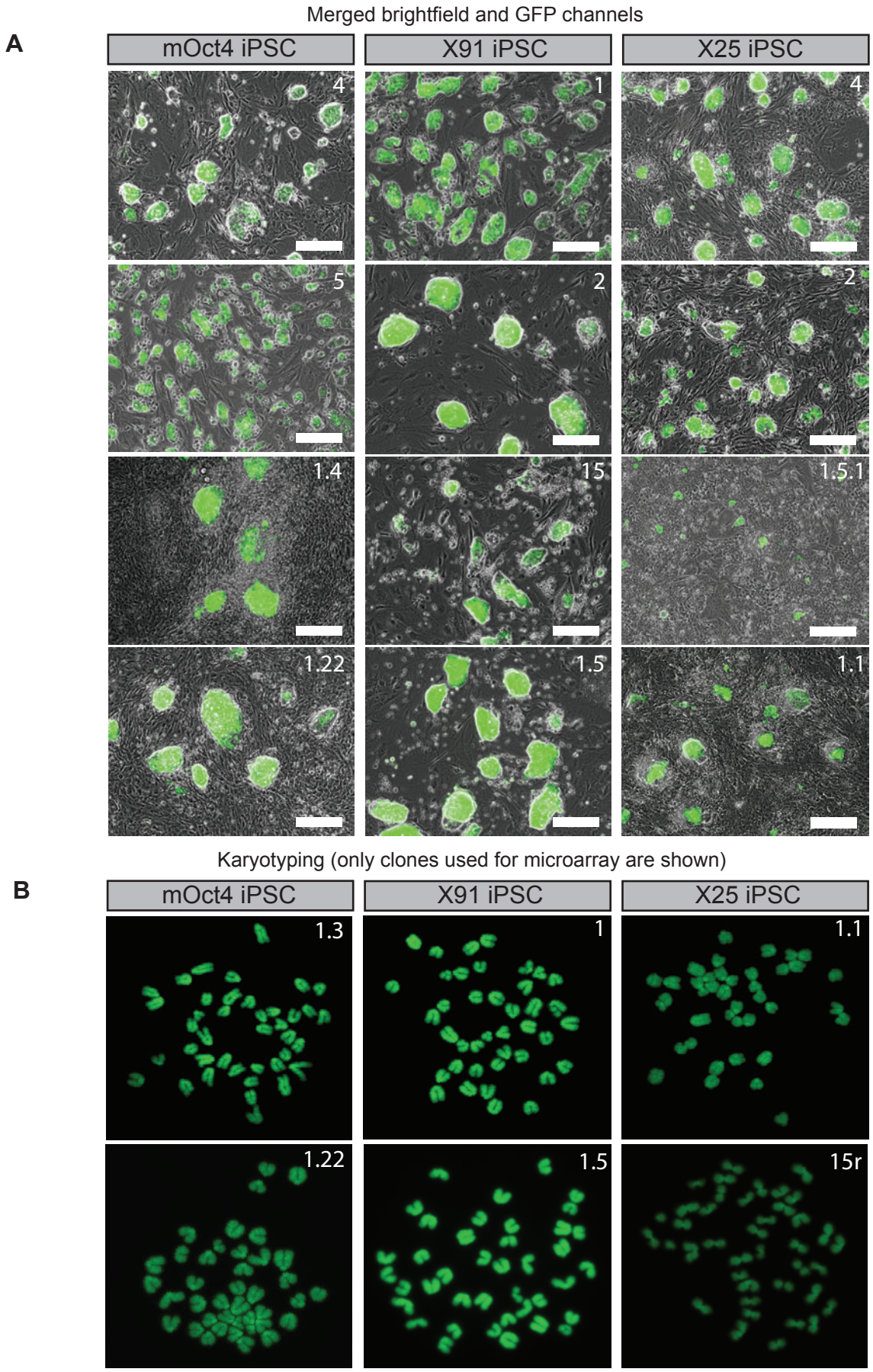


Figure 4.3 Derivation of iPSC clonal cell lines generated by Xlpou91 or Xlpou25 and karyotyping. A) Data shows the merged images of brightfield and GFP marking Nanog expression. All iPSC clonal lines shown here were derived from two independent experiments and expanded until passage 6 for imaging. B) Karyotyping was done for all derived clones (5 clones each) before further analysis. The chromosomes were stained with DAPI and at least 10 cells were imaged. Data shows only the result of clones used for global gene expression analysis by microarray.

During the retrovirus-based iPSC generation, the retrovirus integrate their reverse transcribed genome carrying an exogenous gene construct into the host (MEF) cell chromosomes. The high amount of retrovirus used for induction might potentially cause chromosome breaks. Moreover, the generation of iPSCs is a selective process and this process can also select for abnormal, growth advantaged, karyotypes. Thus, all established iPSC lines were karyotyped before further characterization (Figure 4.3B). I found that most of the screened cells from Xlpou91 and mOct4 iPSC lines had a correct karyotype (40 chromosomes) whereas some Xlpou25 derived lines contained a significant proportion of cells with an abnormal karyotype, consisting of an increase in the number of chromosomes (around 42-44). In total 50% of the Xlpou25 derived lines had an abnormal karyotype and they were excluded from further analysis.

Most defined, stable, pluripotent cell lines can now be grown under defined conditions without feeders. To test the robust nature of the reprogramming by different POUV proteins, I next aimed to examine whether POUV derived iPSC lines could be passaged on gelatin coated plates in the absence of feeders. After two passages, the Nanog-GFP expression in each clone was assessed by flow cytometry. Based on testing three clones, I observed that two of the clones induced by either Oct4 or Xlpou91 maintained robust expression of Nanog-GFP in the absence of feeders, while one clone lost expression. All clones derived from Xlpou25 failed to maintain significant levels of Nanog-GFP expression (Figure 4.4). Taken together this suggests that only Xlpou91, not Xlpou25, is able to induce stable expression of the pluripotency network.

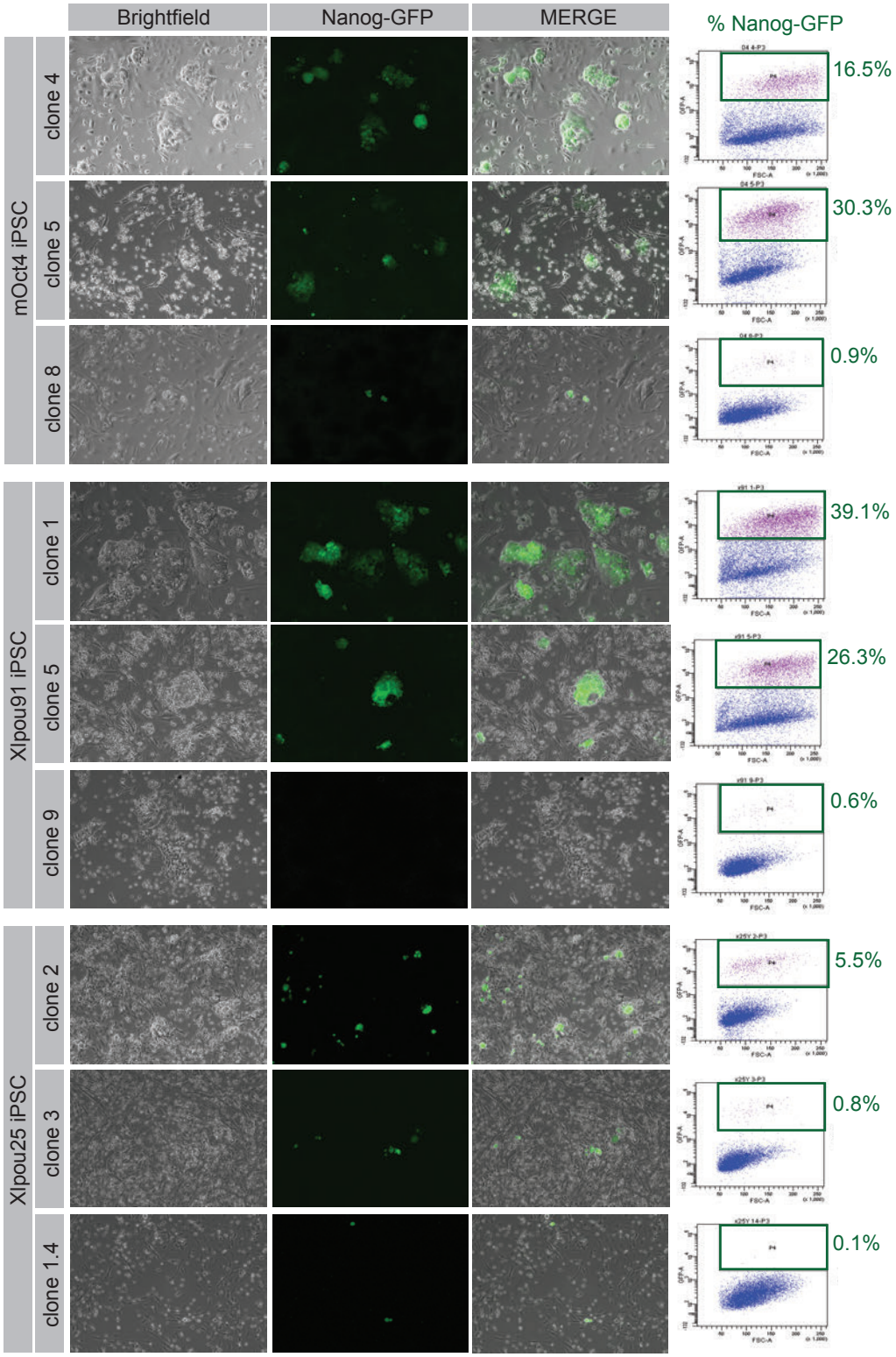


Figure 4.4 Adaptability of iPSC clonal lines to grow on gelatin-coated culture dishes All iPSC clones were picked and cultured on feeders in iPSC medium for 6 passages. The cells were then counted and reseeded onto gelatin-coated culture dishes for two passages. The percentage of Nanog-GFP expressing cells were assessed by flow cytometry.

To determine whether this was a property of the culture or the reprogramming process, I isolated the GFP^+ subpopulation and expanded it with the goal of assessing whether the reprogrammed population induced by Xlpou25 was fundamentally different. To do this I used flow cytometry to sort the Nanog-GFP positive population from Xlpou25 iPSC line alongside a similar population from a mOct4 control. The sorted cells were then expanded on feeders for three passages. Using this strategy, I could establish Xlpou25 iPSC lines capable of growing independent of feeders, suggesting that the initial culture system established by Xlpou25 was unstable, but once isolated it was able to grow efficiently. As shown in figure 4.5, all three clonal lines of Xlpou25 iPSCs could potentially grow on gelatin after removing all Nanog negative cells, and their expression of ESC markers was similar to that of mOct4 iPSC lines, as judged by immunofluorescence and qRT-PCR. During expansion of these gelatin adapted iPSC lines, I also replaced the iPSC medium with ESC medium and found that both mOct4 and Xlpou25 iPSC derived iPSCs exhibited a decrease in Nanog-GFP expression over time, while those in iPSC induction medium maintained the Nanog expression at high level.

Thus, this result suggests that Xlpou91 is capable of restoring the pluripotency gene network similar to mouse Oct4 and that the network probably resembles that of embryonic stem cells, while Xlpou25 also has this capacity, but is less effective at establishing this network stably. Xlpou25 reprogrammed cells appear more sensitive to the differentiated cells around them than do Xlpou91 or Oct4 reprogrammed cells.

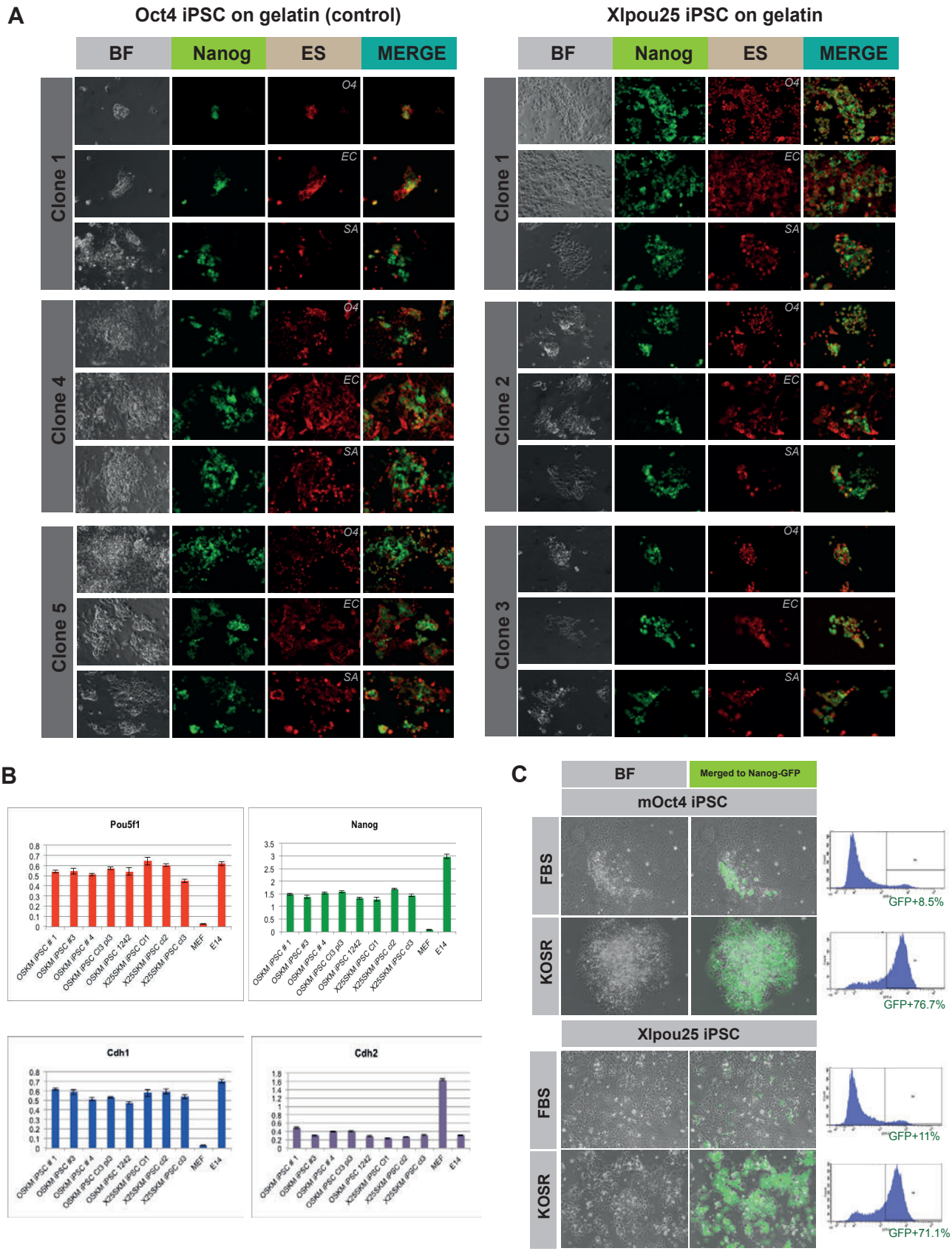


Figure 4.5 The generation of stable Xlpou25 iPSC clonal lines on gelation by serial cell sorting

A) Xlpou25 iPSC growing on feeders showed high differentiation and could not adapt well on gelatin, this might be due to the overgrowth of differentiated cells or partial reprogrammed cells; thus we performed serial cell sorting for three passages to select only Nanog positive population for expansion. After four passages, gelatin-adapted iPSC were stained for immunofluorescence. Both mOct4 and Xlpou25 iPSC showed positive for Oct4 (O4), E-cadherin (EC) and SSEA1 (SA). B) At this passage 4, we also collected cells for qRT-PCR to quantify Oct4, Nanog, Cdh1 (E-cadherin) and Cdh2 (N-cadherin) transcript level . C) All of iPSC clones in A) and B) were maintained in iPSC medium containing Knockout Replacement Serum (KOSR) similar to the iPSC induction. We tested the culture of these iPSC lines in conventional ESC medium with fetal bovine serum (FBS) for two passages and found that Nanog-GFP clearly decreased and showed more heterogeneity.

Section 4.4 Characterization of iPSC clonal lines

Induced pluripotent lines generated by different POUV proteins exhibited ESC-like colony morphology with varied degree of differentiation. To better understand the degree to which they represent true reprogramming, I characterised the ESC-like subpopulation that I have established. I further investigated the expression of naïve pluripotency markers by immunofluorescence, flow cytometry/cell sorting and qRT-PCR.

Immunofluorescence with ESC markers

Immunofluorescent staining of all iPSC lines was performed using antibodies detecting Oct4 and SSEA-1. During reprogramming, iPSCs switch to endogenous Oct4 and are independent from exogenous expression of Oct4 expressed from the retrovirus that becomes silent in the pluripotent state. All clones assessed here were previously screened to confirm that the exogenous Oct4 had been silenced. Stable expression of the endogenous Oct4 is a sign of complete reprogramming. The result of immunofluorescent staining is shown in Figure 4.6 and demonstrates that all iPSC lines expressed endogenous Oct4. I also observed the expression of Oct4 and Nanog were reversely correlated. In particular, Xlpou91 iPSC clone 2 and 3 exhibited low levels of Oct4 but high levels of Nanog expression, while Xlpou91 iPSC clone 1 exhibited high levels of Oct4 but low levels of Nanog expression. This is consistent with published reprogramming experiments on the role of Oct4 dose, that show low levels of Oct4 are required to establish high levels of Nanog expression (Radzisheuskaya et al., 2013). In addition, all iPSC lines were also stained with SSEA-1 antibody, which is an ESC/germ cell surface marker. Interestingly, SSEA-1 was present only in some Xlpou91 and mOct4 induced iPSCs, not those generated with Xlpou25. SSEA-1 expression appeared highest in lines generated with Xlpou91 (figure 4.6).

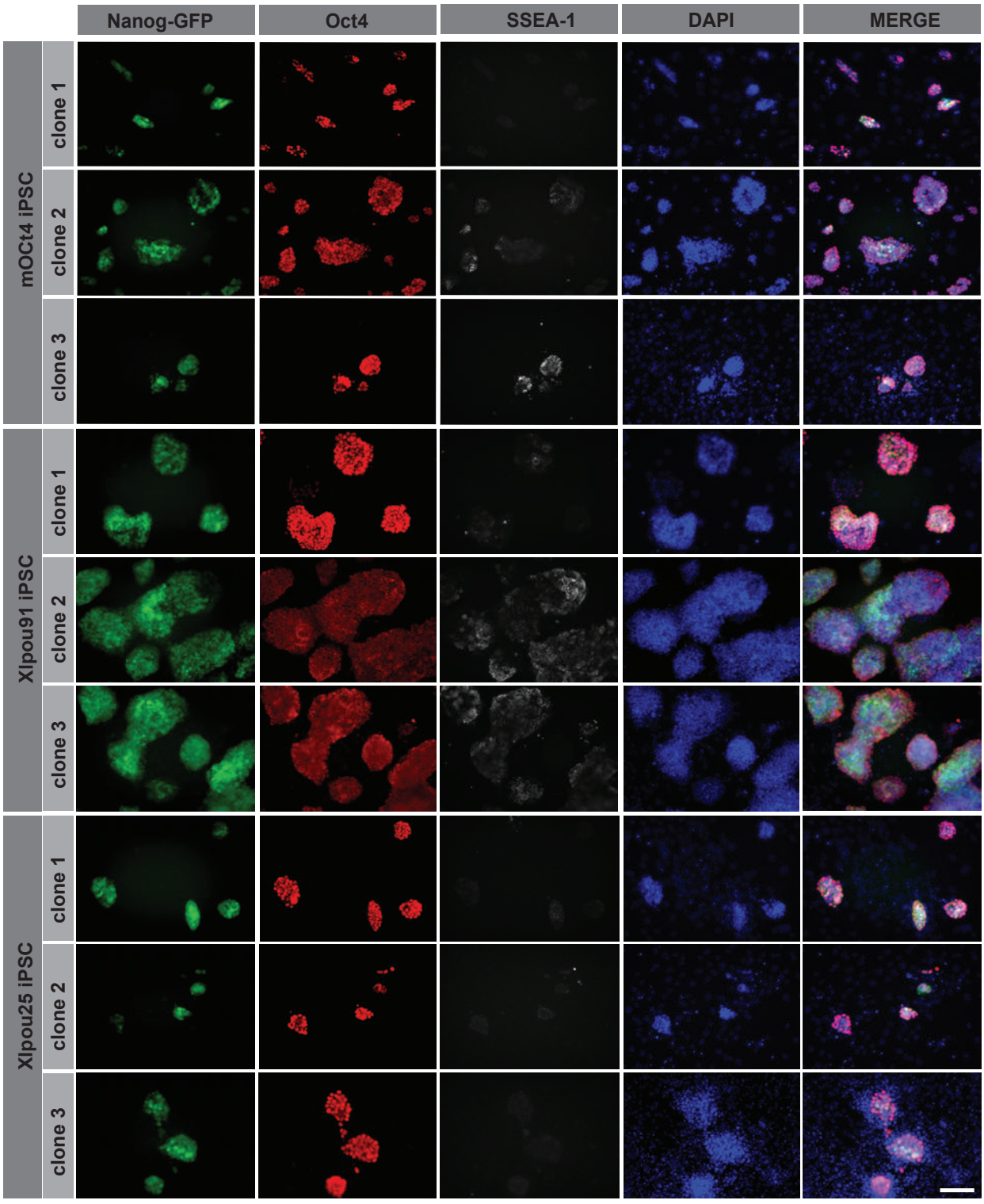


Figure 4.6 Characterization of *Xenopus* POUV-derived iPSC clonal lines by immunofluorescence
iPSC clonal lines at passage 6 were seeded onto feeders and cultured for two days before fixing with 4%PFA. The fixed cells were stained with mouse Oct4 and SSEA-1 antibodies. All imaging and image processes were performed at the same time. Scale bar: 250 μm.

Flow cytometry with iPSC/ESC markers

James O'Malley and colleagues have identified protein surface markers, which can be used to follow murine reprogramming behavior. During reprogramming, they observed a unique pattern of ICAM-1 and CD-44 cell phenotypes as assessed by flow cytometry. In the final state, iPSC exhibiting a fully reprogrammed state (similar to ESC) are ICAM1 positive and CD44 negative in both Nanog positive and negative populations (O'Malley et al., 2013), see chapter 1 figure 1.6. To better characterise the phenotype of iPSCs derived from different POUV factors, I used flow cytometry to characterise the expression of Nanog-GFP, ICAM1 and CD44 in four clonal lines of mOct4 iPSCs, X91 iPSCs and X25 iPSCs. The control was the TNG-B line, an ESC line carrying GFP expression driven by the Nanog promoter (ESCs used to generate the mice from which the Nanog-GFP MEFs were derived (Chambers et al., 2007)).

The results of flow cytometry are shown in figure 4.7 and 4.8. The Nanog positive and negative populations from the control ESCs, TNG-B, exhibited an ICAM-1 positive and CD44 negative profile, confirming the expected cellular phenotype. In newly established iPSC lines, all Nanog positive cells (with the exception of those from Xlpou25 iPSC clone 1.5.1) were $ICAM1^{+}/CD44^{-}$, consistent with this population exhibiting complete reprogramming. However, unlike ESCs, Nanog negative cells of iPSC clones contained two major subpopulations: $ICAM1^{+}/CD44^{-}$ and $ICAM^{+}/CD44^{+}$. The latter subpopulation marks differentiated cells including feeders. As feeders were seeded before culturing iPSCs and were mitotically inactivated, each culture should contain a constant quantity of feeders, and variations in the $ICAM^{+}CD44^{+}$ populations should reflect the overall levels of differentiation in the culture. The cell phenotyping of the different iPSCs is summarised below:

(a) mOct4 iPSC lines: there were three clones with high amounts of $ICAM1^{+}CD44^{-}$ and three clones with high amounts of $ICAM1^{+}CD44^{+}$ cells (Figure 4.7).

(b) Xlpou91 iPSC lines: there were three clones with high amounts of $ICAM1^{+}CD44^{-}$ and two clones with high amounts of $ICAM1^{+}CD44^{+}$ cells (Figure 4.8).

(c) Xlpou25 iPSC lines: there was one clone with high amounts of $ICAM1^{+}CD44^{-}$ and four clones with high amounts of $ICAM^{+}CD44^{+}$ cells (Figure 4.8).

From the observation of phenotypes in the Nanog negative population, we could confirm that Xlpou25 iPSCs had the highest number of differentiated cells, while Xlpou91 iPSCs had the lowest amount of differentiation. However, despite the presence of these differentiated cells, all clones had $ICAM1^{+}CD44^{-}$ populations of fully reprogrammed cells, although the balance between these populations and differentiated cells was different in the different iPSCs.

A second set of cell phenotyping tools available are the surface markers Pecam-1 and c-Kit regularly used for characterization of the naïve pluripotency, germ cell phenotypes and final states of iPSC generation (De Felici et al., 2005; Polo et al., 2012). As a result, I assessed the expression of these markers in different clones derived from either Xlpou25 or Xlpou91. Figure 4.9 shows flow cytometry for these two markers in different iPSC lines. While Xlpou25 derived iPSCs showed expression of both markers, there was a more significant negative population than that observed in either Xlpou91 or mOct4. Interestingly, with the exception of one clone, Xlpou91 induced cells appeared to have both the highest percentage of Pecam-1^{+} and c-Kit^{+} cells and the highest mean level of expression (Figure 4.9).

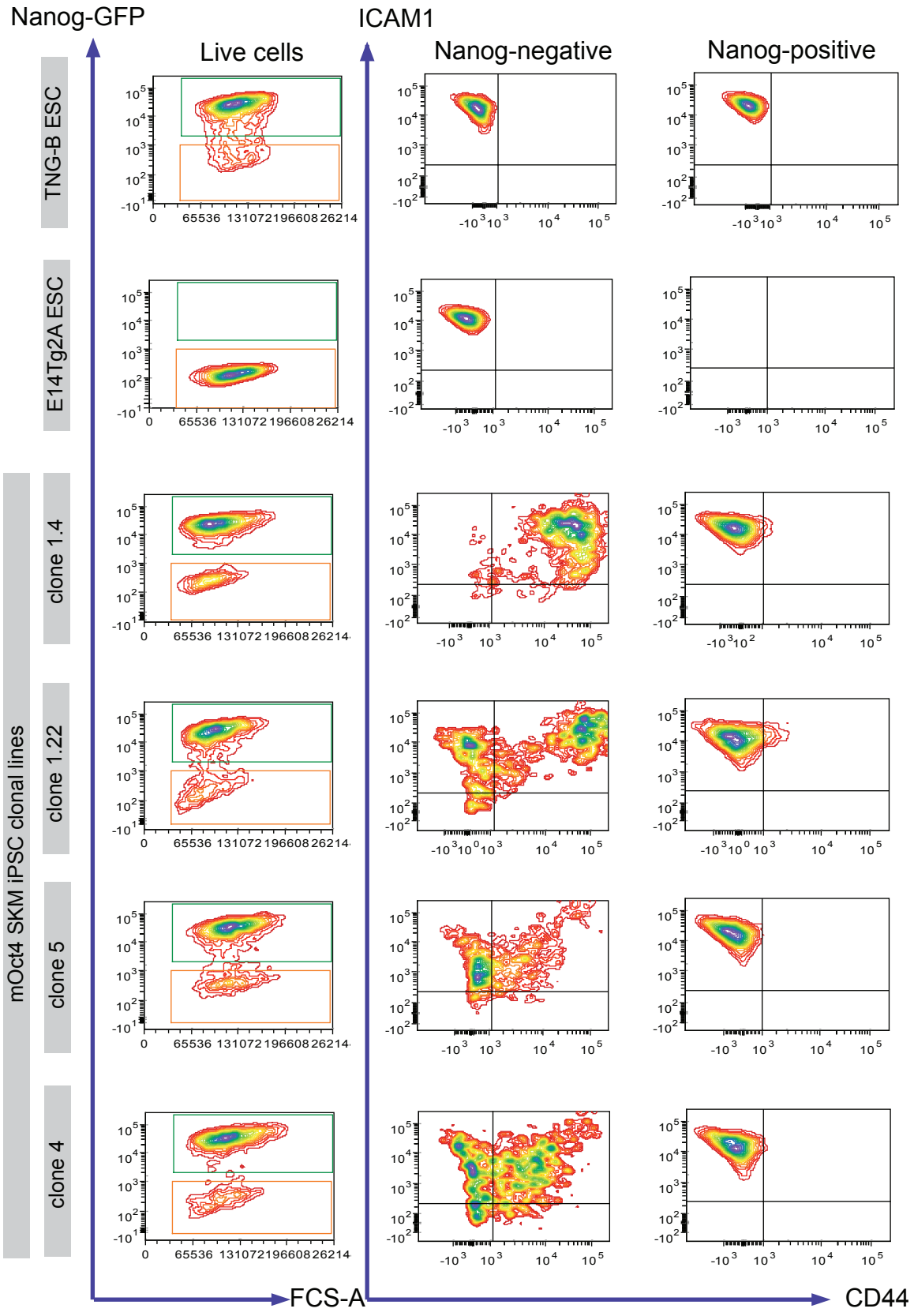


Figure 4.7 Characterization of ESC phenotype in iPSC clonal lines by flow cytometry
All iPSC clonal lines from two independent experiments were expanded to passage 10 and the ESC phenotype assessed by flow cytometry. ESCs and fully reprogrammed cells exhibit ICAM1^{+}CD44^{-} .

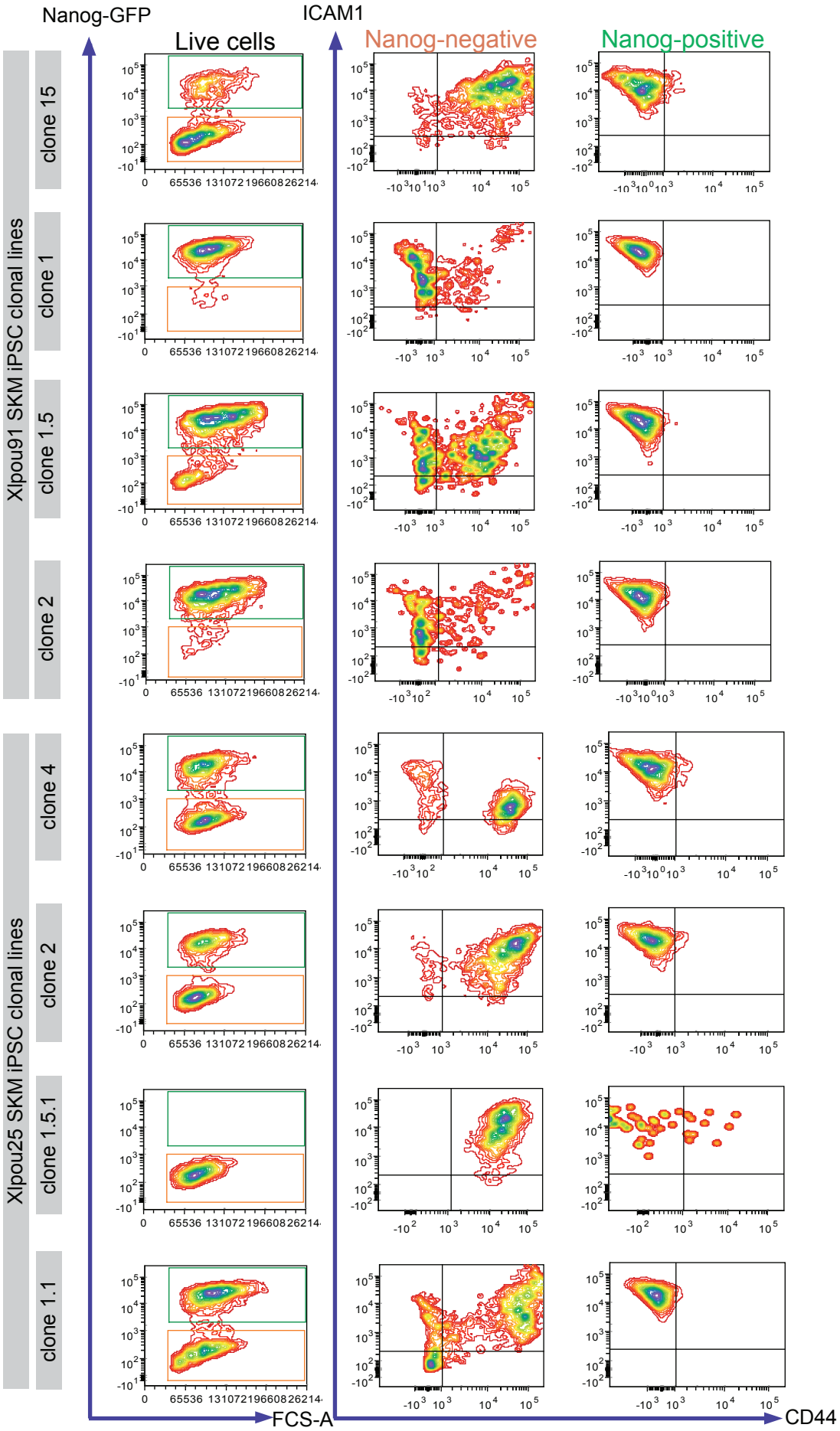


Figure 4.8 Characterization of ESC phenotype in Xlpou91 and Xlpou25-derived iPSC clonal lines by flow cytometry See figure 4.7 for description.

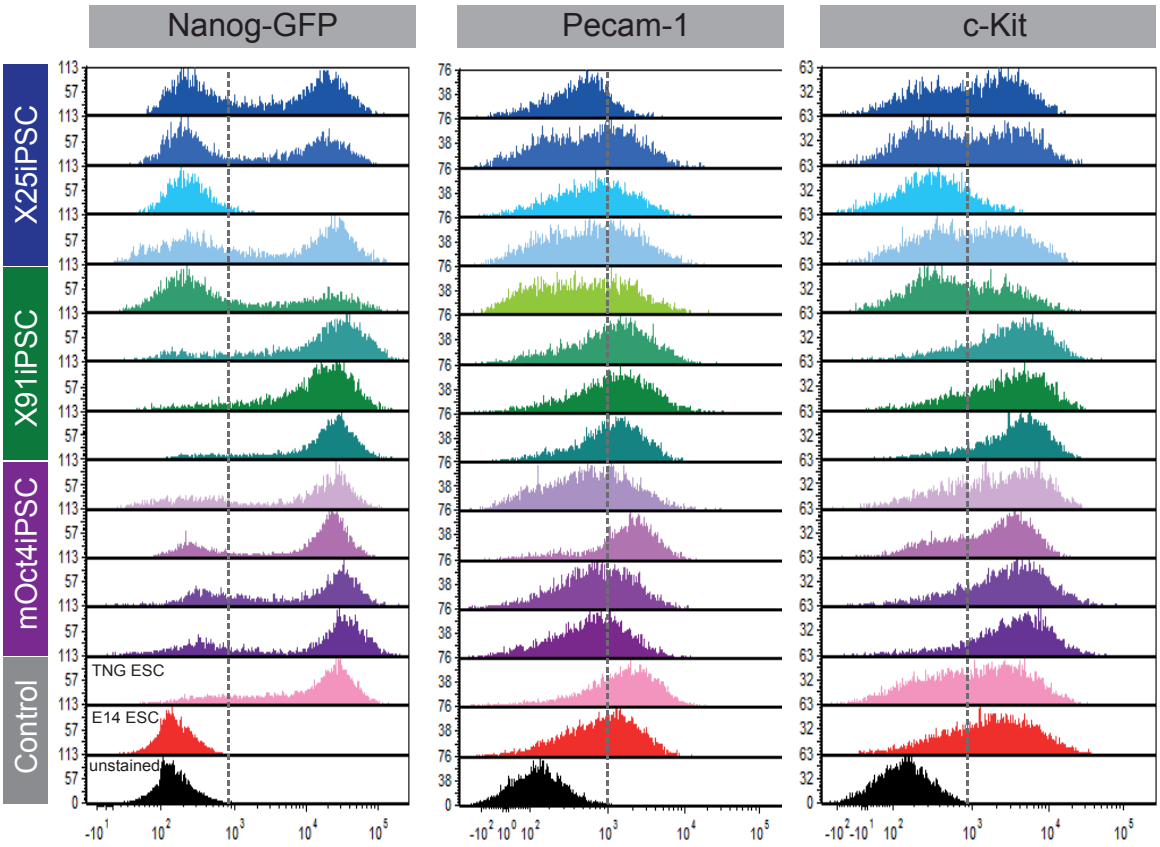


Figure 4.9 Characterization of iPSC clonal lines by flow cytometry with some ESC/germ cell specific markers. Four clones of each iPSC line were analysed by flow cytometry with cell surface marker Pecam-1 or c-Kit alongside with GFP expression driven by Nanog promoter.

To determine whether the different POUV proteins have truly produced a completely reprogrammed stable population of cells, I decided to perform molecular analysis on the population of cells that expresses both Nanog-GFP and c-Kit as this is thought to better represent homogeneously reprogrammed cell populations, since that both markers are expressed in naïve ESCs. Figure 4.10 represents this data in terms of double and single positive populations. Consistent with all the previous analysis, I observed similar levels of this double positive population in Oct4 and Xlpou91 induced iPSCs, and this was reduced in the majority of Xlpou25 induced cell lines.

Gene expression analysis of ESC-like population in the iPSC lines by qRT-PCR

To assess whether these populations were the same or different, in cell lines derived from different POUV proteins, the Nanog^{+} c-Kit^{+}population was sorted and RNA made from these samples for qRT-PCR. Figure 4.11 shows the results of this qRT-PCR for the expression of Prdm14, Esrrb and Klf4. While the overall levels of these genes in the culture (total population of live cells purified by flow cytometry) were different, the expression of these genes was similar in the sorted populations derived from the different iPSCs. This observation suggests that in all cases full reprogramming is obtained in a portion of the culture, but the extent to which this population is maintained dynamically over time is influenced by the starting POUV protein.

Why should the reprogrammed population in one culture be more stable than another? Why should Xlpou25 induce the same state, but a state that appears more prone to differentiate? To address these questions, I sorted the Nanog^{+}c-Kit^{+} positive population from different iPSCs and determined the transcriptome established during reprogramming within this pluripotent sub-fraction. These results of transcriptome analysis are shown in section 4.5.

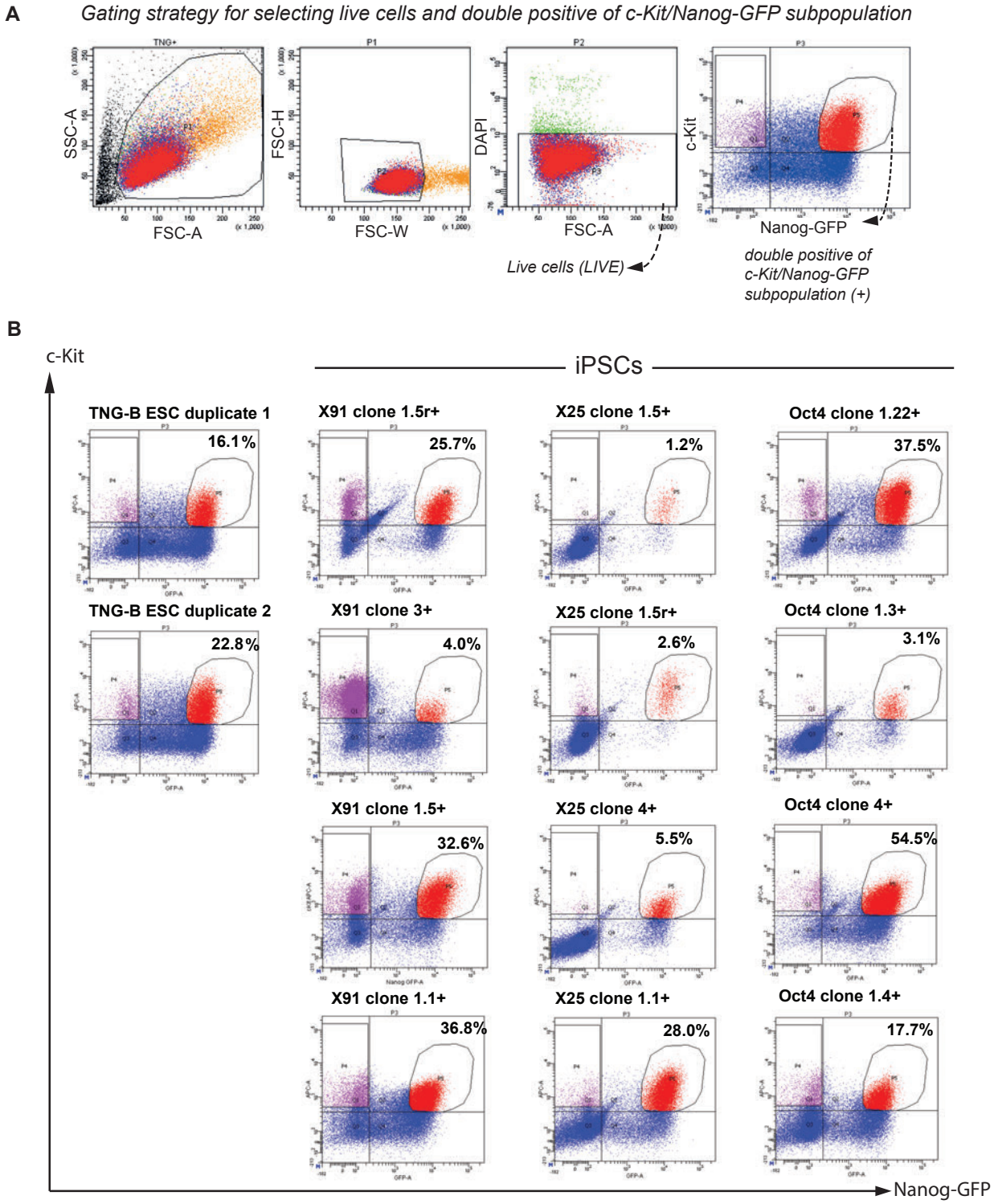


Figure 4.10 FACS sorting strategy and analysis of selected iPSC clonal lines for microarray analysis There is some degree of correlation of Nanog and c-Kit expression, as shown in figure 4.9. Thus, I used c-Kit and Nanog as markers to sort naïve ESC-like population in the iPSC clonal lines for further characterizing global gene expression profile (figure 4.12). The gating strategy to collect double positive (abbreviation, "+") of c-Kit and Nanog subpopulation and live cells (abbreviation, "LIVE") are shown in A. B) four clones of each iPSC lines and two duplicates of TNG-B were FACS-sorted as strategy in A. Gene expression of "+" and "LIVE" subpopulations were preliminarily analyzed by qRT-PCR (figure 4.11)

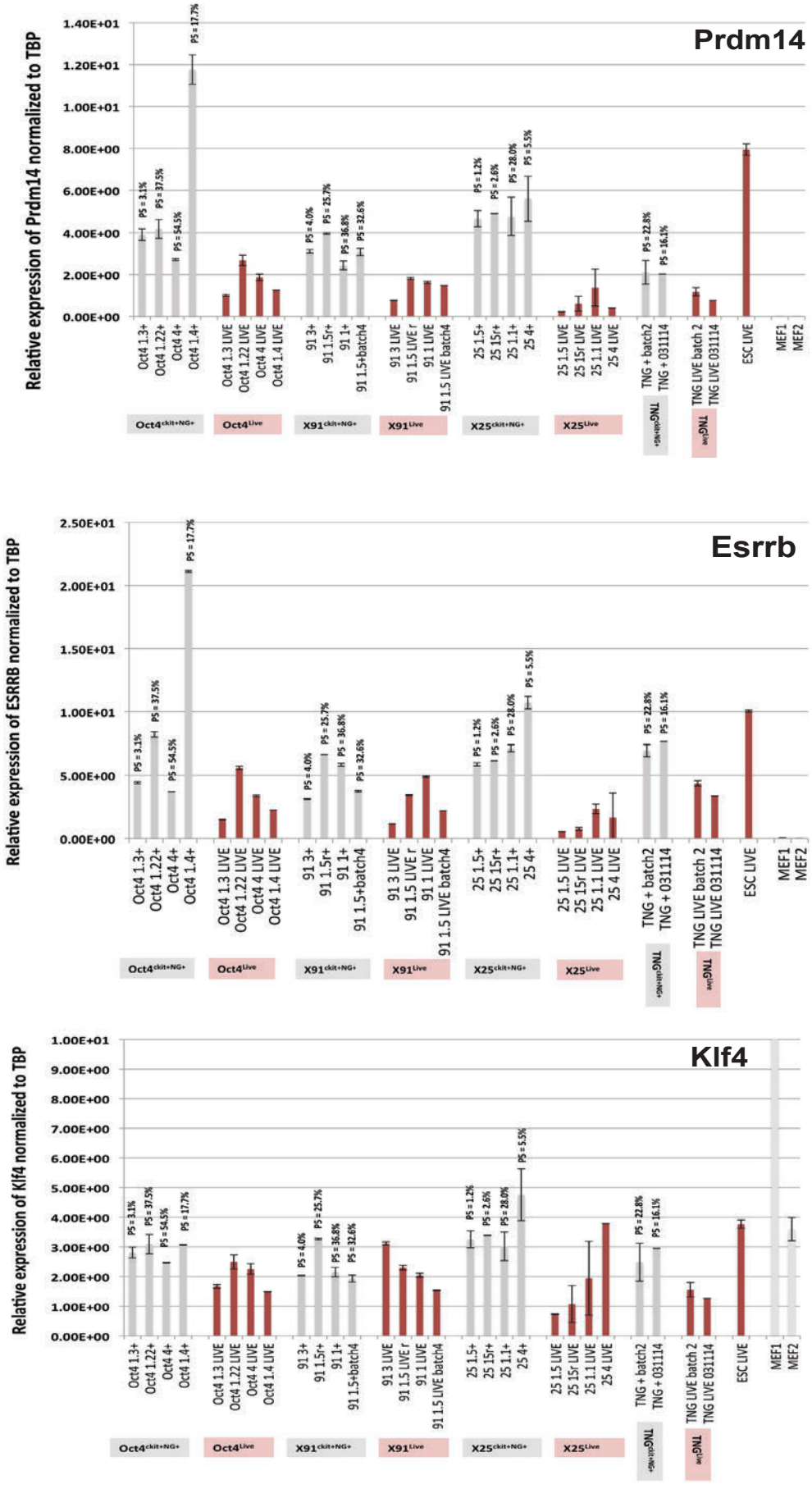


Figure 4.11 Characterizarion of c-Kit⁺/Nanog-GFP⁺ subpopulation of iPSC clonal line by qPCR. IPSC cells at passage 12 were FACS-sorted, see figure 4.10. Live cells (LIVE) and c-Kit/Nanog double positive cells (+) were collected for RNA extraction and analyzed by qRT-PCR. The transcript levels of genes of interest were normalized to house keeping gene TBP. The numbers on the bar chart show the percentage of sorted c-Kit/Nanog-GFP subpopulation, according to Figure 4.10. MEF1 and MEF2 are Nanog-GFP MEF cell lines used to establish iPSCs. Data represents technical replicates.

Section 4.5 Microarray-based global gene expression profile of iPSC clonal line generated by different Oct4 homologues

To determine global gene expression differences between iPSC lines, we carried out one-colour based microarray on two biological replicates of each iPSC line. The biological replicates were selected based on the following criteria: (1) similar ESC phenotypes (ICAM-1/CD44 profile), (2) correct karyotype, and (3) similar ESC/germ cell-related gene expression profile of sorted double Nanog/c-Kit double positive population. Full details of the microarray strategy is described in chapter 2. Briefly, we used high quality RNA (RNA integration number (RIN) = 10) to generate Cyt3-labelled cRNAs, which were then hybridised to whole genome-oligonucleotide microarray slides (Agilent Technologies). The hybridised slides were scanned to obtain global gene expression profiles. The raw datasets of samples were combined and uploaded to the online analysis platform, NIA Array Analysis (http://lgsun.grc.nia.nih.gov/ANOVA/). All inputs were normalised by the platform to equalise multiple quantiles of the probability distribution of gene expression. Global gene analysis was based on 1) false discovery rate threshold (FDR) $\leq$ 0.05 2) fold-change threshold $\geq$ 2.

First, I investigated how global gene expression profiles are different between each iPSC line by performing pair-wise comparison (Figure 4.12). The pairwise comparison of gene expression profiles between each pair of the iPSC lines allowed the identification of genes that are differentially expressed at a significance level FDR < 0.05, in both overexpressed and underexpressed directions. Results of the pairwise comparisons of sorted iPSC lines indicate very little difference between the purified fractions from either line. Intriguingly, the largest difference was observed between mOct4 and Xlpou91 iPSCs (127 overexpressed and 230 underexpressed). The smallest differences in gene expression were

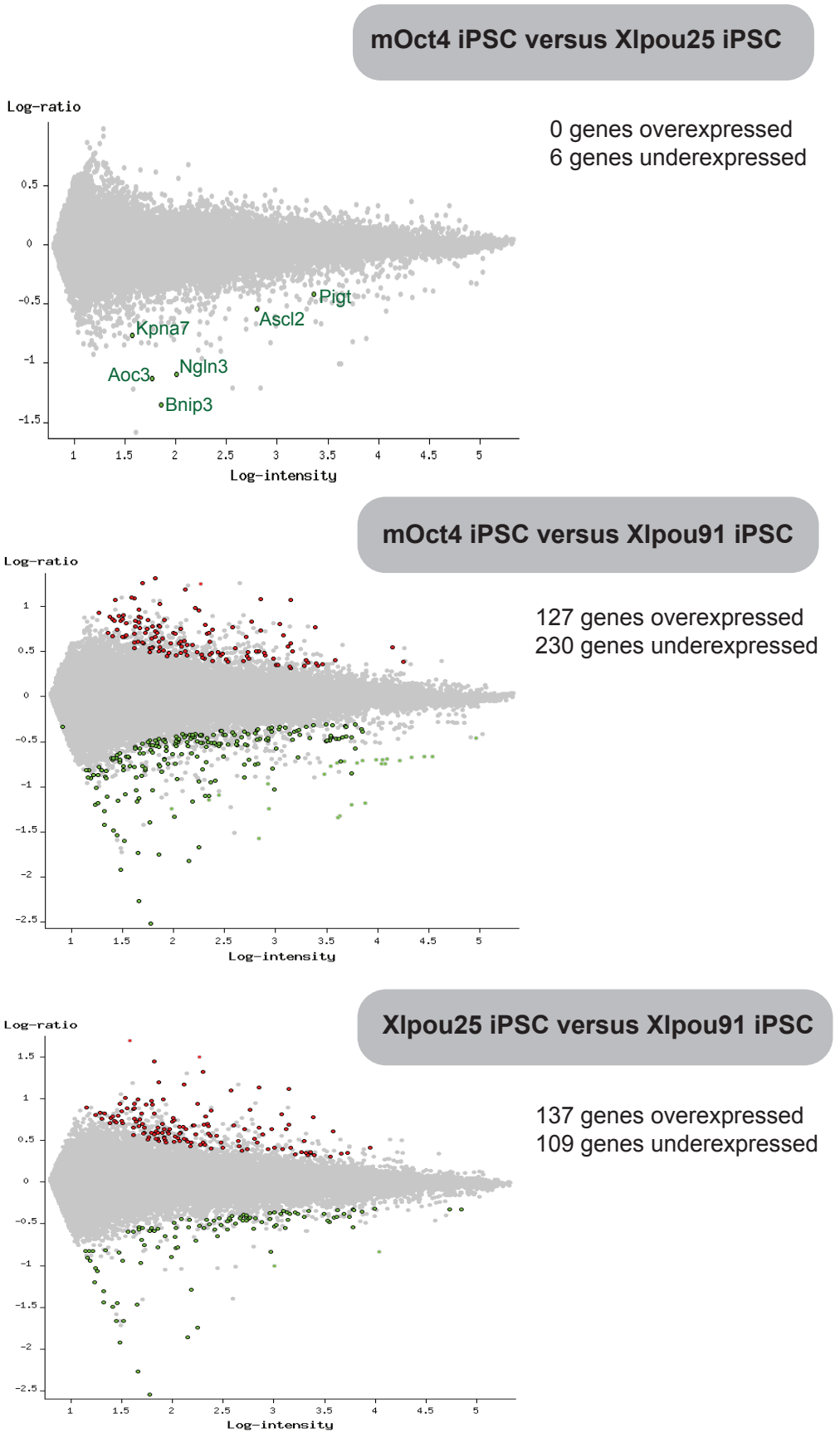


Figure 4.12 DNA microarray-based global gene expression analysis of c-kit^{+}Nanog-GFP^{+} subpopulation of iPSC clonal lines generated by different *Xenopus* POUV proteins
Pairwise comparison of gene expression profiles were assessed by NIA array online tool. The data represents two biological replicates

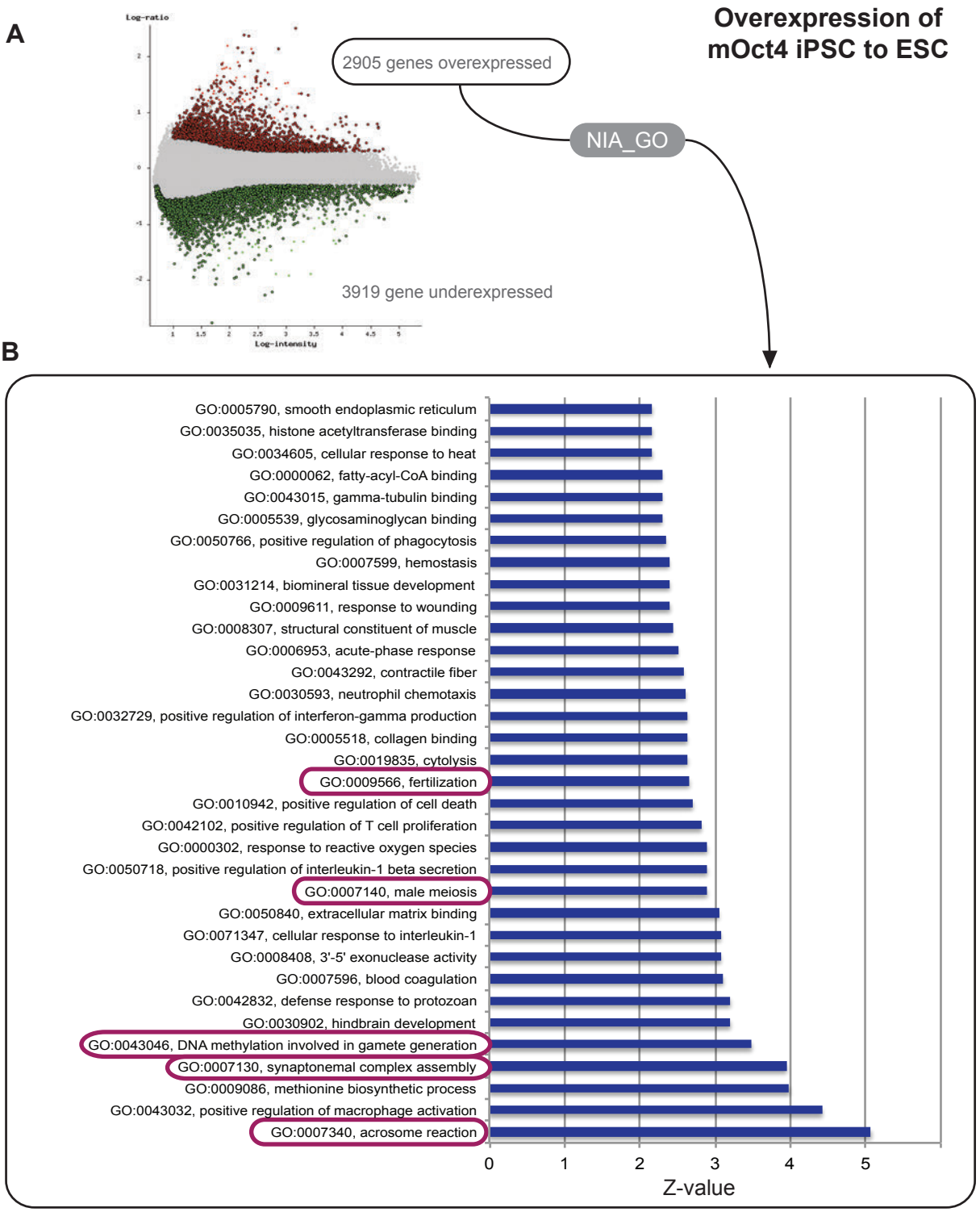
observed between Xlpou25 iPSCs and mOct4, with whole genome expression that exhibited very few greater than 2-fold differences (only 6 genes were differentially expressed). Taken together, these comparisons suggest all these cultures contain very similar populations of reprogrammed cells.

In order to perform more sensitive tests for understanding differential gene expression, I thought to compare all three iPSCs expression profiles to ESCs. As gene expression differences between different iPSCs might be masked by the 2-fold cut off, I thought that perhaps the differences between these iPSCs and ESCs might produce different lists of genes that could point to subtle differences between these reprogrammed cells. As a result I compared global gene expression profiles of iPSC lines to E14Tg2a ESC by pairwise comparison (Figure 4.13-4.15). Genes differentially expressed in iPSC lines in comparison to the ESCs were extracted and subjected to GO enrichment analysis in ExATLAS (NIA webtool). As these comparisons produced significant differences (even for Oct4 the number of gene expression changes was around 5,000, Figure 4.13A), I could ask whether there were differences in specific GO terms. The GO clusters enriched in both Xlpou91 and mOct4 induced $Nanog^{+}c\text{-}kit^{+}$ populations showed significant enrichment in several germ cell categories relative to ESCs, whereas Xlpou25 did not. These catagories are highlighted in red in figures 4.13-4.15. The term reminiscent of germ cells identified in Xlpou25 induced cells was meiosis.

In addition, the expression of some pluripotency and differentiation genes was compared between sorted iPSC lines, ESC and EpiSC (Figure 4.16). Several naïve pluripotency genes (e.g. *Sox2*, *Klf2*, *Klf4*, *Klf5*, *Esrrb*, *Pecam-1*, *Icam-1*, *Zfp42*) that are downregulated in EpiSC were not significantly different between sorted iPSC lines and the ESCs. Among differentiation markers, the sorted iPSCs expressed lower levels of epiblast genes (e.g *Fgf5*, *Brachyury*, *Wnt3*) and trophoblast genes (e.g. *Cdx2*, *Krt18*) than the ESCs.

This result reveals that the c-Kit and Nanog double positive population resembles naïve ESCs and my choice to use these two markers to compare iPSC lines was a valid one.

Taken together, my results so far have pointed to subtle differences in the levels of germ cell markers expressed within the reprogrammed cells induced by either Oct4 or Xlpou91. This is revealed both through the Pecam staining and pairwise comparisons. Are these changes sufficient to explain the differences in stability of these populations in terms of differentiation? Or is there something different about the route by which these cells are established that leads to a less stable pluripotent state that is more prone to dynamically differentiate? To address these questions, I compared the set of genes induced during the reprogramming process by these different POUV proteins.



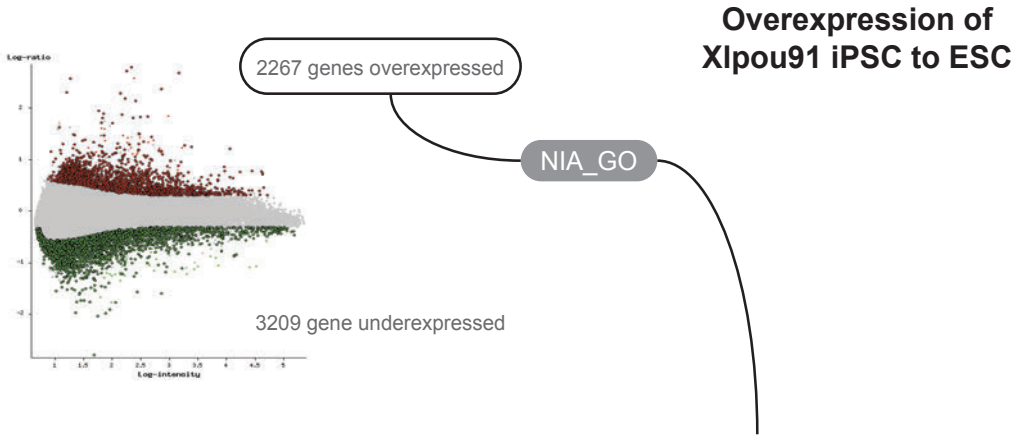
C

GO:0007340, acrosome reaction	GO:0007140, male meiosis	GO:0043046, DNA methylation involved in gamete generation
Acr,Glra1,Plcd4,Spesp1,Syt6, Syt8,Trpc2	Asz1,Mael,Tdrd9,Tex11,Tex15	Asz1,Ctcfl,Mael,Piwil2,Tdrd9

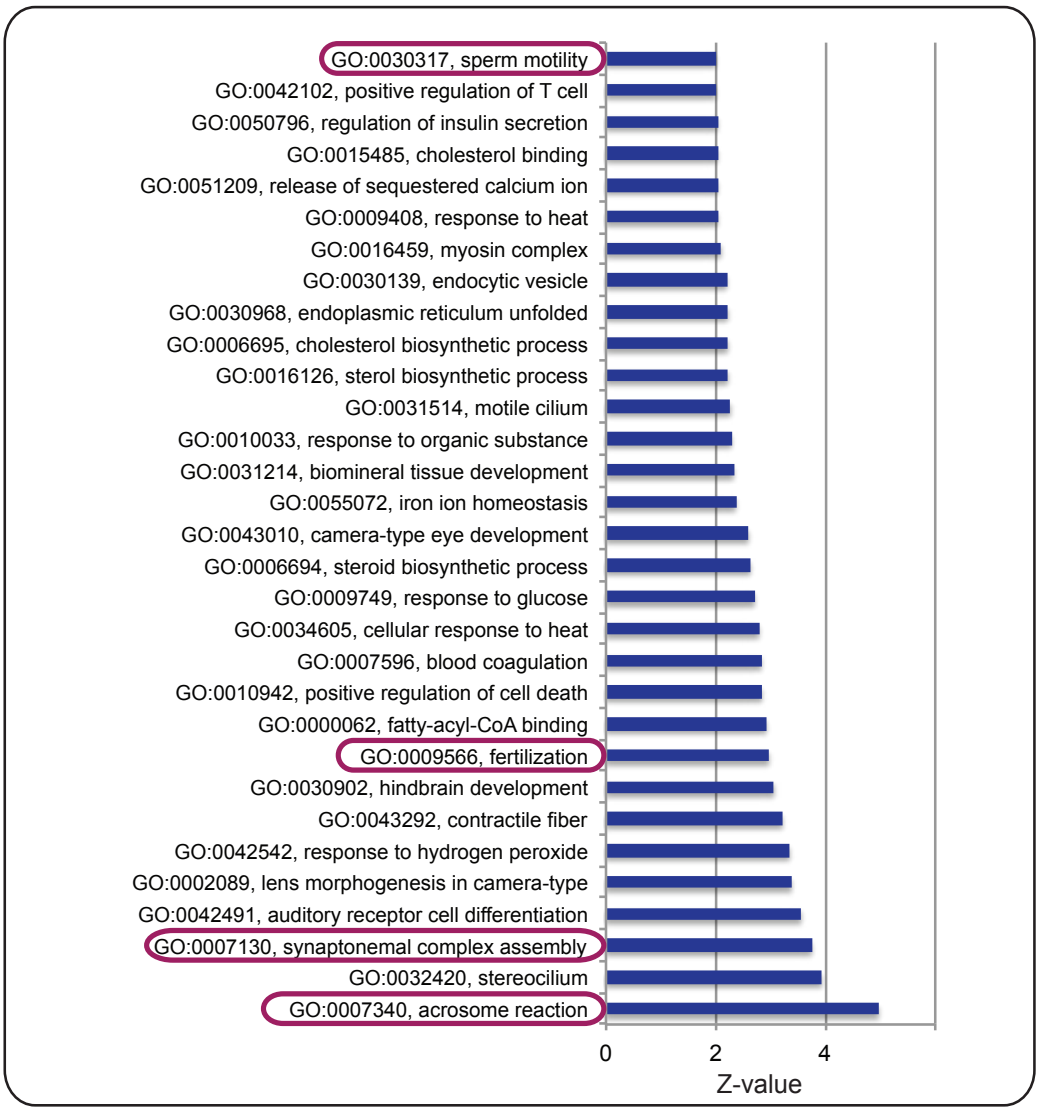
GO:0007130, synaptonemal complex assembly	GO:0009566, fertilization
Hormad1,Syce1,Sycp1,Tex11, Tex12,Tex15	Atp1a4,Klhl10,Klk14,Mael,Prdm14, Spef2,Stra8,Tdrd9,Tex11,Tex15

Figure 4.13 Gene annotation enrichment analysis of mouse Oct4 iPSC overexpressed genes compared to ESCs. A) Gene expression of double positive c-Kit/ Nanog subpopulation of mOct4 iPSCs were used to compare to E14 ESCs in NIA Array Tool Analysis. There are 2905 genes overexpressed in mOct4 iPSCs to the ESCs. B) The list of 2905 genes were further analyzed for the GO enrichment in ExATLAS, NIA webtool. The red block indicate the genes related to germ cell development and reproduction. C) List of genes from GO clusters related to germ cell development are shown.

A



B



C

GO:0007340, acrosome reaction	GO:0030317, sperm motility
Acr,Plcd4,Spesp1,Syt6,Syt8,Trpc2	Catsper4,Dnaja1,Hist1h1t,Ldhc, Lrrc6,Rabl2,Rhox5
GO:0007130, synaptonemal complex assembly	**GO:0009566, fertilization**
Hormad1,Syce1,Tex11,Tex12, Tex15	Klhl10,Klk14,Mael,Spef2,Stra8,Svs2, Tex11,Tex15,Tsx

Figure 4.14 Gene annotation enrichment analysis of Xlpou91 iPSC overexpressed genes compared to ESCs. A) Gene expression of double positive c-Kit/ Nanog subpopulation of mOct4 iPSCs were used to compare to E14 ESCs in NIA Array Tool Analysis. There are 2267 genes overexpressed in mOct4 iPSCs to the ESCs. B) The list of 2267 genes were further analyzed for the GO enrichment in ExATLAS, NIA webtool. The red block indicate the genes related to germ cell development and reproduction. C) List of genes from GO clusters related to germ cell development are shown.

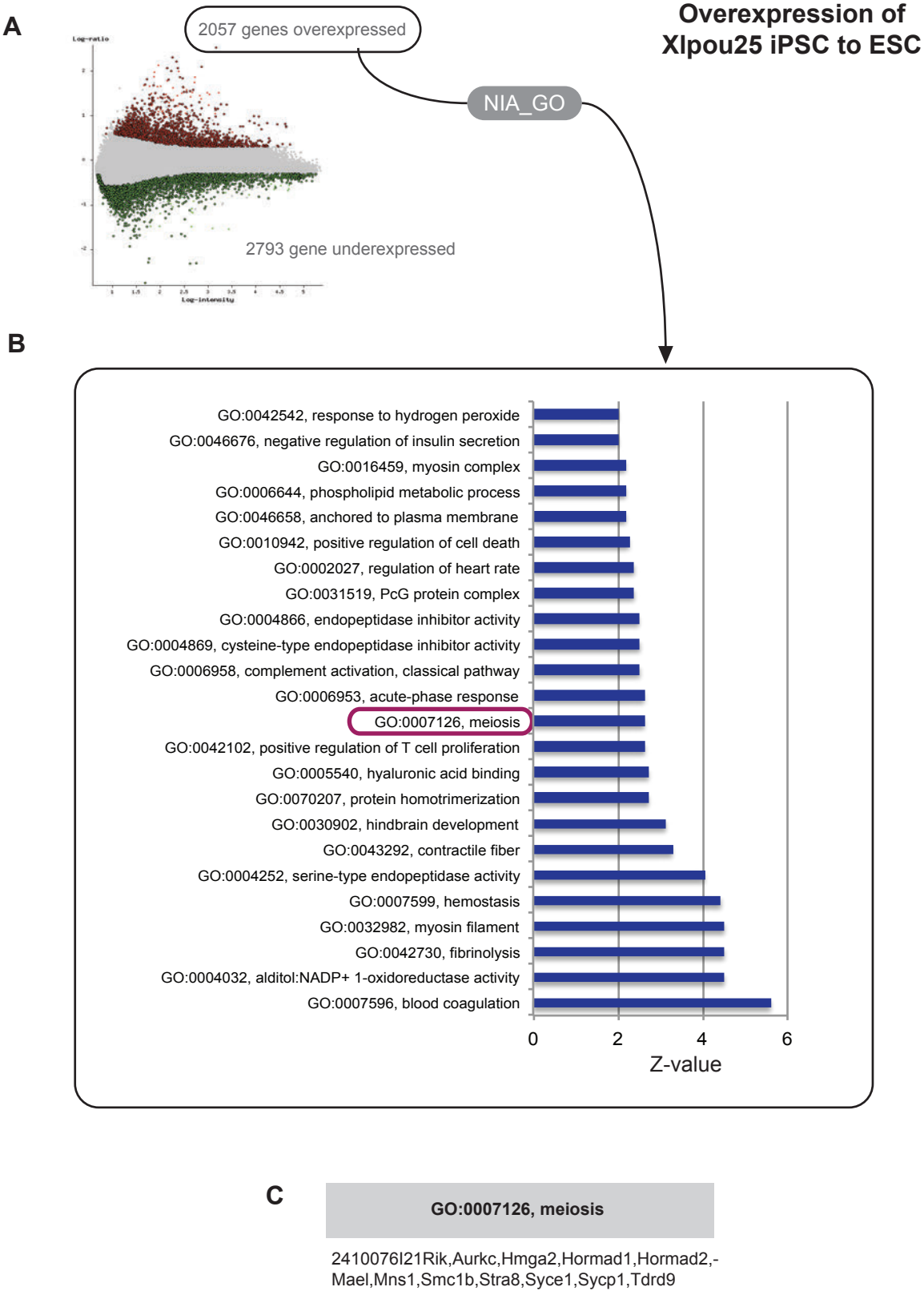


Figure 4.15 Gene annotation enrichment analysis of Xlpou25 iPSC overexpressed genes compared to ESCs. A) Gene expression of double positive c-Kit/ Nanog subpopulation of mOct4 iPSC were used to compare to E14 ESCs in NIA Array Tool Analysis. There are 2057 genes overexpressed in mOct4 iPSCs to the ESCs. B) The list of 2057 genes were further analyzed for the GO enrichment in ExATLAS, NIA webtool. The red block indicate the genes related to germ cell development and reproduction. C) List of genes from GO clusters related to germ cell development are shown.

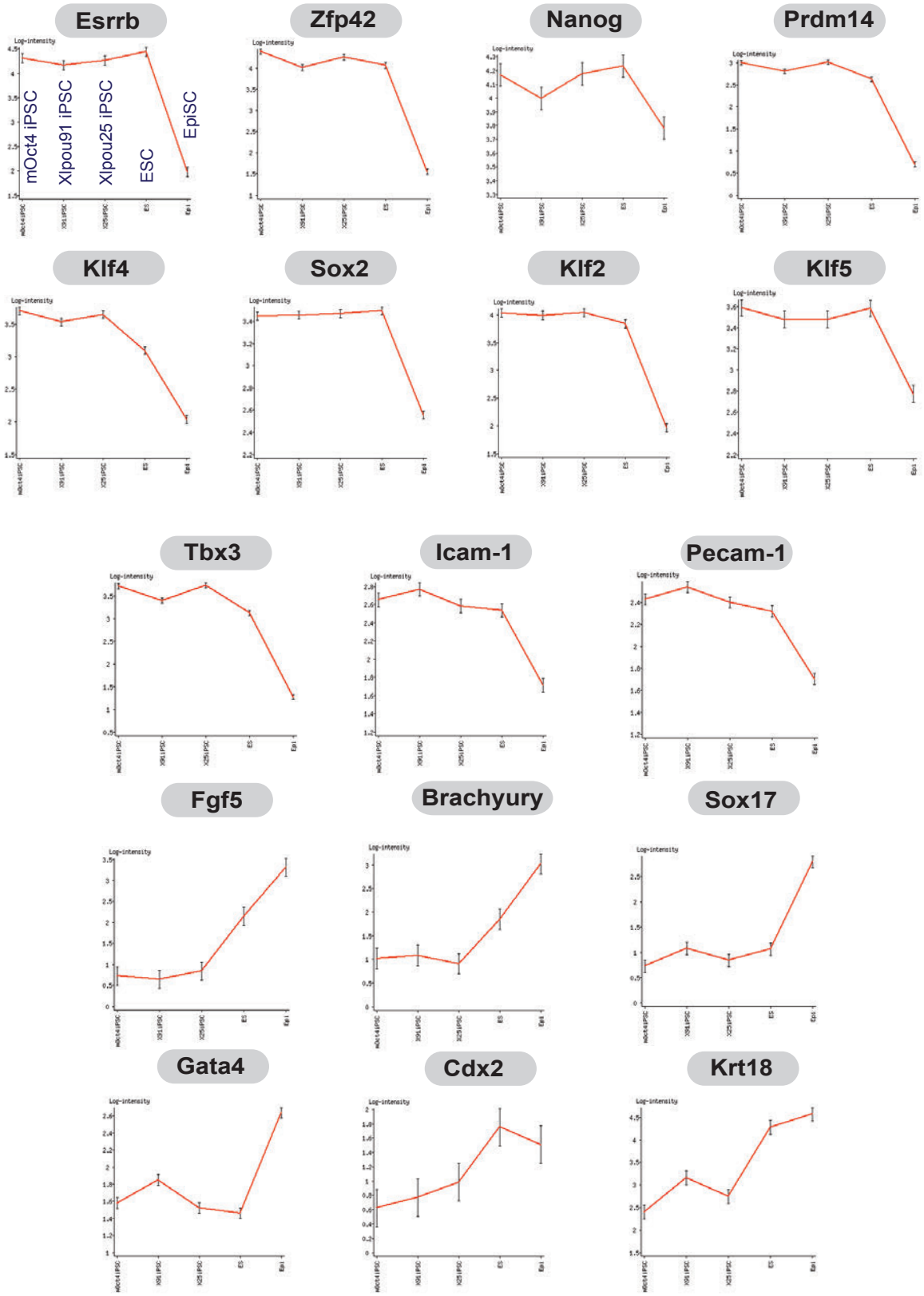


Figure 4.16 Microarray-based expression profiles of some selected genes involving in germ cell specification, iPSC generation and ESC/EpiSC pluripotency. ESC-like subpopulation of sorted iPSCs (c-Kit^{+}Nanog^{+}) were compared to unsorted ESCs and EpiSCs. Graphs show mean-log intensity representing transcript levels. Data represents two biological replicates.

Section 4.6 Mechanisms of cellular reprogramming driven by different Oct4/POUV homologs

As shown in previous sections, Xlpou91 and mouse Oct4 have the potential to overcome the barrier to reprogramming with similar kinetics and based on response to a similar dose of POUV protein. Xlpou25 is required at a higher level, had slower reprogramming kinetics and produces a less stable pluripotent population. To assess the basis for these differences I performed a time course on reprogramming by these factors using qRT-PCR.

The effects of different doses of POUV homologs on early embryonic gene expression were analyzed at day 4 to day 9 of reprogramming. This time period was chosen because I would like to monitor the effects of exogenous POUV expression to the reprogramming process as it is expressed early and just prior to the emergence of endogenous Oct4 expression around days 10-13 (based on the induction by mouse Oct4).

In the first experiment, I assessed the immediate early effects of all three POUV proteins. I collected RNA for qRT-PCR from day 4 post the first infection, which was the day of reseeding induced MEF cells onto the irradiated feeders. I analyzed the expression of genes that have been associated with reprogramming in other studies. In particular, embryonic ectodermal genes, mesodermal genes and EMT/MET related genes were examined. I observed no expression of pluripotency genes (e.g. Prdm14 and Nanog) at day 4 post infection. In figure 4.17A, I found that different transcriptional levels of POUV homologs affected the expression of these markers differently. Interestingly, the high dose of Xlpou25 (which can generate iPSCs) induced high level of Lhx5, an ectodermal gene marker, regulator of adhesion and direct Oct4 target that is conserved in *Xenopus* gastulation

(Livigni et al., 2013). High doses of Xlpou25 also induced the highest level of N-cadherin (Cdh2), a mesenchymal/mesodermal and primed pluripotency marker, and the primitive endoderm marker Sox7. The ectodermal marker Dix3 was also induced by Xlpou25, but E-cadherin appeared to respond best to low doses of all three POUV proteins.

Based on these intriguing observations indicating gastrulation stage gene expression downstream of Xlpou25, I investigated its activity in reprogramming more closely, using Oct4 as a positive control. Here I explored the expression of early embryonic genes at day 4 and 9. I found that expression of a second primitive endoderm marker Gata6 was preferentially induced by Xlpou25, as well as the primitive endoderm and epiblast marker Fgf5 (figure 4.17B). Strikingly, Xlpou25 indeed induced higher level of Gata6 than mOct4 and the transcript level of Gata6 increased during reprogramming. In particular, high dose Xlpou25 induces the higher level of Gata6 than the low dose. It has been shown that primitive endoderm markers like Gata6, Gata3 and Sox7 are potent in replacing mouse Oct4 in murine cellular reprogramming (Shu et al., 2013). Thus it is possible that the underlying gene regulatory network responding to reprogramming could be different in Xlpou25 induced cells, and that these cells progress into reprogramming with the help of primitive endoderm transcription factors. If these factors remain expressed throughout the reprogramming process, this would explain the potential instability of the pluriptotent populations induced by Xlpou25. These experiments are only preliminary and require repeating before extensive interpretation.

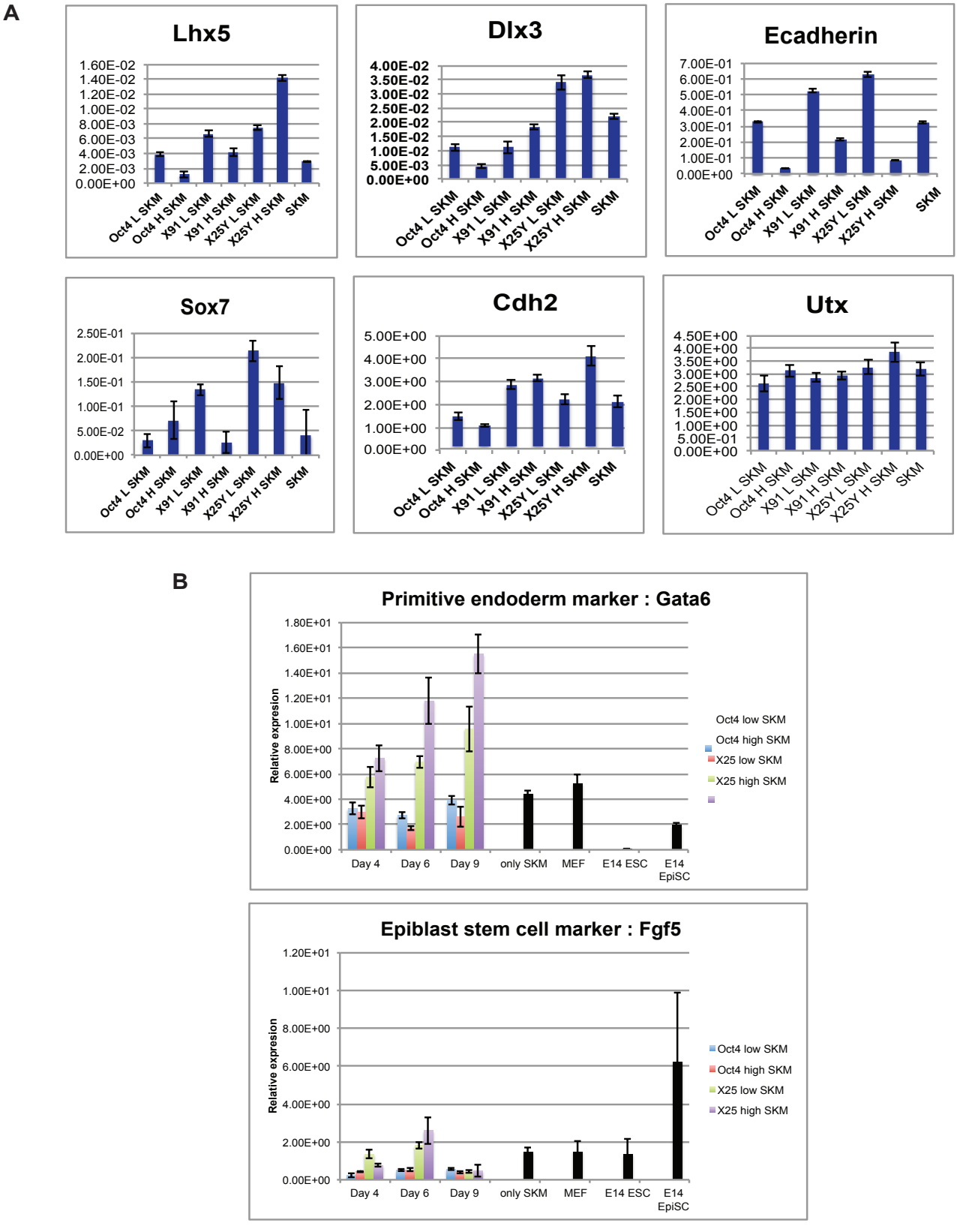


Figure 4.17 Mechanism of cellular reprogramming induced by different POUV homologs A) gene expression analysis by qRT-PCR of POUV-SKM induced MEF cells at day 4 post transduction. L, H and SKM stands for low Oct4/POUV dose, high Oct4/POUV dose, and Sox2-Klf4-c-Myc, respectively. B) time-course gene expression profile of Gata6 and Fgf5 from Oct4 and Xlpou25 SKM-induced MEFs collected at day 4, 6 and 9 post transduction.

Section 4.7 Discussions

In this chapter, I have shown that the roles of POUV proteins in the induction of pluriptoency are conserved. I showed that two different Pou5f3 proteins, with different ESC rescue activity, could both induce iPSCs. While these proteins induced very similar iPSC populations, the stability of these populations to differentiation did correlate with the ability of these proteins to support ESCs. I also showed that this stability might be explained by differences in the sequence of factors induced by POUV proteins during reprogramming.

Based on the rescue assay shown in Morrison and Brickman, 2006 and this study, *Xenopus* Pou91 is a germ cell specific Pou5f3 protein with Pou5f1-like activities. Low levels of Pou91 were capable of reprogramming murine somatic cells towards a naïve pluripotent state. This can be compared to the other *Xenopus* homolog with more obvious Pou5f3 like activities in gastrulation, Xlpou25. Higher levels of Xlpou25 expression were required for reprogramming of somatic cells towards pluripotent states. Xlpou25 represents a primed state/epiblast specific POUV protein, regulating gastrulation in *Xenopus* and required for cell integrity during the epiboly process (Livigni et al., 2013). The discovery that high doses of Pou25 could achieve reprogramming implies that both naïve/germ cell and primed/epiblast-like activity of vertebrate POUV are able to induce pluripotency. As shown in chapter 3, Xlpou25 rescue phenotypes are similar to those of all POU5F3-rescued lines. Are all POU5F3 required at the high dose to achieve reprogramming? I have shown that coelacanth POU5F3 and turtle POU5F3 are required at higher levels to maintain partial ESC self-renewal. Perhaps these cells have lower affinity for naïve ESC targets. Future work in reprogramming of mouse somatic cells with these POU5F3s and a comparison to their binding affinities on a few key target genes might explain some of these differences.

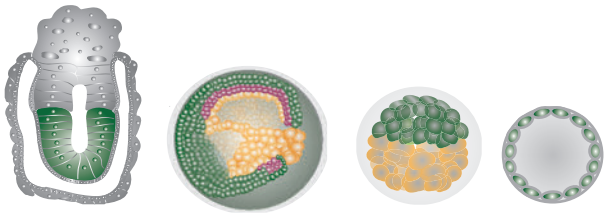
Tapia et al 2012 has shown that conserved POUV activities in reprogramming varied amongst different POU5F1 and POU5F3 encoded factors. However, they performed iPSC generation at a constant dose of different POUV proteins and found that the fish POUVs (POU5F3s) were very poor in reprogramming. However, based on my data, it is likely that these proteins could all reprogram at a sufficient dose.

The final states of reprogramming induced by different POUV homologs was very similar, in that they all contain a population of $ICAM1^{+}CD44^{-}$ (O'Malley et al., 2013), Nanog positive cells. Analysis of the sorted $Nanog^{+}c\text{-}Kit^{+}$ subpopulation suggests that these populations are very similar in the different iPSC lines, however cell lines derived from Xlpou25 appeared more prone to differentiate and contained higher levels of differentiated cells in their cultures. Is this related to the stability of the pluripotency network in these iPSCs ? Transcriptomics indicates only subtle differences, but perhaps the Xlpou25 route to induce reprogramming is more prone to drive differentiation. Expression of Xlpou25 is required at high levels early in reprogramming and these levels are sufficient to induce expression of endoderm genes such as Sox7 and Gata6, and endoderm/epiblast genes Fgf5 and N-cadherin. Perhaps the robust induction of these genes can induce the pluripotency network but at the same time promote endoderm differentiation, so that sub-fractions of the cultures produced by Xlpou25 contain a mixture of pluripotent cells and differentiatied cells.

The one subtle difference between the fully reprogrammed population of $Nanog^{+}c\text{-}Kit^{+}$ is the increase in GO terms associated with germ cells in the Xlpou91- and Oct4-induced iPSCs when compared to ESCs. While mOct4 has both germ cell and epiblast like function based on its expression profile, the duplication of *pou5f3* in *Xenopus* created an opportunity for specialization and Xlpou91 exhibits early germ cell expression, but is not observed in Xlpou25. While this list of GO terms is not a striking observation, perhaps when combined with the enhanced stability of Xlpou91 induced pluripotent cells, and the data in

Chapter 3, it makes a strong case for functional specialization. After all, I showed that Xlopou91 appears unique in the POU5F3 family as having Oct4-like activity in ESC rescue assays and this correlates well with its capacity to induce germ cell specific markers. Taken together this suggests that after duplication in Anuran amphibians, Pou5f3 underwent revised specialization and regained germ cell activity similar to that present in its lost Pou5f1 paralog.

Is the reprogramming capacity of POUV genes related to their embryonic localization? Does the POUV gene expressed in germ cells always have the best reprogramming capacity? However, this specialization observed in *Xenopus* might be uniquely due to the fact that only Xlpou91 is expressed in PGCs. In the axolotl, both POU5F1 and POU5F3 are expressed in germ cells. Interestingly, axolotl OCT4/POU5F1 was as potent in maintaining ESCs as Xlpou91, but is not as effective at reprogramming as Xlpou91 and is actually slightly less efficient in iPSC generation than axolotl POU5F3 (Tapia et al., 2012). Based on the fact that some form of reactivation of the germ line program similar to reprogramming is a key event of germ cell development, the co-expression of both axolotl POU5F1 and POU5F3 in germ cells might have led to further sub-specialization of the germ cell network, such that the net activity is the same, but each protein only encodes partial regulatory activity. Future work on the combinatorial activities of axolotl POU5F1 and POU5F3 might provide the clue as to whether they act synergistically to induce pluripotency and why it is an advantage to promote this level of sub-specialization. Clearly in *Xenopus*, while there is some co-expression of these two proteins, Xlpou91 appears to need no contribution from Xlpou25 primed activity to faithfully reprogram somatic cells.



CHAPTER 5
FINAL DISCUSSION

CHAPTER 5: FINAL DISCUSSION

In this thesis, I investigated the functional conservation and divergence of vertebrate class V POU proteins. Two forms of POUV genes called *POU5F1* and *POU5F3* have been identified among jawed vertebrates (gnathostomes) (Frankenberg and Renfree, 2013). My study showed that there is only one *POUV* gene in lamprey (agnathan), suggesting that both *POU5F1* and *POU5F3* may have originated from a genome duplication event that occurred after the divergence of agnathan and gnathostome lineages. We can further group gnathostomes into three categories based on presence/absence of *POU5F1* and *POU5F3*: the presence of only *POU5F1* (e.g. mammals, squamates), only *POU5F3* (actinopterygian fish, birds), or both *POUV* forms (coelacanths, turtles, axolotls, marsupials). Through literature review and the analysis in this thesis, I can classify conserved and diverged POUV activities into four different levels.

a) Roles at organ or tissue level:

Species carrying a single *POUV* gene express it in two embryonic cell types: epiblast cells and germ cells (Fuhrmann et al., 1999; Lavial et al., 2007; Pelton et al., 2002; Sabour et al., 2010; Sánchez-Sánchez et al., 2009; Stebler et al., 2004). In addition, eutherians, which have lost *POU5F3* during evolution, show *POU5F1* expression in another unique structure called the inner cell mass (ICM) of the mammalian blastocyst, from which ESCs can be isolated (Nichols et al., 1998). The ICM shares expression of a large group of genes with the germs cells, so called naïve pluripotency (Leitch et al., 2013). Human ESCs and EpiSCs express genes more characteristic of later stages of development and exhibit a gene expression state referred to as primed pluripotency (Brons et al., 2007; Nichols and Smith, 2009; Rossant, 2008; Tesar et al., 2007). In species carrying more than one paralogs, POUV proteins are expressed at different tissues, as evidences in *Xenopus* and tammar wallaby embryos (Frankenberg et al., 2013; Morrison and Brickman, 2006; Venkatarama et

al., 2010). What are the distinct roles of POU5F1 and POU5F3 in embryos? Recently, there is no *in vivo* experiment to delineate the distinct roles of POUV paralogs at tissue level, due to the limitation of embryonic manipulation (e.g. tammar wallaby and turtle embryos) and unavailability of samples (e.g. coelacanth embryo). In this thesis, I harboured mouse embryonic stem cells as an *in vitro* tool to delineate the functional roles of different POUV homologs, and here I presented evidence that POU5F1 homologs in several vertebrate taxa play conserved roles related to a naïve or germ cell-like pluripotency, while POU5F3 homologs are more relevant to establishment and maintenance of a primed pluripotent state as found in the gastrulation stage epiblast. The extent to which POUV proteins exhibit different activities with respect to germ cell versus epiblast-related pluripotency varies from taxon to taxon, and likely depends on the actual expression sites.

b) Genetic functions and gene regulatory network:

Global transcriptome analysis showed that rescue of Pou5f1-null ESCs with coelacanth POU5F1 produces similar expression profiles to rescue with mouse Oct4. Several genes related to reproduction, stem cell maintenance and differentiation were specific to the POU5F1 lines. In contrast, rescue with coelacanth POU5F3 showed upregulation of genes involved in various cell differentiation programs, including cell integrity, thus representing the properties of the primed state. This clearly shows that POU5F1 and POU5F3 can induce different gene networks; however, some pluripotency genes are upregulated by both POUV paralogs, suggesting some degree of functional conservation. Future works on mass spectrometry analysis and ChIP-seq of POUV rescued ESC lines might unravel the pattern of transcriptional network cooperated with both POUVs or either POU5F1 or POU5F3. In addition, future work on global transcriptome on Oct4-null ESCs rescued by POUV from other species e.g. turtle and tammar wallaby might give higher resolution of how POUV network has evolved in different vertebrate taxa.

c) Cellular functions:

Based on the gene annotation analysis from microarray-based global transcriptome, coelacanth POU5F1 induces the expression of genes related to cell proliferation, chromatin binding, regulation of cell shape, stem cell differentiation, stem cell maintenance, cell differentiation, metal ion binding etc. Whereas, the POU5F3 induce expression of genes mostly related to cell adhesion, cell-cell adherens junction, extracellular matrix, basement membrane, integrin complex, cell migration etc. In addition, It is suggested by gene target analysis of *Xenopus* POUV, mouse and human ESCs that POUV network shared by both mammalian and non-mammalian species are cell adhesion, cell migration, cell proliferation and transcriptional regulation (Livigni et al., 2013). Based on this observation, it is interesting that species retaining only one POUV homologs seem to have both sets of POUV activities while species retaining both paralogs have segregated POUV proteins to perform functions at distinct cellular levels.

d) Molecular functions:

What could be the driving force behind the functional segregation of POUV proteins? One possibility is variations in the affinity of these proteins for DNA and its partners (Aksoy et al., 2013; Esch et al., 2013; Pardo et al., 2010; Remenyi, 2003). Thus variability in the expression level required to regulate a specific set of target genes enables a specific level of Oct4/Pou5f1 to support self-renewal with driving differentiation (Karwacki-Neisius et al., 2013; Radzisheuskaya et al., 2013). This is consistent with observations that comparisons of Oct4 fusion proteins with a greater capacity to activate transcription can maintain ESCs at lower levels (Hammachi et al., 2012). Below a threshold level of Oct4/Pou5f1 expression required to activate its targets, ESCs differentiate towards trophectoderm and mesoderm/primitive endoderm, respectively (Nichols et al., 1998; Niwa et al., 2000; Radzisheuskaya et al., 2013). In addition, the functional distinction of POUV paralogs might be due to differences in miRNA-related mRNA degradation and protein turnover rate.

Evolutionary trends of POUV activities

The genome duplication event that underlies POUV divergence has also produced homologs of several other transcription factor families, and animals have evolved several strategies to cope with the resulting functional redundancy. As the rate of evolution in the coelacanth protein-coding genes is very slow (Amemiya et al., 2013), the functional segregation observed here might probably be close to that achieved with the initial duplication event. Taken together, these observations suggest that the functional segregation of POU5F1 and POU5F3 may be an ancient trait that coincided with the birth of POUV paralogs when the gnathostome lineage first evolved. As POU5F1 and POU5F3 show a high degree of similarity in their conserved domains, it is plausible that functional segregation has become blurred over time in some vertebrate lineages. Thus, POU5F3 could be expected to regain partial germ cell function, while POU5F1 could regain some epiblast activity, depending on the species. Here I summarize the evolutionary trends of POUV activity into five patterns as follows: (summarised in figure 5.1)

Trend 1: Vertebrate ancestors might experience a phenomenon called "POUV subfunctionalization" after genome duplication, in which the ancient roles of a single POUV protein have been segregated into separated roles for POU5F1 and POU5F3, creating distinct temporal and spatial expression patterns of these POUV proteins during embryonic development, to reduce pressure created as a consequence of excess POUV expression from two alleles. Subsequently, the ancestral POUV targets have segregated, so that each protein regulates its own network. In this case, the germ cell network is regulated by POU5F1 and the epiblast by POU5F3. Several living vertebrate taxa display this trend of POUV activity, including coelacanth (predited), turtle (predicted), and tammar wallaby.

Trend 2: Axolotl POUV expression in the early embryo is intriguing in that both POUV proteins are expressed in its germ cells and epiblast (Tapia et al., 2012). My results

revealed that axolotl POU5F1 was more efficient at rescuing naïve ESCs than its paralog, POU5F3. This observation is still in line with the functional segregation between POU5F1 and POU5F3 described in trend 1. If complete functional segregation of POUV proteins (trend 1) is an ancient trait, the emergence of POU5F3 in axolotl germ cells could be a secondary derived trait. Perhaps this is the reason that axolotl POU5F3 induces Prdm14 (germ cell marker) at a higher level than do other POU5F3s, thus corroborating the notion that axolotl POU5F3 regains at least some control over the germ cell network.

Trend 3 and 4: Most vertebrate taxa eliminate the redundancy by losing one of the POUV paralog. In species that have lost either POU5F1 or POU5F3 during evolution, the remaining POUV protein has to regulate the complete set of genes to support both levels of pluripotency. For example, mouse Oct4 (a POU5F1) is expressed in both naïve ESC/ICM/germ cells and primed epiblast/EpiSC (Fuhrmann et al., 1999; Nichols et al., 1998; Pelton et al., 2002; Sabour et al., 2010), and chicken POUV (a POU5F3) is expressed in both gastrulating epiblast and germ cells (Lavial et al., 2007).

Trend 5: A special case of POUV evolution occurred in the frog lineage. Frogs have lost *pou5f1* and retain only *pou5f3*, but further duplication events have led to the emergence of three *pou5f3* paralogs in these species. This vertebrate lineage has also coped with POUV redundancy through functional segregation, creating distinct spatial and temporal expression patterns of the three Pou5f3 proteins. These Pou5f3s are called Xlpou91 (Pou5f3.1), Xlpou25 (Pou5f3.2) and Xlpou60 (Pou5f3.3). Xlpou60 is expressed predominantly prior mid-blastula transition (Hinkley et al., 1992). Xlpou25 is expressed predominantly during gastrulation, but later also in neural specification and in the posterior neural tube (Cao, 2004). Xlpou25 has also been shown to be an important regulator of gastrulation (Livigni et al., 2013), while Xlpou91 is expressed in similar embryonic region to Xlpou25, but has an additional domain of expression in the germ cells, suggesting its role as germ cell regulator (Hinkley et al., 1992). The rescue assay showed that Xlpou25 has a similar ESC phenotype to other POU5F3s, while Xlpou91 has a similar capacity to Oct4 and other POU5F1 protein

to support ESC self-renewal. I also found that the naïve ESC-like populations in iPSCs derived with different *Xenopus* POUVs are very similar, but exhibit two striking differences. First Xlpou91 and mOct4 induce more genes associated with reproduction than does Xlpou25. Second, the Xlpou91 and mOct4 derived iPSCs contain an ESC-like fraction that is more stable and less differentiated. The Oct4-like function of Xlpou91 constitutes a clear example of convergent evolution, a rare evolutionary event that took place in POUV history.

Taken together, these evolutionary trends in POUV activity (functional segregation, respecialization, and reduction of POUV activities based on redundancy level) emphasise that POUV proteins are flexible in switching between two modes of activity: "Germ cell activity" and "Epiblast activity". This flexibility is likely managed through cooperative binding, so that minor amino acid changes alter the set of partners preferred by a particular POU protein, but no so drastically that it is no longer able to see the remaining set. In support of this is the notion that the tested POU5F1 and POU5F3 proteins are mostly able to activate transcription from Luciferase reporters driven by different promoter sequences defined based on Oct4 targets (Hammachi et al., 2012; Morrison and Brickman, 2006; Niwa et al., 2000), suggesting that affinity differences that distinguish POU5F1 and POU5F3 are subtle. However, these subtle and easy to engineer changes in POUV proteins might have had a key influence on shaping the pattern of vertebrate embryonic development, supporting their success in species radiation. Interestingly, vertebrate taxa that have lost POU5F1 switch from an ancient inductive mode of germ cell specification towards a predetermined mode. Are these events related? I have shown that POU5F1 induces more reproduction-related genes than its paralog, and many of those reproduction-related genes (e.g Prdm14, Nanos3) are involved in the initiation of germ cell programs induced through paracrine signaling in the epiblast. The loss of POU5F1 may affect the competency of epiblast cells to respond to these signals, which could have provided an evolutionary force to drive animals without POU5F1 to acquire another source of germ cell network induction.

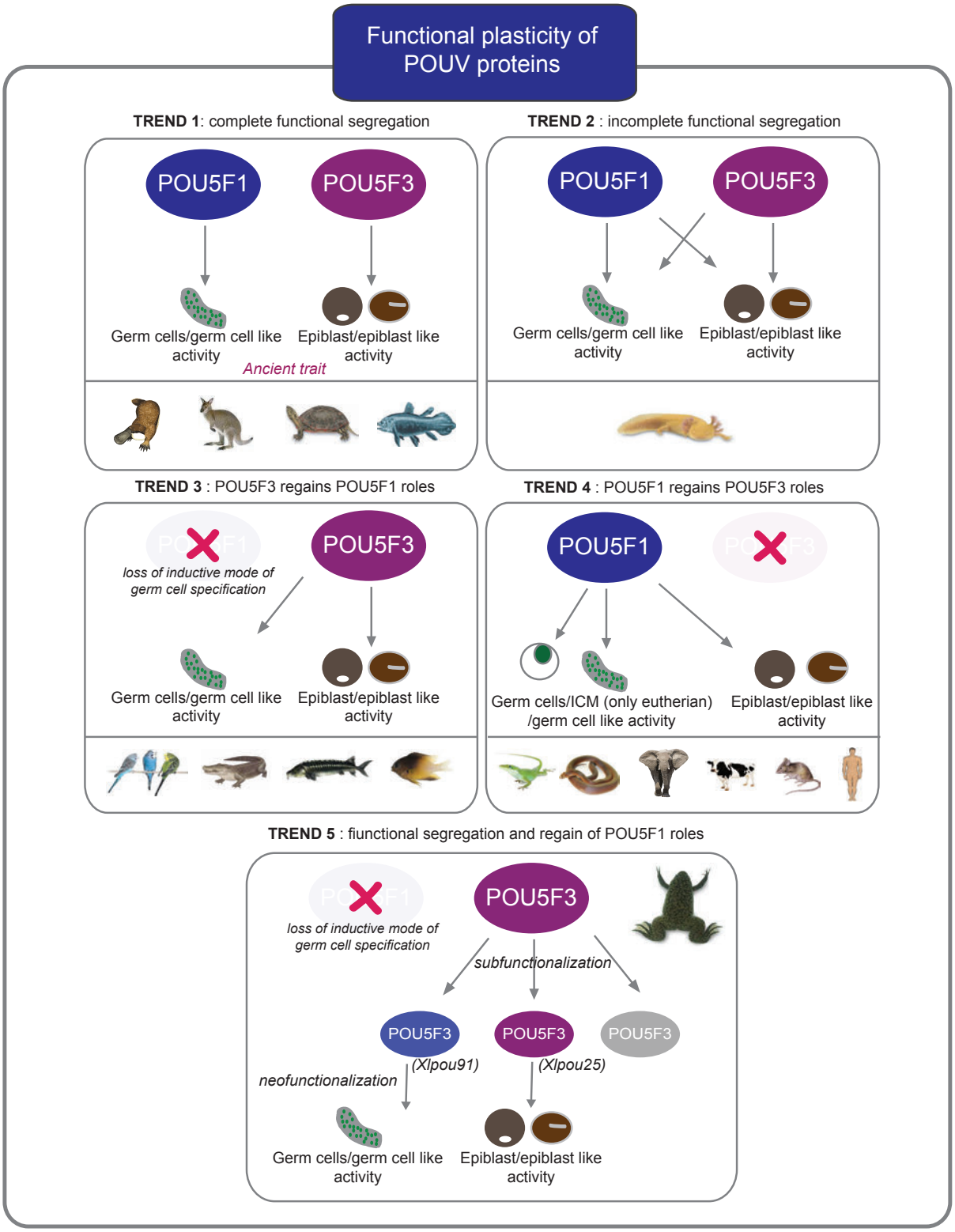


Figure 5.1 Modes of vertebrate POUV activities and the plasticity of POUV network

Vertebrate ancestors might experience a phenomenon called "POUV subfunctionalization" after genome duplication, in which the ancient roles of a single POUV protein have been segregated into separated roles for POU5F1 and POU5F3 proteins. In this thesis, I have presented the evidences that in species carrying both POUV forms, POU5F1 roles are related to germ cell network and reproduction while POU5F3 roles are more related to cell adhesion and integrity of gastrulation-stage epiblast **(Trend 1)**. As POU5F1 and POU5F3 show a high degree of similarity in their conserved domains, it is plausible that functional segregation has become blurred over time in some vertebrate lineages. For example, axolotl embryo expresses POU5F1 and POU5F3 in both germ cells and epiblast **(Trend 2)**. In the species carrying only single POUV form, their POUV proteins regain the conserved activity of the lost paralog and perform both germ cell and epiblast like activities **(Trend 3 and 4)**. In eutherians (using mouse as a model), pluripotent stem cells isolated from inner cell mass and germ cells share similar gene expression profile and epigenetic status and both tissues express Pou5f1/Oct4. This leads to the idea that conserved germ cell network regulated by Pou5f1 might evolve to be a part of the transcriptional network regulating inner cell mass/ESC pluripotency **(Trend 4)**. There is an exceptional trend of POUV activities in frog lineage (using *Xenopus laevis* as a model). In *Xenopus*, there are three Pou5f3 genes which probably originated through tandem gene duplication from a single Pou5f3 gene. Frog Pou5f3 proteins have underwent further subfunctionalization, in that Xlpou91 (Pou5f3.1) and Xlpou25 (Pou5f3.2) have distinct roles in germ cells and epiblast, respectively **(Trend 5)**. The re-emergence of POU5F1 roles in *Xenopus* Pou91 (through neofunctionalization) represents an example of convergent evolution.

REFERENCES

REFERENCES

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. The EMBO Journal *32*, 938–953.

Amemiya, C.T., Alföldi, J., Lee, A.P., Fan, S., Philippe, H., Maccallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. Nature *496*, 311–316.

Amemiya, C.T., Dorrington, R., and Meyer, A. (2014). The coelacanth and its genome. J. Exp. Zool. B Mol. Dev. Evol. *322*, 317–321.

Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J.K., and Lonai, P. (1998). Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. Proceedings of the National Academy of Sciences *95*, 5082–5087.

Arnason, U., Gullberg, A., and Janke, A. (2001). Molecular phylogenetics of gnathostomous (jawed) fishes: old bones, new cartilage. Zoologica Scripta *30*, 249–255.

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol *10*, 91–103.

Artus, J., Kang, M., Cohen-Tannoudji, M., and Hadjantonakis, A.-K. (2013). PDGF signaling is required for primitive endoderm cell survival in the inner cell mass of the mouse blastocyst. Stem Cells *31*, 1932–1941.

Artus, J., Piliszek, A., and Hadjantonakis, A.-K. (2011). The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. Developmental Biology *350*, 393–404.

Assou, S., Le Carrour, T., Tondeur, S., Ström, S., Gabelle, A., Marty, S., Nadal, L., Pantesco, V., Réme, T., Hugnot, J.-P., et al. (2007). A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells *25*, 961–973.

Atlasi, Y., Noori, R., Gaspar, C., Franken, P., Sacchetti, A., Rafati, H., Mahmoudi, T., Decraene, C., Calin, G.A., Merrill, B.J., et al. (2013). Wnt Signaling Regulates the Lineage Differentiation Potential of Mouse Embryonic Stem Cells through Tcf3 Down-Regulation. PLoS Genet *9*, e1003424.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & Development *17*, 126–140.

Bachvarova, R.F., Crother, B.I., Manova, K., Chatfield, J., Shoemaker, C.M., Crews, D.P., and Johnson, A.D. (2009). Expression of Dazl and Vasa in turtle embryos and ovaries: evidence for inductive specification of germ cells. Evolution & Development *11*, 525–534.

Baker, M. (2009). A new role for Oct4. Nature Reports Stem Cells.

Belting, H.-G., Hauptmann, G., Meyer, D., Abdelilah-Seyfried, S., Chitnis, A., Eschbach, C., Söll, I., Thisse, C., Thisse, B., Artinger, K.B., et al. (2001). spiel ohne grenzen/pou2 is required during establishment of the zebrafish midbrain-hindbrain boundary organizer. Development *128*, 4165–4176.

Blair, J.E., and Hedges, S.B. (2005). Molecular phylogeny and divergence times of deuterostome animals. Mol. Biol. Evol. *22*, 2275–2284.

Boland, M.J., Nazor, K.L., and Loring, J.F. (2014). Epigenetic regulation of pluripotency and differentiation. Circulation Research *115*, 311–324.

Boroviak, T., and Nichols, J. (2014). The birth of embryonic pluripotency. Philosophical Transactions of the Royal Society B: Biological Sciences *369*, 20130541–20130541.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat Cell Biol *16*, 516–528.

Bourillot, P.-Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., Hubner, N., and Savatier, P. (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.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell *122*, 947–956.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Brook, F.A., and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. Proceedings of the National Academy of Sciences *94*, 5709–5712.

Broughton, R.E., Betancur-R, R., Li, C., Arratia, G., and Ortí, G. (2013). Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Currents.

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

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. Cell *150*, 1209–1222.

Cai, K.Q., Capo-Chichi, C.D., Rula, M.E., Yang, D.-H., and Xu, X.-X. (2008). Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis. Dev. Dyn. *237*, 2820–2829.

Campbell, P.A., Perez-Iratxeta, C., Andrade-Navarro, M.A., and Rudnicki, M.A. (2007). Oct4 Targets Regulatory Nodes to Modulate Stem Cell Function. PLoS ONE *2*, e553.

Canham, M.A., Sharov, A.A., Ko, M.S.H., and Brickman, J.M. (2010). Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol. *8*, e1000379.

Cañestro, C., Yokoi, H., and Postlethwait, J.H. (2007). Evolutionary developmental biology and genomics. Nature Reviews Genetics *8*, 932–942.

Cao, Y. (2004). The POU Factor Oct-25 Regulates the Xvent-2B Gene and Counteracts Terminal Differentiation in Xenopus Embryos. Journal of Biological Chemistry *279*, 43735–43743.

Cao, Y., Siegel, D., Oswald, F., and Knochel, W. (2008). Oct25 Represses Transcription of Nodal/Activin Target Genes by Interaction with Signal Transducers during Xenopus Gastrulation. Journal of Biological Chemistry *283*, 34168–34177.

Casane, D., and Laurenti, P. (2013). Why coelacanths are not “living fossils”: a review of molecular and morphological data. Bioessays *35*, 332–338.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Developmental Cell *10*, 615–624.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet *45*, 34–42.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008a). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106–1117.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008b). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106–1117.

Cheng, A.M., Saxton, T.M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C.G., Cardiff, R.D., Cross, J.C., and Muller, W.J. (1998). Mammalian Grb2 Regulates Multiple Steps in Embryonic Development and Malignant Transformation. Cell *95*, 793–803.

Chenoweth, J.G., McKay, R.D.G., and Tesar, P.J. (2010). Epiblast stem cells contribute new insight into pluripotency and gastrulation. Develop. Growth Differ. *52*, 293–301.

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., et al. (2005). Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. Molecular and Cellular Biology *25*, 6031–6046.

Chiari, Y., Cahais, V., Galtier, N., and Delsuc, F. (2012). Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biol *10*, 65.

Chu, G.C. (2004). Differential requirements for Smad4 in TGF -dependent patterning of the early mouse embryo. Development *131*, 3501–3512.

Covello, K.L. (2006). HIF-2 regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. Genes & Development *20*, 557–570.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K., and Glenn, T.C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. *8*, 783–786.

David Archibald, J. (2003). Timing and biogeography of the eutherian radiation: fossils and molecules compared. Mol. Phylogenet. Evol. *28*, 350–359.

De Felici, M., Scaldaferri, M.L., and Farini, D. (2005). Adhesion molecules for mouse primordial germ cells. Front. Biosci. *10*, 542–551.

de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejčić, M., Torruella, G., Domazet-Lošo, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proceedings of the National Academy of Sciences *110*, E4858–E4866.

Deshpande, A.M., Dai, Y.S., Kim, Y., Kim, J., Kimlin, L., Gao, K., and Wong, D.T. (2008). Cdk2ap1 Is Required for Epigenetic Silencing of Oct4 during Murine Embryonic Stem Cell Differentiation. Journal of Biological Chemistry *284*, 6043–6047.

Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature *439*, 965–968.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P.W., et al. (2013). Oct4 Is Required ~E7.5 for Proliferation in the Primitive Streak. PLoS Genet *9*, e1003957.

Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W., et al. (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes & Development *27*, 1378–1390.

Donoghue, P.C.J., and Keating, J.N. (2014). Early vertebrate evolution. Palaeontology *57*, 879–893.

Dores, R.M. (2011). Hagfish, Genome Duplications, and RFamide Neuropeptide Evolution. Endocrinology *152*, 4010–4013.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. *29*, 1969–1973.

Dunn, N.R. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. Development *131*, 1717–1728.

Durcovahills, G., and Mclaren, D. (2004). Isolation and Maintenance of Murine Embryonic Germ Cell Lines. In Handbook of Stem Cells, (Elsevier), pp. 451–457.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Bruch, Vom, H., Han, D., Drexler, H.C.A., Araúzo-Bravo, M.J., Ng, C.K.L., et al. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nature Publishing Group *15*, 295–301.

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., et al. (2010). Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. Cell Stem Cell *6*, 71–79.

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

Extavour, C.G. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development *130*, 5869–5884.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol *8*, 188–194.

Frankenberg, S., Pask, A., and Renfree, M.B. (2010). The evolution of class V POU domain transcription factors in vertebrates and their characterisation in a marsupial. Developmental Biology.

Frankenberg, S., and Renfree, M.B. (2013). On the origin of POU5F1. BMC Biol *11*, 56.

Frankenberg, S., Shaw, G., Freyer, C., Pask, A.J., and Renfree, M.B. (2013). Early cell lineage specification in a marsupial: a case for diverse mechanisms among mammals. Development *140*, 965–975.

Frum, T., Halbisen, M.A., Wang, C., Amiri, H., Robson, P., and Ralston, A. (2013). Oct4 Cell-Autonomously Promotes Primitive Endoderm Development in the Mouse Blastocyst. Developmental Cell *25*, 610–622.

Fuhrmann, G., Sylvester, I., and Scholer, H.R. (1999). Repression of Oct-4 during embryonic cell differentiation correlates with the appearance of TRIF, a transiently induced DNA-binding factor. Cell. Mol. Biol. (Noisy-Le-Grand) *45*, 717–724.

Gauthier, J., Kluge, A.G., and Rowe, T. (1988). Amniote phylogeny and the importance of fossils. Cladistics *4*, 105–209.

Gerbe, F., Cox, B., Rossant, J., and Chazaud, C. (2008). Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. Developmental Biology *313*, 594–602.

Gold, D.A., Gates, R.D., and Jacobs, D.K. (2014). The Early Expansion and Evolutionary Dynamics of POU Class Genes. Mol. Biol. Evol. *31*, 3136–3147.

Greer Card, D.A., Hebbar, P.B., Li, L., Trotter, K.W., Komatsu, Y., Mishina, Y., and Archer, T.K. (2008). Oct4/Sox2-Regulated miR-302 Targets Cyclin D1 in Human Embryonic Stem Cells. Molecular and Cellular Biology *28*, 6426–6438.

Grubb, B.J. (2006). Developmental Biology, Eighth Edition. Scott F. Gilbert, editor. Integrative and Comparative Biology *46*, 652–653.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.

Gurdon, J.B., ELSDALE, T.R., and FISCHBERG, M. (1958). Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. Nature *182*, 64–65.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., and Grotewold, L. (2009). Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell.

Hamburger, V. (1988). The Heritage of Experimental Embryology (Oxford University Press on Demand)

Hammachi, F., Morrison, G.M., Sharov, A.A., Livigni, A., Narayan, S., Papapetrou, E.P., O'Malley, J., Kaji, K., Ko, M.S.H., Ptashne, M., et al. (2012). Transcriptional Activation by Oct4 Is Sufficient for the Maintenance and Induction of Pluripotency. Cell Reports *1*, 99–109.

Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T., et al. (2010). Epiblast Stem Cell Subpopulations Represent Mouse Embryosof Distinct Pregastrulation Stages. Cell 1–17.

Hatano, S.-Y., Tada, M., Kimura, H., Yamaguchi, S., Kono, T., Nakano, T., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Pluripotential competence of cells associated with Nanog activity. Mechanisms of Development *122*, 67–79.

Hayashi, K., Lopes, S.M.C. de S., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell *3*, 391–401.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell *146*, 519–532.

Heard, E. (2004). Recent advances in X-chromosome inactivation. Current Opinion in Cell Biology *16*, 247–255.

Hedges, S.B., and Kumar, S. (2009). The Timetree of Life (Oxford University Press).

Hinkley, C.S., Martin, J.F., Leibham, D., and Perry, M. (1992). Sequential expression of multiple POU proteins during amphibian early development. Molecular and Cellular Biology *12*, 638–649.

Hirate, Y., Hirahara, S., Inoue, K.-I., Suzuki, A., Alarcon, V.B., Akimoto, K., Hirai, T., Hara, T., Adachi, M., Chida, K., et al. (2013). Polarity-Dependent Distribution of Angiomotin Localizes Hippo Signaling in Preimplantation Embryos. Current Biology *23*, 1181–1194.

Hochedlinger, K., and Jaenisch, R. (2006). Nuclear reprogramming and pluripotency. Nature *441*, 1061–1067.

Holland, L.Z., Albalat, R., Azumi, K., Benito-Gutiérrez, E., Blow, M.J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L.J., et al. (2008). The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. *18*, 1100–1111.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., et al. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science *341*, 651–654.

Hugall, A.F., Foster, R., and Lee, M.S.Y. (2007). Calibration Choice, Rate Smoothing, and the Pattern of Tetrapod Diversification According to the Long Nuclear Gene RAG-1. Systematic Biology *56*, 543–563.

Hurley, I.A., Mueller, R.L., Dunn, K.A., Schmidt, E.J., Friedman, M., Ho, R.K., Prince, V.E., Yang, Z., Thomas, M.G., and Coates, M.I. (2007). A new time-scale for ray-finned fish evolution. Proc. Biol. Sci. *274*, 489–498.

Ijiri, K., Narita, T., and Mizuno, R. (1996). Primordial germ cells in the embryos of Medaka fish. Biol. Sci. Space *10*, 156–157.

Inoue, J.G., Miya, M., Tsukamoto, K., and Nishida, M. (2003). Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol. Phylogenet. Evol. *26*, 110–120.

Inoue, J.G., Miya, M., Venkatesh, B., and Nishida, M. (2005). The mitochondrial genome of Indonesian coelacanth Latimeria menadoensis (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene *349*, 227–235.

Iwamatsu, T. (2004). Stages of normal development in the medaka Oryzias latipes. Mechanisms of Development *121*, 605–618.

Janecka, J.E., Miller, W., Pringle, T.H., Wiens, F., Zitzmann, A., Helgen, K.M., Springer, M.S., and Murphy, W.J. (2007). Molecular and genomic data identify the closest living relative of primates. Science *318*, 792–794.

Jang, H., Kim, T.W., Yoon, S., Choi, S.-Y., Kang, T.-W., Kim, S.-Y., Kwon, Y.-W., Cho, E.-J., and Youn, H.-D. (2012). O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. Cell Stem Cell *11*, 62–74.

Jen, Y., Manova, K., and Benezra, R. (1997). Each member of the Id gene family exhibits a unique expression pattern in mouse gastrulation and neurogenesis. Dev. Dyn. *208*, 92–106.

Ji, Q., Luo, Z.-X., Yuan, C.-X., Wible, J.R., Zhang, J.-P., and Georgi, J.A. (2002). The earliest known eutherian mammal. Nature *416*, 816–822.

Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., and Ng, H.-H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353–360.

Johanson, Z., Long, J.A., Talent, J.A., Janvier, P., and Warren, J.W. (2006). Oldest coelacanth, from the Early Devonian of Australia. Biol. Lett. *2*, 443–446.

Johnson, A.D., Crother, B., White, M.E., Patient, R., Bachvarova, R.F., Drum, M., and Masi, T. (2003). Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. Philosophical Transactions of the Royal Society B: Biological Sciences *358*, 1371–1379.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Bioinformatics *8*, 275–282.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771–775.

Kang, M. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Development *140*, 267–279.

Karwacki-Neisius, V., Göke, J., Osorno, R., Halbritter, F., Ng, J.-H., Weiße, A.Y., Wong, F.C.K., Gagliardi, A., Mullin, N.P., Festuccia, N., et al. (2013). Reduced Oct4 Expression Directs a Robust Pluripotent State with Distinct Signaling Activity and Increased Enhancer Occupancy by Oct4 and Nanog. Cell Stem Cell *12*, 531–545.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Izpisúa Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140–1144.

Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomelí, H., Nagy, A., McLaughlin, K.J., Schöler, H.R., et al. (2004). Oct4 is required for primordial germ cell survival. EMBO Reports *5*, 1078–1083.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell *132*, 1049–1061.

Kim, J.B., Greber, B., Araúzo-Bravo, M.J., Meyer, J., Park, K.I., Zaehres, H., and Schöler, H.R. (2009). Direct reprogramming of human neural stem cells by OCT4. Nature *461*, 649–653.

Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T., and Kumagai, H. (2003). Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. Experimental Hematology *31*, 1007–1014.

Kobayashi, S., Amikura, R., and Mukai, M. (1998). Localization of mitochondrial large ribosomal RNA in germ plasm of Xenopus embryos. Current Biology *8*, 1117–1120.

Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A.F., Jones, V., Hor, A., de Alencastro, G., Logan, G.J., et al. (2014). The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak. Cell Stem Cell *14*, 107–120.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells *26*, 903–911.

Kraushaar, D.C., and Zhao, K. (2013). The Epigenomics of Embryonic Stem Cell Differentiation. International Journal of Biological Sciences *9*, 1134–1144.

Kumar, S., and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. Nature *392*, 917–920.

Kuraku, S., and Kuratani, S. (2006). Time Scale for Cyclostome Evolution Inferred with a Phylogenetic Diagnosis of Hagfish and Lamprey cDNA Sequences. Zoological Science *23*, 1053–1064.

Kuraku, S. (2008). Insights into Cyclostome Phylogenomics: Pre-2R or Post-2R. Zoological Science *25*, 960–968.

Kuraku, S., Meyer, A., and Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol. Biol. Evol. *26*, 47–59.

Lachnit, M., Kur, E., and Driever, W. (2008). Alterations of the cytoskeleton in all three embryonic lineages contribute to the epiboly defect of Pou5f1/Oct4 deficient MZspg zebrafish embryos. Developmental Biology *315*, 1–17.

Laurin, M., and Reisz, R.R. (1997). A New Perspective on Tetrapod Phylogeny. In Amniote Origins, (Elsevier), pp. 9–59.

Lavial, F., Acloque, H., Bertocchini, F., MacLeod, D.J., Boast, S., Bachelard, E., Montillet, G., Thenot, S., Sang, H.M., Stern, C.D., et al. (2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. Development *134*, 3549–3563.

Lawson, K.A. (1999). Fate mapping the mouse embryo. Int. J. Dev. Biol. *43*, 773–775.

Lee, M.S. (1993). The origin of the turtle body plan: bridging a famous morphological gap. Science *261*, 1716–1720.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Leon Hughes, R., and Hall, L.S. (1998). Early development and embryology of the platypus. Philosophical Transactions of the Royal Society of London B: Biological Sciences *353*, 1101–1114.

Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., et al. (2007a). Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Molecular and Cellular Biology *27*, 8748–8759.

Li, J., Pan, G., Cui, K., Liu, Y., Xu, S., and Pei, D. (2007b). A dominant-negative form of mouse SOX2 induces trophectoderm differentiation and progressive polyploidy in mouse embryonic stem cells. J. Biol. Chem. *282*, 19481–19492.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. Stem Cell *7*, 51–63.

Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., Ge, J., Liu, C., Zhang, W., Zhang, X., et al. (2010). Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. Cell Res *21*, 196–204.

Li, W., Wei, W., Zhu, S., Zhu, J., Shi, Y., Lin, T., Hao, E., Hayek, A., Deng, H., and Ding, S. (2009). Generation of Rat and Human Induced Pluripotent Stem Cells by Combining Genetic Reprogramming and Chemical Inhibitors. Cell Stem Cell *4*, 16–19.

Lichner, Z., Páll, E., Kerekes, A., Pállinger, É., Maraghechi, P., Bősze, Z., and Gócza, E. (2011). The miR-290-295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells. Differentiation *81*, 11–24.

Livigni, A., Peradziryi, H., Sharov, A.A., Chia, G., Hammachi, F., Migueles, R.P., Sukparangsi, W., Pernagallo, S., Bradley, M., Nichols, J., et al. (2013). A Conserved Oct4/POUV-Dependent Network Links Adhesion and Migration to Progenitor Maintenance. Current Biology *23*, 2233–2244.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet *42*, 1113–1117.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431–440.

Lowe, C.B., Kellis, M., Siepel, A., Raney, B.J., Clamp, M., Salama, S.R., Kingsley, D.M., Lindblad-Toh, K., and Haussler, D. (2011). Three Periods of Regulatory Innovation During Vertebrate Evolution. Science *333*, 1019–1024.

Lunde, K., Belting, H.-G., and Driever, W. (2004). Zebrafish pou5f1/pou2, Homolog of Mammalian Oct4, Functions in the Endoderm Specification Cascade. Current Biology *14*, 48–55.

Luo, Z.-X. (2007). Transformation and diversification in early mammal evolution. Nature *450*, 1011–1019.

Lyson, T.R., Bever, G.S., Bhullar, B.A.S., Joyce, W.G., and Gauthier, J.A. (2010). Transitional fossils and the origin of turtles. Biol. Lett. *6*, 830–833.

M, K., A, N., S, M., and Y, K. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature *447*, 714–719.

Ma, G.T., Roth, M.E., Groskopf, J.C., Tsai, F.Y., Orkin, S.H., Grosveld, F., Engel, J.D., and Linzer, D.I. (1997). GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. Development *124*, 907–914.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nature Publishing Group *15*, 905–915.

Mallatt, J., and Winchell, C.J. (2007). Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. Mol. Phylogenet. Evol. *43*, 1005–1022.

Mannen, H., and Li, S.S.L. (1999). Molecular Evidence for a Clade of Turtles. Mol. Phylogenet. Evol. *13*, 144–148.

Marjanovic, D., and Laurin, M. (2007). Fossils, Molecules, Divergence Times, and the Origin of Lissamphibians. Systematic Biology *56*, 369–388.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625–635.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.A., and Ko, M.S.H. (2006). Dissecting Oct3/4-Regulated Gene Networks in Embryonic Stem Cells by Expression Profiling. PLoS ONE *1*, e26.

Meilhac, S.M., Adams, R.J., Morris, S.A., Danckaert, A., Le Garrec, J.-F., and Zernicka-Goetz, M. (2009). Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. Developmental Biology *331*, 210–221.

Millane, R.C., Kanska, J., Duffy, D.J., Seoighe, C., Cunningham, S., Plickert, G., and Frank, U. (2011). Induced stem cell neoplasia in a cnidarian by ectopic expression of a POU domain transcription factor. Development *138*, 2429–2439.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. (IEEE), pp. 1–8.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Miyagi, S., Masui, S., Niwa, H., Saito, T., Shimazaki, T., Okano, H., Nishimoto, M., Muramatsu, M., Iwama, A., and Okuda, A. (2008). Consequence of the loss of Sox2 in the developing brain of the mouse. FEBS Lett. *582*, 2811–2815.

Moody, S.A. (1987). Fates of the blastomeres of the 32-cell-stage Xenopus embryo. Developmental Biology *122*, 300–319.

Morgani, S.M., Canham, M.A., Nichols, J., Sharov, A.A., Migueles, R.P., Ko, M.S.H., and Brickman, J.M. (2013). Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. Cell Reports *3*, 1945–1957.

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

Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S., and Miller, W. (2007). Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. *17*, 413–421.

Nakaya, Y., and Sheng, G. (2008). Epithelial to mesenchymal transition during gastrulation: an embryological view. Develop. Growth Differ. *50*, 755–766.

Nakaya, Y., Sukowati, E.W., and Sheng, G. (2013). Epiblast integrity requires CLASP and Dystroglycan-mediated microtubule anchoring to the basal cortex. J. Cell Biol. *202*, 637–651.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., and Smith, W.L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the National Academy of Sciences *109*, 13698–13703.

Ng, H.-H., and Surani, M.A. (2011). The transcriptional and signalling networks of pluripotency. Nature Publishing Group *13*, 490–496.

Nichols, J., and Smith, A. (2009). Perspective. Stem Cell *4*, 487–492.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. Cell *95*, 379–391.

Niimi, T., Hayashi, Y., Futaki, S., and Sekiguchi, K. (2004). SOX7 and SOX17 Regulate the Parietal Endoderm-specific Enhancer Activity of Mouse Laminin 1 Gene. J. Biol. Chem. *279*, 38055–38061.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372–376.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

Oatley, J.M., and Brinster, R.L. (2008). Regulation of Spermatogonial Stem Cell Self-Renewal in Mammals. Annu. Rev. Cell Dev. Biol. *24*, 263–286.

Oisi, Y., Ota, K.G., Kuraku, S., Fujimoto, S., and Kuratani, S. (2013). Craniofacial development of hagfishes and the evolution of vertebrates. Nature *493*, 175–180.

Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M., and Hamada, H. (1991). The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. The EMBO Journal *10*, 2997–3005.

Ovitt, C.E., and Scholer, H.R. (1998). The molecular biology of Oct-4 in the early mouse embryo. Mol. Hum. Reprod. *4*, 1021–1031.

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88–91.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. Stem Cell *6*, 382–395.

Paton, T., Haddrath, O., and Baker, A.J. (2002). Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc. Biol. Sci. *269*, 839–846.

Pelton, T.A., Sharma, S., Schulz, T.C., Rathjen, J., and Rathjen, P.D. (2002). Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. J. Cell. Sci. *115*, 329–339.

Peng, Z., He, S., Wang, J., Wang, W., and Diogo, R. (2006). Mitochondrial molecular clocks and the origin of the major Otocephalan clades (Pisces: Teleostei): A new insight. Gene *370*, 113–124.

Pereira, S.L. (2006). A Mitogenomic Timescale for Birds Detects Variable Phylogenetic Rates of Molecular Evolution and Refutes the Standard Molecular Clock. Mol. Biol. Evol. *23*, 1731–1740.

Phillips, B.T., Gassei, K., and Orwig, K.E. (2010). Spermatogonial stem cell regulation and spermatogenesis. Philosophical Transactions of the Royal Society B: Biological Sciences *365*, 1663–1678.

Plickert, G., Frank, U., and Müller, W.A. (2012). Hydractinia, a pioneering model for stem cell biology and reprogramming somatic cells to pluripotency. Int. J. Dev. Biol. *56*, 519–534.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development *135*, 3081–3091.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Potter, I.C., Gill, H.S., Renaud, C.B., and Haoucher, D. (2014). The Taxonomy, Phylogeny, and Distribution of Lampreys. In Lampreys: Biology, Conservation and Control, (Dordrecht: Springer Netherlands), pp. 35–73.

Price, F.D., Yin, H., Jones, A., van Ijcken, W., Grosveld, F., and Rudnicki, M.A. (2013). Canonical Wnt signaling induces a primitive endoderm metastable state in mouse embryonic stem cells. Stem Cells *31*, 752–764.

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

Radzisheuskaya, A., and Silva, J.C.R. (2013). Do all roads lead to Oct4? The emerging concepts of induced pluripotency. Trends in Cell Biology 1–10.

Radzisheuskaya, A., Chia, G.L.B., Santos, dos, R.L., Theunissen, T.W., Castro, L.F.C., Nichols, J., and Silva, J.C.R. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nature Publishing Group *15*, 579–590.

Ralston, A., and Rossant, J. (2009). The genetics of induced pluripotency. Reproduction *139*, 35–44.

Rathjen, P.D., Toth, S., Willis, A., Heath, J.K., and Smith, A.G. (1990). Differentiation inhibiting activity is produced in matrix-associated and diffusible forms that are generated by alternate promoter usage. Cell 62, 1105–1114.

Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W., and Besser, D. (2011). scientific report. EMBO Reports 1–7.

Reim, G., and Brand, M. (2002). spiel-ohne-grenzen/pou2 mediates regional competence to respond to Fgf8 during zebrafish early neural development. Development *129*, 917–933.

Reim, G., Mizoguchi, T., Stainier, D.Y., Kikuchi, Y., and Brand, M. (2004). The POU Domain Protein Spg (Pou2/Oct4) Is Essential for Endoderm Formation in Cooperation with the HMG Domain Protein Casanova. Developmental Cell *6*, 91–101.

Remenyi, A. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes & Development *17*, 2048–2059.

Rizzino, A. (2009). Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells. Wiley Interdisciplinary Reviews: Systems Biology and Medicine *1*, 228–236.

Roca, A.L., Bar-Gal, G.K., Eizirik, E., Helgen, K.M., Maria, R., Springer, M.S., O'Brien, S.J., and Murphy, W.J. (2004). Mesozoic origin for West Indian insectivores. Nature *429*, 649–651.

Romero-Lanman, E.E., Pavlovic, S., Amlani, B., Chin, Y., and Benezra, R. (2012). Id1 Maintains Embryonic Stem Cell Self-Renewal by Up-Regulation of Nanog and Repression of Brachyury Expression. Stem Cells and Development *21*, 384–393.

Rossant, J., and Cross, J.C. (2001). Placental development: lessons from mouse mutants. Nature Reviews Genetics *2*, 538–548.

Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527–531.

Rossello, R.A., Chen, C.C., Dai, R., Howard, J.T., Hochgeschwender, U., and Jarvis, E.D. (2013). Mammalian genes induce partially reprogrammed pluripotent stem cells in non-mammalian vertebrate and invertebrate species. eLife *2*, e00036–e00036.

Rugg-Gunn, P.J., Cox, B.J., Lanner, F., Sharma, P., Ignatchenko, V., McDonald, A.C.H., Garner, J., Gramolini, A.O., Rossant, J., and Kislinger, T. (2012). Cell-Surface Proteomics Identifies Lineage-Specific Markers of Embryo-Derived Stem Cells. Developmental Cell *22*, 887–901.

Sabour, D., Arauzo-Bravo, M.J., Hubner, K., Ko, K., Greber, B., Gentile, L., Stehling, M., and Scholer, H.R. (2010). Identification of genes specific to mouse primordial germ cells through dynamic global gene expression. Human Molecular Genetics *20*, 115–125.

Saito, T., Pšenička, M., Goto, R., Adachi, S., Inoue, K., Arai, K., and Yamaha, E. (2014). The Origin And Migration Of Primordial Germ Cells In Sturgeons. PLoS ONE *9*, e86861.

Saitou, M., and Yamaji, M. (2010). Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences. Reproduction *139*, 931–942.

Saitou, M., and Yamaji, M. (2012). Primordial Germ Cells in Mice. Cold Spring Harbor Perspectives in Biology *4*, a008375–a008375.

Salamat, M., Miosge, N., and Herken, R. (1995). Development of Reichert's membrane in the early mouse embryo. Anat. Embryol. *192*, 275–281.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning (CSHL Press).

Sarkar, A., and Hochedlinger, K. (2013). The Sox Family of Transcription Factors: Versatile Regulators of Stem and Progenitor Cell Fate. Cell Stem Cell *12*, 15–30.

Saxe, J.P., Tomilin, A., Schöler, H.R., Plath, K., and Huang, J. (2009). Post-Translational Regulation of Oct4 Transcriptional Activity. PLoS ONE *4*, e4467.

Sánchez-Sánchez, A.V., Camp, E., García-España, A., Leal-Tassias, A., and Mullor, J.L. (2009). Medaka Oct4 is expressed during early embryo development, and in primordial germ cells and adult gonads. Dev. Dyn. *239*, 672–679.

Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. The EMBO Journal *9*, 2185–2195.

Sémon, M., Schubert, M., and Laudet, V. (2012). Programmed genome rearrangements: in lampreys, all cells are not equal. Curr. Biol. *22*, R641–R643.

Sharov, A.A., Dudekula, D.B., and Ko, M.S.H. (2005). A web-based tool for principal component and significance analysis of microarray data. Bioinformatics *21*, 2548–2549.

Sharov, A.A., Masui, S., Sharova, L.V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M.S. (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics *9*, 269.

Sheng, G. (2014). Epiblast morphogenesis before gastrulation. Developmental Biology.

Shi, Y., Tae Do, J., Desponts, C., Hahm, H.S., Schöler, H.R., and Ding, S. (2008). A Combined Chemical and Genetic Approach for the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell *2*, 525–528.

Shimeld, S.M., and Donoghue, P.C.J. (2012). Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development *139*, 2091–2099.

Shimeld, S.M., and Holland, P.W. (2000). Vertebrate innovations. Proceedings of the National Academy of Sciences *97*, 4449–4452.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of Pluripotency in Mouse Somatic Cells with Lineage Specifiers. Cell *153*, 963–975.

Singh, A.M., Hamazaki, T., Hankowski, K.E., and Terada, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells *25*, 2534–2542.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

Stebler, J., Spieler, D., Slanchev, K., Molyneaux, K.A., Richter, U., Cojocaru, V., Tarabykin, V., Wylie, C., Kessel, M., and Raz, E. (2004). Primordial germ cell migration in the chick and mouse embryo: the role of the chemokine SDF-1/CXCL12. Developmental Biology *272*, 351–361.

Stern, C.D., and Downs, K.M. (2012). The hypoblast (visceral endoderm): an evo-devo perspective. Development *139*, 1059–1069.

Stuart, H.T., van Oosten, A.L., Radzisheuskaya, A., Martello, G., Miller, A., Dietmann, S., Nichols, J., and Silva, J.C.R. (2014). NANOG amplifies STAT3 activation and they synergistically induce the naive pluripotent program. Curr. Biol. *24*, 340–346.

Sumi, T., Oki, S., Kitajima, K., and Meno, C. (2013). Epiblast Ground State Is Controlled by Canonical Wnt/β-Catenin Signaling in the Postimplantation Mouse Embryo and Epiblast Stem Cells. PLoS ONE *8*, e63378.

Taberlay, P.C., Kelly, T.K., Liu, C.-C., You, J.S., De Carvalho, D.D., Miranda, T.B., Zhou, X.J., Liang, G., and Jones, P.A. (2011). Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell *147*, 1283–1294.

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.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.

Takeda, H., Matsuzaki, T., Oki, T., Miyagawa, T., and Amanuma, H. (1994). A novel POU domain gene, zebrafish pou2: expression and roles of two alternatively spliced twin products in early development. Genes & Development *8*, 45–59.

Takeuchi, M., Takahashi, M., Okabe, M., and Aizawa, S. (2009). Germ layer patterning in bichir and lamprey; an insight into its evolution in vertebrates. Developmental Biology *332*, 90–102.

Tapia, N., Reinhardt, P., Duemmler, A., Wu, G., Araúzo-Bravo, M.J., Esch, D., Greber, B., Cojocaru, V., Rascon, C.A., Tazaki, A., et al. (2012). Reprogramming to pluripotency is an ancient trait of vertebrate Oct4 and Pou2 proteins. Nat Commun *3*, 1279.

Taranova, O.V., Magness, S.T., Fagan, B.M., Wu, Y., Surzenko, N., Hutton, S.R., and Pevny, L.H. (2006). SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes & Development *20*, 1187–1202.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124–1128.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Theunissen, T.W., Costa, Y., Radzisheuskaya, A., van Oosten, A.L., Lavial, F., Pain, B., Castro, L.F.C., and Silva, J.C.R. (2011). Reprogramming capacity of Nanog is functionally conserved in vertebrates and resides in a unique homeodomain. Development *138*, 4853–4865.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909–918.

Uchikawa, M., Yoshida, M., Iwafuchi-Doi, M., Matsuda, K., Ishida, Y., Takemoto, T., and Kondoh, H. (2011). B1 and B2 Sox gene expression during neural plate development in chicken and mouse embryos: Universal versus species-dependent features. Develop. Growth Differ. *53*, 761–771.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W.B., Cho, C.H.-H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339–1349.

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nature Reviews Genetics *10*, 725–732.

van den Berg, D.L.C., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2008). Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. Molecular and Cellular Biology *28*, 5986–5995.

van Rheede, T. (2005). The Platypus Is in Its Place: Nuclear Genes and Indels Confirm the Sister Group Relation of Monotremes and Therians. Mol. Biol. Evol. *23*, 587–597.

Venkatarama, T., Lai, F., Luo, X., Zhou, Y., Newman, K., and King, M.L. (2010). Repression of zygotic gene expression in the Xenopus germline. Development *137*, 651–660.

Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., Ohta, Y., Flajnik, M.F., Sutoh, Y., Kasahara, M., et al. (2014). Elephant shark genome provides unique insights into gnathostome evolution. Nature *505*, 174–179.

Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. Genes & Development *17*, 1646–1662.

Viotti, M., Nowotschin, S., and Hadjantonakis, A.-K. (2014). SOX17 links gut endoderm morphogenesis and germ layer segregation. Nature Publishing Group *16*, 1146–1156.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013a). Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. Developmental Cell *25*, 69–80.

Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., Li, C., White, S., Xiong, Z., Fang, D., et al. (2013b). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat Genet *45*, 701–706.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. Proceedings of the National Academy of Sciences *111*, 4484–4489.

Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grützner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., et al. (2008). Genome analysis of the platypus reveals unique signatures of evolution. Nature *453*, 175–183.

Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J.M., Schafer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H.J., et al. (2009). Critical Function of AP-2gamma/TCFAP2C in Mouse Embryonic Germ Cell Maintenance. Biology of Reproduction *82*, 214–223.

Wei, F., Schöler, H.R., and Atchison, M.L. (2007). Sumoylation of Oct4 Enhances Its Stability, DNA Binding, and Transactivation. J. Biol. Chem. *282*, 21551–21560.

Williams, M., Burdsal, C., Periasamy, A., Lewandoski, M., and Sutherland, A. (2012). Mouse primitive streak forms in situ by initiation of epithelial to mesenchymal transition without migration of a cell population. Dev. Dyn. *241*, 270–283.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810–813.

Woodburne, M.O., Rich, T.H., and Springer, M.S. (2003). The evolution of tribospheny and the antiquity of mammalian clades. Mol. Phylogenet. Evol. *28*, 360–385.

Wu, G., Gentile, L., Fuchikami, T., Sutter, J., Psathaki, K., Esteves, T.C., Araúzo-Bravo, M.J., Ortmeier, C., Verberk, G., Abe, K., et al. (2010). Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. Development *137*, 4159–4169.

Wu, G., Han, D., Gong, Y., Sebastiano, V., Gentile, L., Singhal, N., Adachi, K., Fischedick, G., Ortmeier, C., Sinn, M., et al. (2013). Establishment of totipotency does not depend on Oct4A. Nat Cell Biol *15*, 1089–1097.

Xu, H.M., Liao, B., Zhang, Q.J., Wang, B.B., Li, H., Zhong, X.M., Sheng, H.Z., Zhao, Y.X., Zhao, Y.M., and Jin, Y. (2004). Wwp2, an E3 ubiquitin ligase that targets transcription factor Oct-4 for ubiquitination. J. Biol. Chem. *279*, 23495–23503.

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. (2009). MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. Cell *137*, 647–658.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet *40*, 1016–1022.

Yamanoue, Y., Miya, M., Inoue, J.G., Matsuura, K., and Nishida, M. (2006). The mitochondrial genome of spotted green pufferfish Tetraodon nigroviridis (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. Genes Genet. Syst. *81*, 29–39.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881–894.

Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nat Cell Biol *13*, 762–770.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Young, R.A. (2011). Control of the Embryonic Stem Cell State. Cell *144*, 940–954.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917–1920.

Yuan, X., Wan, H., Zhao, X., Zhu, S., Zhou, Q., and Ding, S. (2011). Brief Report: Combined Chemical Treatment Enables Oct4-Induced Reprogramming from Mouse Embryonic Fibroblasts. Stem Cells *29*, 549–553.

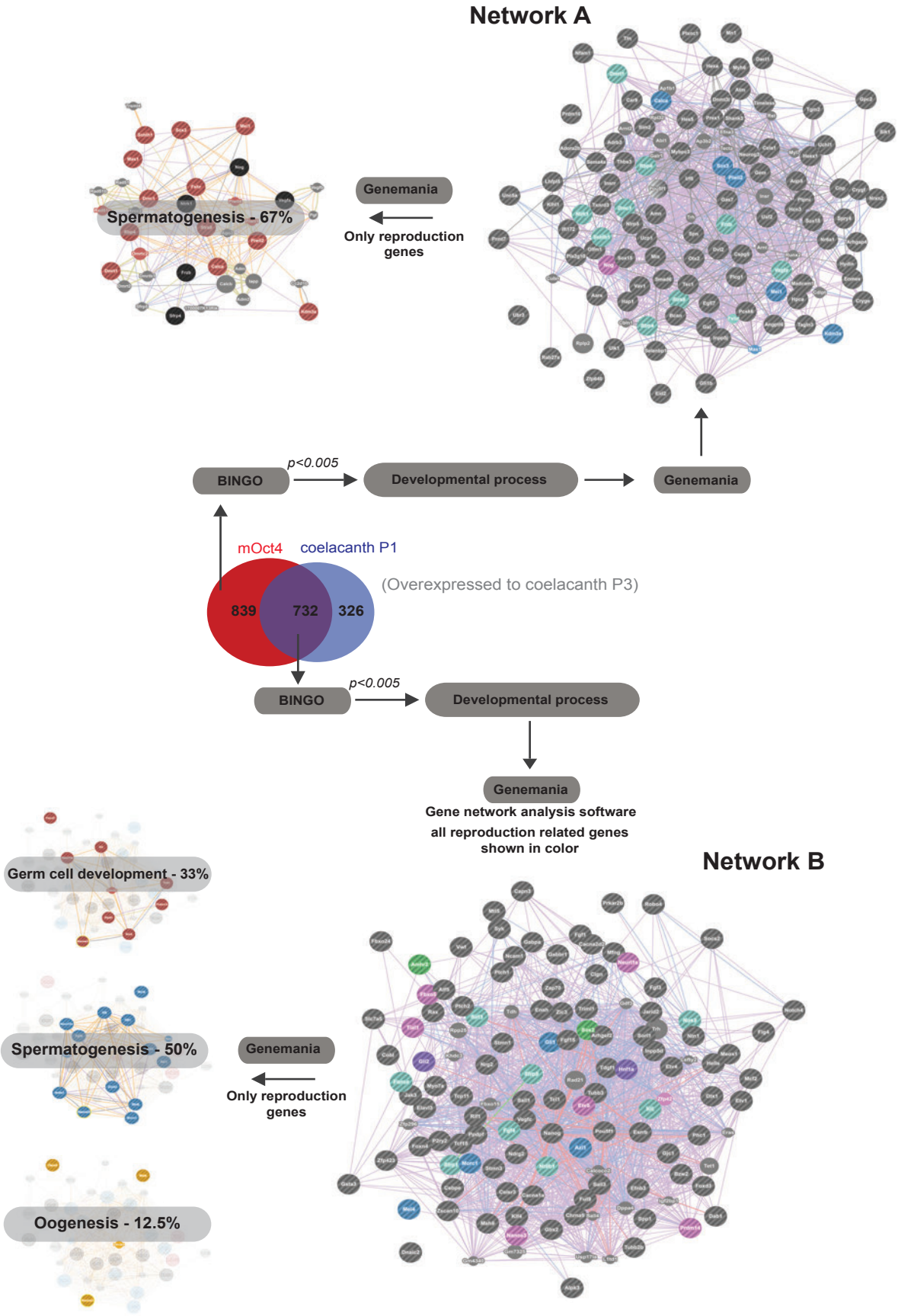
Zhang, P., Papenfuss, T.J., Wake, M.H., Qu, L., and Wake, D.B. (2008). Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. *49*, 586–597.

Zhang, P., Zhou, H., Chen, Y.-Q., Liu, Y.-F., and Qu, L.-H. (2005). Mitogenomic Perspectives on the Origin and Phylogeny of Living Amphibians. Systematic Biology *54*, 391–400.

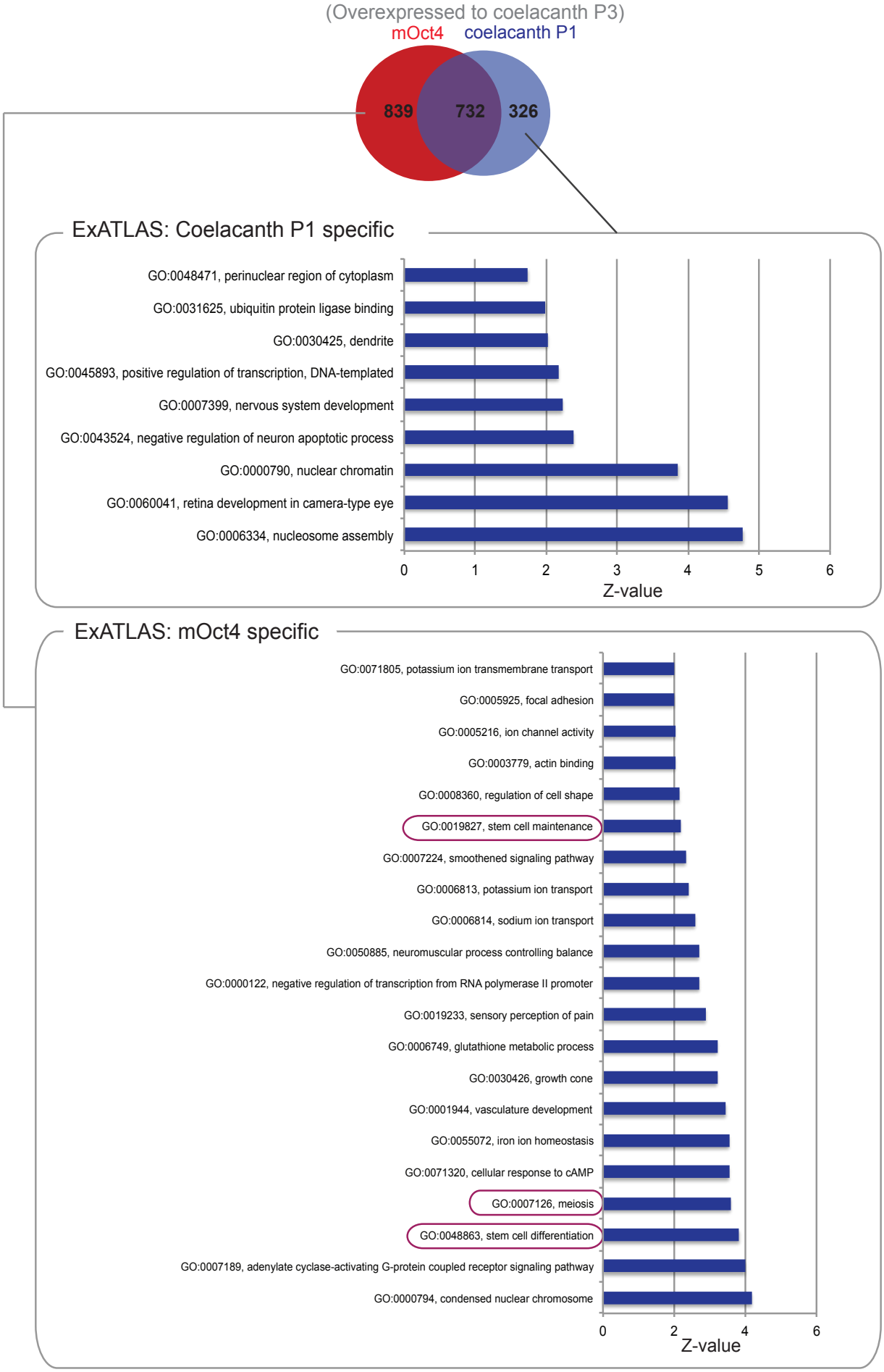
Zhao, R., Deibler, R.W., Lerou, P.H., Ballabeni, A., Heffner, G.C., Cahan, P., Unternaehrer, J.J., Kirschner, M.W., and Daley, G.Q. (2014). A nontranscriptional role for Oct4 in the regulation of mitotic entry. Proceedings of the National Academy of Sciences *111*, 15768–15773.

Zorn, A.M., and Wells, J.M. (2009). Vertebrate Endoderm Development and Organ Formation. Annu. Rev. Cell Dev. Biol. *25*, 221–251.

APPENDIX



Appendix 1 Comparative reproduction gene network between mOct4 and coelacanth POU5F1-rescued ESC lines. Overexpressed genes shared by both mOct4 and coelacanth Pou5f1 and only specific to mOct4 were analyzed by BINGO, as described in figure 3.13. From BINGO, only genes in GO: developmental process were analyzed further by Genemania, which is online software predicting gene network. Network A is developmental gene network from mOct4 specific cluster overexpressed to coelacanth POU5F3. Network B is developmental gene network shared by both mOct4 and coelacanth POU5F1 cluster overexpressed to coelacanth POU5F3. In both network A and B, genes involving in reproduction processes (e.g. gonad development, oogenesis, spernatogenesis etc) were marked in color. Then I selected only reproduction genes from each network to further analyse by Genemania. Most genes of reproduction process in network A are related to spermantogenesis while reproduction genes in network B are related to spermatogenesis and basic germ cell specification network (e.g. Prdm14, Nanos3, Zfp42).



Appendix 2 Gene annotation (GO)-term analysis of genes overexpressed in mOct4 or coelacanth POU5F1 compared to coelacanth POU5F3 Overexpressed genes in mOct4 specific cluster or coelacanth POU5F1 specific cluster were analyzed for gene ontology in web-based analysis tool ExATLAS. The parameter is FDR<0.05.

Appendix 3
GO-term analysis of 732 overexpressed genes shared by both mouse Oct4 and coelacanth POU5F1 (related to figure 3.14)
FE, fold enrichment and NG, number of genes

Title	z-value	FDR	FE	NG
GO:0006644, phospholipid metabolic process	5.5415	0	6.7964	6
Acp6,Aspg,Oc90,Pla2g1b,Pla2g2c,Pla2g5				
GO:0048863, stem cell differentiation	5.2789	0	6.3277	6
Etv4,Jarid2,Kit,Mtf2,Sox2,Zscan10				
GO:0019827, stem cell maintenance	4.8474	0	4.6164	8
Esrrb,Fgf4,Klf4,Mtf2,Rif1,Sox2,Tcl1,Tet1				
GO:0030154, cell differentiation	3.7814	0.0002	1.7935	38
Azi1,Bzw2,Camk2b,Clgn,Dab1,Efnb3,Elavl3,Enah,Fgf1,Fgf3,Fgf4,Foxn4, Gli1,Gli2,Jarid2,Kit,Klf4,M1ap,Morc1,Mtl5,Nanos3,Ndrg2,Notch4,Nr0b1, Ppdpf,Robo4,Sfrp1,Sfrp5,Sirt1,Slc7a5,Stmn1,Tcp11,Tdgf1,Tenm4, Tex40,Vegfc,Zfp423,Zic3				
GO:0006909, phagocytosis	3.753	0.0002	4.4976	5
Cebpe,Coro1a,Hck,Myo7a,Tub				
GO:0007389, pattern specification process	3.2304	0.0012	2.9128	8
Foxd3,Foxn4,Foxo6,Gli2,Mfng,Ptch1,Rax,Zic3				
GO:0005518, collagen binding	3.2171	0.0013	3.7297	5
Dpp4,Itga9,Map1a,Pcolce2,Vwf				
GO:0043565, sequence-specific DNA binding	3.0943	0.002	1.7022	31
Atf5,Cebpe,Cux2,Esrrb,Etv1,Etv4,Etv5,Foxd3,Foxn4,Foxo6,Gabpa, Gbx1,Gbx2,Gli2,Hnf1a,Jdp2,Klf2,Klf4,Meox1,Nfe2,Nkx6-3,Nr0b1,Nr5a2,Rax,Sox2,Stat4,Tfap4,Zfp42,Zfp423,Zic3,Zscan10				
GO:0006461, protein complex assembly	3.0854	0.002	3.2194	6
Capn3,Clgn,Dlgap3,Mdm4,Nupr1,Tfap4				
GO:0003682, chromatin binding	2.9991	0.0027	1.8725	21
Cdc5l,Gabpa,Gli1,Gli2,Hells,Jarid2,Jdp2,Meox1,Mkrn1,Msh6,Mybl2, Nupr1,Phc1,Pola1,Prkcb,Rcor2,Sall1,Shmt2,Sox2,Ticrr,Ube2t				
GO:0042102, positive regulation of T cell proliferation	2.9614	0.0031	3.3982	5
Cd80,Coro1a,Itgal,Jak3,Zp3				
GO:0046872, metal ion binding	2.8004	0.0051	1.2435	131
2610305D13Rik,9830147E19Rik,Acap1,Acy1,Adamts14,Adamts4,Adap1,Amhr2,Apobec3,Arhgef2, C030039L03Rik,Cabp1,Cacna1a,Cacna2d2,Capn3,Car13,Cars,Casz1,Chd5,Cpsf4l,Dgke,Dtx1,Dtx3l, Ece2,Efhb,Esco2,Esrrb,Fblim1,Fbxo5,Fgd1,Fgg,Gli1,Gli2,Gm10324,Gm13152,Gm13154,Gm13242, Gm13251,Gm14124,Gm14420,Gm5165,Hbq1a,Hpdl,Itgal,Itpk1,Kat6b,Kdm5b,Kit,Klf2,Klf4,				

Klf8,Lap3,Mast1,Mdm4,Mex3a,Mfng,Mical1,Mkrn1,Mllt6,Mmp17,Morc1,Mtf2,Nanos3, Necab2,Neil2,Neurl1a,Ninl,Nos3,Nr5a2,Oas2,Padi4,Parp1,Pck2,Pfkp,Phc1,Phf11a,Phf11d,Phf21b, Phyhd1,Pir,Pla2g1b,Pla2g2c,Pla2g5,Pola1,Ppm1n,Prdm14,Prkcb,Psph,Rasa3,Rasa4,Rasgrp2,Rnft2, Sall1,Sall3,Sap30,Sirt1,Stk38,Tcea3,Tet1,Tns1,Trim2,Trim45,Trim63,Triml1,Usp51,Xaf1, Zbtb8a,Zfp119b,Zfp13,Zfp286,Zfp42,Zfp423,Zfp428,Zfp454,Zfp459,Zfp532,Zfp534,Zfp59,Zfp607, Zfp661,Zfp689,Zfp808,Zfp811,Zfp931,Zfp936,Zfp947,Zfyve28,Zic3,Zmat4,Zranb3,Zscan10				
GO:0003690, double-stranded DNA binding	2.5595	0.0105	2.2562	9
Foxd3,Foxn4,Hnf1a,Jdp2,Klf4,Msh6,Nr0b1,Nr5a2,Pola1				
GO:0005516, calmodulin binding	2.3719	0.0177	2.0526	10
Cacna1a,Camk2b,Camkv,Itpka,Kcnn2,Myh3,Myo1f,Myo1g,Myo7a,Nos3				
GO:0035556, intracellular signal transduction	2.3344	0.0196	1.6991	18
Arhgef2,Dab1,Dgke,Inpp5d,Jak3,Kit,Mcf2l,Pdk1,Prkcb,Rasa3,Rasa4,Rasgrp2,Shc2,Socs2,Stk10, Syk,Zap70,Zp3				
GO:0043197, dendritic spine	2.3143	0.0207	2.302	7
Asic1,Dlgap3,Fbxo2,Gabbr1,Itpka,Kcnn2,Neurl1a				
GO:0007275, multicellular organismal development	2.2434	0.0249	1.3802	44
Alpk3,Azi1,Bzw2,Celsr3,Clgn,Dab1,Efnb3,Elavl3,Enah,Fgf1,Fgf3,Fgf4,Foxd3,Gli1,Hells,Itgb7, Jarid2,Meox1,Mfng,Morc1,Mtl5,Nanos3,Ndrg2,Notch4,Phc1,Plxdc1,Ppdpf,Rax,Ripply3,Robo4, Sfrp1,Sfrp5,Sirt1,Slc7a5,Sorl1,Sox2,Stmn1,Tcf15,Tcp11,Tenm4,Triml1,Vegfc,Zfp423,Zic3				
GO:0007283, spermatogenesis	2.2281	0.0259	1.711	16
Ak7,Azi1,Clgn,Ggt1,Gli1,Hmga1,Kit,M1ap,Mei4,Morc1,Mtl5,Nanos3,Nr0b1,Sirt1,Tcp11,Tex40				
GO:0007166, cell surface receptor signaling pathway	2.2275	0.0259	2.0389	9
Bai1,Cd79b,Cd97,Celsr3,Gpr133,Itgal,Ncam1,Pdk1,Syk				
GO:0003774, motor activity	2.1265	0.0335	2.4664	5
Dnaic2,Myh3,Myo1f,Myo1g,Myo7a				
GO:0042127, regulation of cell proliferation	2.1097	0.0349	1.9661	9
Atf5,Cdca7,Fanca,Foxo6,Jarid2,Klf4,Nr5a2,Ptch1,Sirt1				
GO:0005089, Rho guanyl-nucleotide exchange factor activity	2.0864	0.0369	2.4273	5
Arhgef10l,Arhgef2,Fgd1,Mcf2l,Plekhg5				
GO:0030027, lamellipodium	2.0461	0.0407	2.0989	7
Coro1a,Dpp4,Enah,Fgd1,Mcf2l,Plekhg5,Wasf3				
GO:0008360, regulation of cell shape	2.0181	0.0436	2.0785	7
Brwd1,Coro1a,Fblim1,Fgd1,Hck,Kit,Wasf3				
GO:0001843, neural tube closure	2.0006	0.0454	2.1846	6
Cecr2,Cobl,Enah,Ptch1,Sall1,Sfrp1				
GO:0045666, positive regulation of neuron differentiation	2.0006	0.0454	2.1846	6
Dab1,Etv5,Gli2,Sall1,Socs2,Sox2				
GO:0008152, metabolic process	1.9607	0.0499	1.4592	25

Acoxl,Acsl1,Acss1,Acy1,Aldh18a1,Aldh3b1,Alg13,Dgke,Dhtkd1,Ece2,Engase,Gfpt2,Glb1,Gsta3, Gsta4,Manba,Mthfd2,Nadk2,Neil2,Npl,Pfkp,Psat1,Psph,Qdpr,Zranb3

Appendix 4

GO-term analysis of 839 overexpressed genes specific in mouse Oct4, compared to coelacanth POU5F3 (related to appendix 2)

FE, fold enrichment and NG, number of genes

Title	z-value	FDR	FE	NG
GO:0000794, condensed nuclear chromosome	4.1719	0	4.5678	6
Dmc1,Dnmt3l,Hormad1,Smc1b,Sycp3,Ttn				
GO:0007189, adenylate cyclase-activating G-protein coupled receptor signaling pathway	3.9928	0.0001	4.3209	6
Adcy5,Adm2,Adora2b,Adrb3,Calca,Ptgir				
GO:0048863, stem cell differentiation	3.8248	0.0001	4.594	5
A2m,Nanog,Phf19,Pou5f1,Zscan4c				
GO:0007126, meiosis	3.5803	0.0003	3.0355	9
2410076I21Rik,Dmc1,Hormad1,Mei1,Piwil2,Rsph1,Smc1b,Stra8,Sycp3				
GO:0071320, cellular response to cAMP	3.5366	0.0004	4.1633	5
Akap7,Egr4,Gpd1,Itpr3,Rapgef3				
GO:0055072, iron ion homeostasis	3.5366	0.0004	4.1633	5
Bdh2,Hfe2,Mfi2,Sfxn4,Trf				
GO:0001944, vasculature development	3.4483	0.0006	4.0372	5
Calca,Frzb,Sfrp4,Sox18,Vegfa				
GO:0030426, growth cone	3.2032	0.0014	2.7251	9
Arhgap4,Hap1,Inpp5j,Npcd,Otx2,Prex1,Trpv4,Tsc1,Twf2				
GO:0006749, glutathione metabolic process	3.2029	0.0014	3.7007	5
Ggt6,Gstt1,Gstt2,Mgst1,Txnrd3				
GO:0019233, sensory perception of pain	2.8885	0.0039	2.7838	7
Cnr2,Ephx2,Kcnip3,Nipsnap1,Ntrk1,P2ry1,Uchl1				
GO:0000122, negative regulation of transcription from RNA polymerase II promoter	2.7061	0.0068	1.5702	33
Bcl6b,Cbx7,Cela1,Cry1,Dact1,Dmrt1,Eid2,Eomes,Gata1,Hes5,Hexim2,Kcnip3, Mlxipl,Nanog,Nog,Nr6a1,Pde2a,Phf19,Phf21a,Pou5f1,Prdm16,S100a1,Sik1,Sim2, Sox15,Sox18,Sox3,Spdef,Tagln3,Tfcp2l1,Timeless,Vegfa,Vldlr				
GO:0050885, neuromuscular process controlling balance	2.6948	0.007	2.8048	6
Aars,Adcy5,Grin2c,Hexa,Jph3,Shank3				
GO:0006814, sodium ion transport	2.5911	0.0096	2.184	10
Asic2,Asic4,Atp1a3,Atp1b2,Fxyd2,Hcn3,Slc20a2,Slc38a8,Slc5a2,Slc9b2				
GO:0006813, potassium ion transport	2.3982	0.0165	2.0655	10
Atp1a3,Atp1b2,Fxyd2,Hcn3,Kcnc4,Kcnip3,Kcnj12,Kcnj4,Kcnq2,Tsc1				

GO:0007224, smoothened signaling pathway	2.3273	0.0199	2.6645	5
Evc2,Gpc2,Hes5,Ift172,Tmem231				
GO:0019827, stem cell maintenance	2.1792	0.0293	2.5137	5
Dppa2,Eomes,Nanog,Phf19,Pou5f1				
GO:0008360, regulation of cell shape	2.1494	0.0316	2.0695	8
F2,Fgr,Gas2,Gas7,Icam1,Sema4a,Vegfa,Vil1				
GO:0003779, actin binding	2.0316	0.0422	1.5674	19
Actn3,Actr3b,Cotl1,Hdac6,Hpca,Klhl1,Mib2,Mybpc3,Myh13,Myh6,Myh7b,Myh8, Mypn,Parvb,Tln2,Trpv4,Twf2,Vil1,Wasf1				
GO:0005216, ion channel activity	2.0212	0.0433	1.7981	11
Asic2,Clcnkb,Fxyd2,Grin2c,Hcn3,Itpr3,Kcnc4,Kcnq2,Mlc1,Ryr3,Trpv4				
GO:0005925, focal adhesion	2.0041	0.0451	1.9737	8
Arhgap4,Cidec,Ephx2,Myh6,Myh8,Parvb,Tln2,Trpv4				
GO:0071805, potassium ion transmembrane transport	1.9835	0.0473	1.8882	9
Abcc8,Atp1a3,Atp1b2,Fxyd2,Hcn3,Kcnc4,Kcnip3,Kcnj12,Kcnq2				

Appendix 5
GO-term analysis of 326 overexpressed genes in coelacanth POU5F1 (related to appendix 2) FE, fold enrichment and NG, number of genes

Title	z-value	FDR	FE	NG
GO:0006334, nucleosome assembly	4.7748	0	6.3245	5
H2afx,Hist1h1a,Hist1h1b,Hist1h1t,Tspyl4				
GO:0060041, retina development in camera-type eye	4.5552	0	5.9098	5
Cacna1f,Clcn2,Lama1,Pax6,Pvrl1				
GO:0000790, nuclear chromatin	3.8624	0.0001	4.2	6
H2afx,Hist1h1a,Hist1h1b,Pax6,Pou4f1,Rara				
GO:0043524, negative regulation of neuron apoptotic process	2.3865	0.017	2.7519	5
Mdk,Ngfr,Pou4f1,Six1,Six4				
GO:0007399, nervous system development	2.2302	0.0257	2.047	9
Disc1,Dpf1,Islr2,Ngfr,Ntrk3,Pou4f1,Pura,Sema4g,Sema5b				
GO:0045893, positive regulation of transcription, DNA-templated	2.1727	0.0298	1.7887	13
Atf7ip,Bcl3,Brca2,Cd38,Egf,Mdk,Nolc1,Nos1,Pax6,Rara,Six1,Six4,Taf8				
GO:0030425, dendrite	2.0238	0.043	1.9958	8
Camk2n1,Clcn2,Cpne6,Nos1,Pura,Rara,Rbm3,Tnk2				
GO:0031625, ubiquitin protein ligase binding	1.9818	0.0475	2.3409	5
Eif4e2,Fzd5,Ngfr,Pax6,Pml				

Appendix 6

GO-term analysis of 1250 underexpressed genes shared by both mouse Oct4 and coelacanth POU5F1 or overexpressed genes specific in coelacanth POU5F3. FDR<0.05, top 100 lists, from total 339 GO-terms found. (Related to figure 3.15)

FE, fold enrichment and NG, number of genes

Title	z-value	FDR	FE	NG
GO:0001725, stress fiber	9.5306	0	5.861	21
Acta1,Amot,Anxa2,Cnn2,Daam1,Dst,Flnb,Ilk,Lima1,Myh10,Myh9,Myl12a,Myl12b,Myl9,Nox4,Palld,Pdlim7,Sorbs1,Tpm4,Vcl,Zyx				
GO:0030335, positive regulation of cell migration	8.1757	0	3.6657	32
Bcar1,Creb3,Csf1,Cxcl12,Cxcl16,Cyr61,Dab2,Edn1,Egfr,Epha1,Ets1,Fam110c,Flt1,Furin,Ilk,Irs1,Irs2,Itgav,Itgb1,Lamb1,Mmp14,Myo1c,Pdgfb,Pdpn,Podxl,Rras2,S1pr1,Smad3,Tfap2a,Tgfb2,Thbs1,Zfp703				
GO:0070062, extracellular vesicular exosome	8.024	0	4.5027	22
Acta1,Anxa1,Anxa2,Anxa5,Anxa6,Cd274,Cd47,Flna,Gsn,Hspg2,Lama4,Lamb1,Ltbp2,Msn,Mvp,Myh9,Myo1c,Myof,Rbmx,Stxbp2,Thbs1,Tln1				
GO:0005925, focal adhesion	7.8073	0	3.8375	27
Arpc2,Bcar1,Cdh1,Dag1,Dst,Fhl2,Flnb,Ilk,Irf2,Itgb1,Itgb5,Lasp1,Lima1,Lpp,Msn,Nox4,Pak1,Parva,Pdlim7,Plec,Sdc1,Sdc4,Sorbs1,Tgfb1i1,Tln1,Vcl,Zyx				
GO:0005178, integrin binding	7.8038	0	4.4771	21
Col16a1,Ctgf,Cyr61,Dab2,Egfr,Emp2,Ilk,Itga6,Itgb1,Itgb5,Itgb6,Lama5,Lamb1,Lamb2,Mmp14,Npnt,Ppap2b,Thbs1,Thy1,Timp2,Tln1				
GO:0005515, protein binding	7.6077	0	1.4083	359
Abhd5,Acta1,Acta2,Actr3,Adam15,Adamts1,Adrb2,Afap1l1,Ahr,Akt1,Alox5,Amot,Amotl1,Amotl2,Ankrd1,Ankrd37,Anxa1,Anxa2,Anxa5,App,Arhgap29,Arhgap31,Arhgef28,Arhgef5,Arhgef6,Arid3a,Arpc2,Artn,Atg10,Atn1,Atxn1,B2m,Bach2,Basp1,Baz1a,Bcar1,Bcl2l1,Bcor,Bhlhe40,Bin1,Bloc1s2,Bmp1,Btg2,Bves,Cacna2d1,Cadps2,Camk2d,Capn6,Capzb,Car4,Carhsp1,Cbx4,Cbx8,Cd24a,Cd274,Cd47,Cd81,Cdc42ep1,Cdc42ep5,Cdh1,Cdh2,Cdkn1a,Cdkn1c,Cdon,Cdx2,Cebpa,Chrdl2,Chst14,Cish,Cited1,Cldn4,Cldn6,Cldn7,Clip1,Clu,Cryab,Csf1,Csrp1,Ctsl,Cyfip2,Daam1,Dab2,Dag1,Dcxr,Dkk1,Dlgap4,Dlx2,Dnajb3,Dppa3,Dsp,Dst,Dusp22,Dynll2,Dysf,Ebf3,Edn1,Eea1,Efhd2,Efna5,Efnb1,Efnb2,Egfr,Egr2,Ehd1,Ehd2,Ehd4,Elavl1,Elmo3,Errfi1,Espn,Ets1,Eya2,F11r,Fas,Fbxl20,Fbxl7,Fbxo4,Fgfr1,Fhl2,Fkbp10,Fkbp7,Flna,Flnb,Flot1,Flrt3,Flt3,Fos,Fras1,Furin,Fxyd3,Fzd6,Gabarapl1,Gadd45b,Gadd45g,Gas1,Gata2,Gata3,Gda,Ghr,Gipc2,Gli3,Glrb,Gls,Gnas,Grip1,Gsn,H2-D1,H2-K1,Heg1,Hic1,Hipk3,Hspa2,Hspg2,Id1,Id2,Id3,Ikzf4,Il10rb,Ilk,Inadl,Iqgap1,Irak2,Irs1,Irs2,Isg15,Itga3,Itgb1,Itm2b,Itprip,Jun,Junb,Kctd10,Kiss1r,Krt18,Krt20,Krt7,Krt8,Lama5,Lamb1,Lamc2,Lasp1,Ldlrap1,Lgals1,Lgals3,Lgi3,Lhx2,Lmna,Lpp,Lrrfip1,Lrrk2,Ltbr,Mafk,Magi3,Map2k4,Mapkapk2,Mapre3,Marcks,Mast2,Mc3r,Mdfi,Mdfic,Mdm2,Meig1,Micall1,Mkl1,Mpp1,Msx2,Mxd1,Mxd4,Myh10,Myh9,Myl12a,Myo1c,Nbl1,Nefl,Nfatc2,Nkx2-2,Nkx2-				

5,Notch3,Nr3c1,Nr4a1,Nr4a2,Nrip1,Ocln,Otud7b,Pak1,Pak3,Pan3,Pard6b,Parva,Pawr,Pax3,Pbx1,Pbx3,Pcbd1,Pcdh1,Pcdh18,Pdcd6ip,Pdlim1,Pdlim7,Pdzk1,Pfn1,Pfn2,Phlda1,Pias1,Pid1,Pik3cb,Pitx2,Pja2,Pkp2,Plaur,Plec,Plin2,Pparg,Ppl,Ppp1r13l,Prkcdbp,Prmt2,Prnp,Psen1,Ptpn13,Purb,Pvr,Rab15,Rab27b,Rassf1,Rb1,Rb1cc1,Reep5,Rela,Rfk,Rgs6,Rhob,Ripk4,Rnd3,Rnf11,Rnf41,Rock2,Rras2,Runx1,Ryk,S100a10,S100a4,S100a6,S100a7a,Satb2,Scn1b,Scx,Sdc2,Sema6a,Serpinb6a,Serpinb9,Serpine1,Serpinh1,Sertad1,Sgsh,Sh3glb1,Sh3rf1,Shroom2,Six2,Slc52a3,Smad3,Socs1,Socs5,Sorbs1,Sos2,Sox6,Spg20,Sri,Ssbp2,Ssbp3,Stap2,Stk25,Sync,Syt1,Sytl1,Sytl2,Tax1bp3,Tbx4,Tdrd7,Tfap2a,Tgfb1i1,Tgfb2,Tgfbr3,Thbs1,Ticam1,Tln1,Tmem173,Tmem176b,Tmem33,Tmsb4x,Tnfaip3,Tnfrsf12a,Tnfrsf1a,Tnfrsf23,Tor2a,Tpm2,Trim21,Trp63,Tuba1a,Tuft1,Ubc,Unc5b,Vamp8,Vasn,Vcl,Wls,Wnk1,Wnt3,Wnt4,Wnt7a,Wnt9a,Wt1,Wwtr1,Yaf2,Zfp36,Zfp467,Zhx3,Zyx				
GO:0030056, hemidesmosome	7.3139	0	10.2334	6
Actr3,Dst,Itga6,Itgb1,Itgb4,Plec				
GO:0035329, hippo signaling	7.0513	0	7.6751	8
Amot,Amotl1,Amotl2,Lats2,Mob1b,Tead3,Tead4,Wwtr1				
GO:0016460, myosin II complex	6.9605	0	10.9644	5
Myh10,Myh9,Myl12a,Myl12b,Myl9				
GO:0048514, blood vessel morphogenesis	6.8608	0	5.8225	11
Ahr,Amot,Cdh2,Edn1,Efnb2,Fgfr1,Flt1,Nr2f2,Pdgfb,Shb,Thbs1				
GO:0008305, integrin complex	6.7889	0	6.1401	10
Itga11,Itga3,Itga6,Itga8,Itgav,Itgb1,Itgb4,Itgb5,Itgb6,Myh9				
GO:0005604, basement membrane	6.6524	0	3.5926	22
Adamts1,Anxa2,Ccdc80,Col2a1,Col4a2,Dag1,Fras1,Gsto1,Hspg2,Itga6,Itgb1,Itgb4,Lama4,Lama5,Lamb1,Lamb2,Lamc2,Npnt,Runx1,Tgfb2,Timp2,Timp3				
GO:0030511, positive regulation of transforming growth factor beta receptor signaling pathway	6.5409	0	6.2796	9
Cdkn1c,Cdkn2b,Dab2,Furin,Itga8,Npnt,Tgfb1i1,Tgfbr3,Thbs1				
GO:0008092, cytoskeletal protein binding	6.5345	0	4.4279	15
Anxa2,Capn2,Cryab,Dynll2,Epb4.1l3,Farp1,Farp2,Flnc,Gabarapl1,Msn,Pdlim3,Plec,Sdc1,Sdc2,Sdc4				
GO:0042383, sarcolemma	6.5053	0	3.6988	20
Adcy6,Adrb2,Alox5,Anxa1,Anxa2,Anxa5,Bsg,Car4,Ctsb,Dag1,Dst,Dysf,Flnc,Flot1,Itgb1,Krt19,Krt8,Plec,Sync,Vcl				
GO:0007155, cell adhesion	6.2359	0	2.0366	67
9430020K01Rik,Adam15,Aebp1,App,Bcar1,Bves,Cd24a,Cd47,Cdh1,Cdh2,Cdon,Cgref1,Col16a1,Ctgf,Cyfip2,Cyr61,Ddr1,Dsc2,Dscaml1,Dst,Efnb2,Epha1,F11r,F5,Farp2,Flot2,Hpse,Hspb11,Itga11,Itga3,Itga6,Itga8,Itgav,Itgb1,Itgb4,Itgb5,Itgb6,Lama4,Lama5,Lamb1,Lamb2,Lamc2,Lpp,Mpzl2,Myh10,Myh9,Ninj1,Ninj2,Npnt,Parva,Pcdh1,Pcdh18,Pdpn,Perp,Pik3cb,Podxl,Ppap2b,Psen1,Rhob,Scarf2,Thbs1,Tln1,Tnfrsf12a,Vcan,Vcl,Wisp1,Zyx				
GO:0009986, cell surface	6.1668	0	2.1028	60

Adamts7,Amot,App,Areg,Axl,Car4,Cav2,Cd24a,Cdh1,Cdon,Cryab,Ctsb,Dscaml1,Egfr,F3,Fas,Fgfbp1,Flot2,Flt3,Furin,Fut4,Fzd1,Ghr,Gpr126,H13,Heg1,Hilpda,Hspa2,Hyal2,Ifitm3,Irak2,Itga3,Itga6,Itgav,Itgb1,Itgb4,Kiss1r,Lgals1,Msn,Ocln,Pam,Pdgfb,Pdgfc,Plaur,Prom1,Psen1,Scube3,Sdc1,Sdc4,Slco3a1,Tgfb2,Tgfb3,Tgfbr3,Thbs1,Thy1,Timp2,Tnfrsf12a,Tnfrsf1a,Wnt4,Wnt7a				
GO:0003779, actin binding	6.1325	0	2.2811	48
Actr3,Arpc2,Arpc5,Cald1,Capg,Capzb,Cfl2,Cnn2,Daam1,Dag1,Dst,Dstn,Epb4.1l3,Espn,Flna,Flnb,Flnc,Gsn,Itgb1,Lasp1,Lima1,Marcks,Mkl1,Msn,Myh10,Myh9,Myo1c,Myo1d,Myo1e,Myo5b,Palld,Parva,Pawr,Pfn1,Pfn2,Plec,Pls3,S100a4,Shroom2,Ssh3,Tln1,Tmsb4x,Tpm1,Tpm2,Tpm4,Vcl,Wdr1,Wipf3				
GO:0045216, cell-cell junction organization	6.0713	0	6.1401	8
Crb3,Heg1,Marveld2,Marveld3,Ocln,Smad3,Tgfb2,Tgfb3				
GO:0031012, extracellular matrix	6.0611	0	2.583	35
Adamts1,Adamts7,Aebp1,Ccdc80,Col2a1,Col4a2,Ctsd,Dcn,Ecm1,Efemp1,F3,Hspg2,Ilk,Lama5,Lamb1,Lamb2,Lgals1,Lgals3,Ltbp2,Mmp14,Mmp23,Npnt,Pxdn,Serpine1,Serpine2,Tgfb1i1,Tgfb2,Tgfb3,Thbs1,Timp3,Tinagl1,Vcan,Wnt3,Wnt4,Wnt7a				
GO:0005913, cell-cell adherens junction	6.0253	0	4.6051	12
Cdh1,Cdh2,Dag1,Dsc2,Itga6,Lmo7,Myh9,Ndrg1,Shroom2,Sorbs1,Vcl,Zyx				
GO:0016477, cell migration	5.8504	0	2.9294	25
Atn1,Bcar1,Cd151,Cd24a,Cd47,Cdh2,Ctgf,Flt1,Fndc3b,Itgav,Itgb1,Lama5,Lamb1,Mmp14,Nck2,Nfatc2,Pax3,Podxl,Pvr,Shroom2,St14,Tgfb2,Thbs1,Tnfaip1,Wwc1				
GO:0019897, extrinsic to plasma membrane	5.8102	0	5.3135	9
Anxa2,Eea1,Ppl,Prss8,Rab13,S100a10,St14,Sytl1,Sytl2				
GO:0030054, cell junction	5.807	0	1.872	75
9430020K01Rik,Adam15,Afap1l1,Amot,Amotl1,Anxa2,Arhgap21,Arhgap31,Atn1,Bcar1,Bcl2l1,Bves,Cadps2,Cdh1,Cgn,Cldn4,Cldn6,Cldn7,Cpt1c,Crb3,Cyfip2,Dag1,Dsc2,Dsp,Dst,Epb4.1l3,Espn,Evpl,F11r,Farp1,Gap43,Glrb,Grip1,Heg1,Ilk,Inadl,Iqgap1,Iqsec1,Itgb4,Lgi3,Lpp,Magi3,Marveld2,Marveld3,Mast2,Mpp7,Myo1e,Nox4,Ocln,Pak1,Palld,Pard6b,Pard6g,Parva,Pdlim7,Perp,Pja2,Plec,Ppl,Psd3,Rab13,Rhou,S100a14,Shroom2,Sorbs1,Svop,Syt1,Tbc1d2,Tes,Tgfb1i1,Tln1,Vcl,Wdr1,Zfp185,Zyx				
GO:0042476, odontogenesis	5.5683	0	5.9695	7
Bcor,Csf1,Gas1,Lamb1,Msx2,Pitx2,Wnt10a				
GO:0005923, tight junction	5.5634	0	3.0411	21
Amot,Amotl1,Amotl2,Bves,Cgn,Cldn4,Cldn6,Cldn7,Crb3,F11r,Inadl,Magi3,Marveld2,Marveld3,Mpp7,Ocln,Pard6b,Pard6g,Pmp22,Rab13,Shroom2				
GO:0005911, cell-cell junction	5.5481	0	2.9623	22
9430020K01Rik,Ahnak,Cdh1,Cdh2,Dsp,Epb4.1l3,F11r,Heg1,Ilk,Iqgap1,Itgb1,Krt8,Magi3,Myo1e,Ocln,Pak1,Pcdh1,Pkp2,Prkcd,Rab13,Shroom2,Vcl				
GO:0003700, sequence-specific DNA binding transcription factor activity	5.545	0	1.7485	86
Ahr,Arid3a,Arntl2,Ascl2,Atf3,Bach2,Barx1,Batf3,Cdx2,Cebpa,Cers3,Cited1,Creb3,Creb3l2,Csrnp1,Dlx				

2,Egr1,Egr2,Elf1,Elk3,Esrra,Ets1,Fos,Fosb,Fosl2,Gata2,Gata3,Gli3,Hic1,Hoxa3,Hoxb4,Hoxc10,Hoxc6, Hoxc9,Hoxd8,Hoxd9,Id1,Id3,Irf2,Irf6,Irx3,Jun,Junb,Klf6,Lhx2,Maff,Mafk,Mef2d,Meis2,Mkl1,Mmp14, Msx2,Nfatc2,Nkx2-2,Nkx2-5,Nkx2-9,Nr2f2,Nr3c1,Nr4a1,Nr4a2,Pax3,Pbx1, Pbx3,Pitx2,Pparg,Purb,Rela,Runx1,Satb2,Six2,Smad3,Sox6,Tbx15,Tbx4,Tcf25,Tead3,Tead4,Tfap2a, Tfap2e,Trerf1,Trp63,Wt1,Zeb2,Zfhx3,Zfhx4,Zhx3				
GO:0042127, regulation of cell proliferation	5.4591	0	2.7411	25
Anxa1,Cd81,Cebpa,Cnn2,Creb3,Dusp22,Egfr,Esrra,Fas,Fgfr1,Gli3,Kctd11,Lama5,Mbd2,Nr3c1,Pbx1,Pias1,Pitx2,Pla2g4a,Ptgs2,Sat1,Serpine1,Tes,Tnfsf9,Wnt7a				
GO:0016324, apical plasma membrane	5.4472	0	2.3317	36
Abcb1b,Adrb2,Amotl1,Amotl2,Anxa4,Anxa6,Car4,Cd81,Cdh2,Cldn4,Crb3,Ctsb,Dab2,Egfr,Emp2,Hspa1a,Hyal2,Lmo7,Msn,Nox4,Ocln,Pdpn,Pdzk1,Plec,Podxl,Prom1,Psen1,Rab27b,S100g,Shroom2,Slc46a1, Slco3a1,Stxbp2,Tcirg1,Thy1,Trpm6				
GO:0014704, intercalated disc	5.3839	0	4.2213	11
Anxa5,Camk2d,Capzb,Cdh2,Dsp,Dst,Itgb1,Pak1,Pkp2,Scn1b,Vcl				
GO:0051496, positive regulation of stress fiber assembly	5.3547	0	4.7638	9
Ctgf,Epha1,Kiss1r,Nox4,Pak1,Pfn1,Sdc4,Smad3,Wnt4				
GO:0051216, cartilage development	5.2766	0	3.319	16
Bmp1,Bmp8a,Bmp8b,Chrdl2,Col2a1,Creb3l2,Dlx2,Edn1,Esrra,Gnas,Hoxa3,Msx2,Satb2,Scx,Sox6, Wnt7a				
GO:0001726, ruffle	5.2745	0	3.4366	15
Amot,Anxa2,Bcar1,Eps8l2,Gsn,Myh9,Pak1,Palld,Pdlim7,Pdpn,Podxl,S100a11,S100a6,Tln1,Tnfrsf12a				
GO:0015629, actin cytoskeleton	5.2743	0	2.6085	26
Acta1,Acta2,Cdc42ep1,Cdh1,Cfl2,Csrp1,Cttnbp2nl,Dennd2a,Dst,Dstn,Espn,Fam101b,Fhl2,Flna,Flnc,Gsn,Lima1,Myh9,Nfatc2,Nxf7,Parva,Pdlim3,Pdlim7,Tln1,Vcl,Wdr1				
GO:0031668, cellular response to extracellular stimulus	5.2572	0	6.1401	6
Axl,Cdkn1a,Cdkn2b,Fos,Itga6,Nr4a2				
GO:0031362, anchored to external side of plasma membrane	5.2572	0	6.1401	6
Car4,Cd24a,Efna5,Hyal2,Thy1,Tnfrsf23				
GO:0000904, cell morphogenesis involved in differentiation	5.2347	0	6.9773	5
Dab2,Dkk1,Krt8,Lamb2,Myh9				
GO:0000982, RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity	5.2347	0	6.9773	5
Arntl2,Ddn,Egr1,Msx2,Tfap2a				
GO:0005544, calcium-dependent phospholipid binding	5.2166	0	4.6051	9
Anxa1,Anxa2,Anxa3,Anxa4,Anxa5,Anxa6,Dysf,Pla2g4a,Syt1				
GO:0016337, cell-cell adhesion	5.1968	0	3.07	18
Cdh1,Cdh2,Cdon,Cyfip2,Cyr61,Dsp,Egfr,Itga6,Lmo7,Mpzl2,Myh9,Pdpn,Pkp2,Ppap2b,Psen1,Pvr,Thy1,				

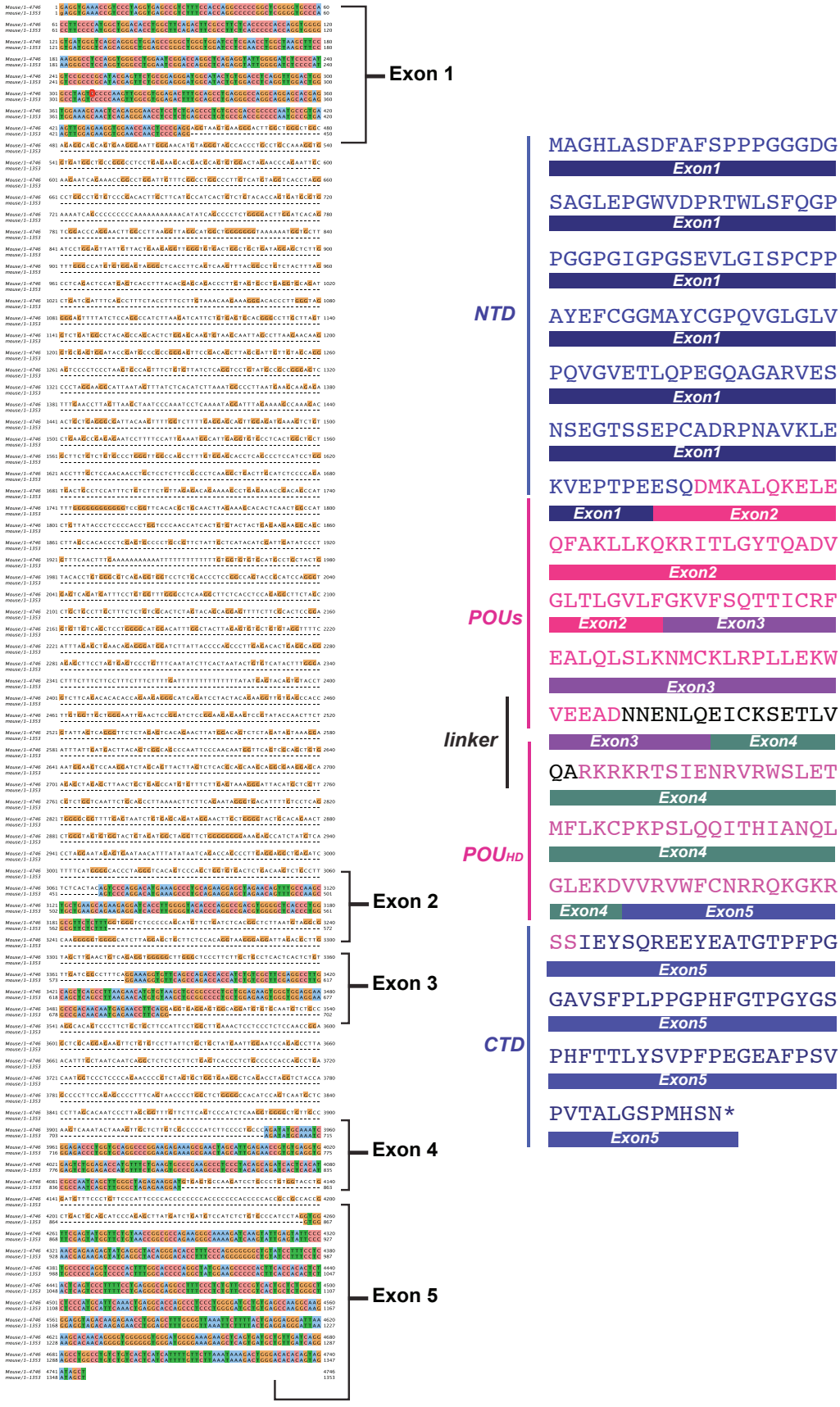
Vnn1				
GO:0043565, sequence-specific DNA binding	5.1813	0	1.8189	66
Ahr,Ascl2,Atf3,Bach2,Barx1,Batf3,Cdx2,Cebpa,Cers3,Creb3,Creb3l2,Dlx2,Egr1,Elf1,Elk3,Esrra,Esx1,Ets1,Fos,Fosb,Fosl2,Gata2,Gata3,Gli3,Hic1,Hoxa3,Hoxb4,Hoxc10,Hoxc6,Hoxc9,Hoxd8,Hoxd9,Irx3,Jun,Junb,Lhx2,Maff,Mafk,Meis2,Msx2,Nfatc2,Nkx2-2,Nkx2-5,Nkx2-9,Nr2f2,Nr3c1,Nr4a1,Nr4a2,Pax3,Pbx1,Pbx3,Pitx2,Pparg,Rela,Satb2,Scx,Six2,Smad3,Sox6,Tfap2a,Trp63,Wt1,Zeb2,Zfhx3,Zfhx4,Zhx3				
GO:0007275, multicellular organismal development	5.133	0	1.6059	102
Akt1,Apaf1,Ascl2,Bin1,Bmp1,Bmp8a,Bmp8b,Bves,Cdx2,Chrdl2,Cited1,Dab2,Dkk1,Dlx2,Dppa3,Dpysl2,Ebf3,Ebf4,Ecscr,Efna5,Efnb1,Efnb2,Egfr,Eya2,Eya4,Fam3c,Farp1,Flt1,Fzd1,Fzd6,Gadd45b,Gadd45g,Gap43,Heg1,Hic1,Hoxa3,Hoxb4,Hoxc10,Hoxc6,Hoxc9,Hoxd8,Hoxd9,Id1,Id2,Id3,Irx3,Itga8,Itgav,Itgb1,Itgb4,Itgb5,Itgb6,Kctd11,Krt8,Mbnl3,Mdfi,Mef2d,Meis2,Msx2,Nkx2-2,Nkx2-5,Nkx2-9,Nnat,Notch3,Npnt,Pak3,Paqr7,Pax3,Pbx1,Pdgfb,Pdgfc,Pdlim7,Pdpn,Pitx2,Plac1,Plxdc2,Ppap2b,Prtg,Rhob,Satb2,Scx,Sema3c,Sema3e,Sema6a,Serpine2,Shb,Shisa3,Shroom2,Six2,Sox6,Tbx4,Tdrd7,Tnfrsf12a,Trp63,Unc5b,Wipf3,Wls,Wnt10a,Wnt3,Wnt4,Wnt7a,Wnt9a				
GO:0048598, embryonic morphogenesis	5.0244	0	5.7563	6
Bmp8b,Cdon,Flt1,Gas1,Gli3,Zeb2				
GO:0005615, extracellular space	5.0107	0	1.6338	91
Adm,Aebp1,Anxa1,Anxa5,Areg,Artn,Axl,B2m,Bmp1,Bmp8a,Bmp8b,Cd109,Cmtm8,Col2a1,Csf1,Ctgf,Ctsb,Ctsd,Ctsl,Cxcl12,Cxcl16,Dag1,Dcn,Defa3,Dkk1,Dmkn,Ecm1,Edn1,Efemp1,F3,F5,Flrt3,Furin,Gcnt1,Gdnf,Ghr,Gm13275,Gpx3,Gsn,Hilpda,Hspg2,Hyal1,Igfbp3,Igfbp6,Itm2b,Lama5,Lamb1,Lamc2,Lgals1,Ltbp2,Nppb,Npy,Pam,Pdgfb,Pdgfc,Podn,Prom1,Prss8,Pxdn,Qsox1,Ren1,Ren2,S100a7a,Sema3c,Serpinb6a,Serpinb9,Serpine1,Serpine2,Serpinh1,Smpd1,Smpdl3a,Soga1,St14,Tacstd2,Tcn2,Tgfb2,Tgfb3,Tgfbr3,Thbs1,Timp2,Tnfrsf1a,Tnfsf15,Tnfsf9,Vcan,Vgf,Vnn1,Wnt10a,Wnt3,Wnt4,Wnt7a,Wnt9a				
GO:0030027, lamellipodium	4.9709	0	2.8593	19
Actr3,Amot,Amotl1,Arhgap31,Arhgef6,Arpc5,Bcar1,Capzb,Cdh2,Dag1,Dysf,Gsn,Ilk,Parva,Pdpn,Podxl,Ptpn13,Rab13,Stxbp2				
GO:0017048, Rho GTPase binding	4.9578	0	4.3172	9
Arhgef16,Daam1,Flna,Iqgap1,Lrrk2,Pak3,Pfn1,Rock2,Vcl				
GO:0071285, cellular response to lithium ion	4.9358	0	6.3959	5
Cdh1,Cebpa,Fas,Id2,Pparg				
GO:0050860, negative regulation of T cell receptor signaling pathway	4.9358	0	6.3959	5
Dusp3,Elf1,Pawr,Prnp,Thy1				
GO:0045893, positive regulation of transcription, DNA-templated	4.825	0	1.7869	61
Afap1l2,Ahr,Bloc1s2,Cdh1,Cdkn1c,Cdx2,Cebpa,Cited1,Cited4,Creb3,Creb3l2,Dab2,Ebf3,Egr1,Egr2,Elf1,Esrra,Ets1,Fos,Fzd1,Gata3,Gli3,Id2,Ilk,Irf6,Jun,Lhx2,Mapre3,Mdfic,Mef2d,Mkl1,Nfatc2,Nkx2-				

5,Nr2f2,Nr4a1,Nr4a2,Pax3,Pcbd1,Pdgfb,Pias1,Pitx2,Pparg,Prmt2,Rbmx,Rela,Runx1,Scx,Sertad1,Smad3,Sox6,Ssbp3,Tfap2a,Tgfb1i1,Tgfb3,Tmsb4x,Trerf1,Trp63,Wnt4,Wnt7a,Wt1,Yaf2				
GO:0043236, laminin binding	4.8104	0	5.4177	6
Ecm1,Itga6,Itgb1,Lgals1,Thbs1,Tinagl1				
GO:0043034, costamere	4.8104	0	5.4177	6
Dag1,Ilk,Krt19,Krt8,Sdc4,Vcl				
GO:0035914, skeletal muscle cell differentiation	4.7642	0	3.4755	12
Ankrd1,Atf3,Btg2,Egr1,Egr2,Fos,Maff,Mef2d,Nr4a1,Pax3,Rb1,Scx				
GO:0001656, metanephros development	4.7329	0	4.3858	8
Gdnf,Gli3,Id2,Itga8,Rdh10,Six2,Wnt4,Wt1				
GO:0051402, neuron apoptotic process	4.7194	0	4.0633	9
Apaf1,App,Atn1,Bcl2l1,Bok,Fas,Psen1,Rb1,Trp63				
GO:0006915, apoptotic process	4.7127	0	1.7634	61
Akt1,Apaf1,App,Bcl2l1,Bnip2,Bok,Cited1,Csrnp1,Ctsc,Cyfip2,Dab2,Dap,Elmo3,Epb4.1l3,Fas,Gadd45b,Gadd45g,Gsn,Gulp1,Hipk3,Krt20,Lgals1,Ltbr,Map2k4,Mef2d,Nek6,Nr4a1,Pak1,Pawr,Pdcd6ip,Pea15a,Peg3,Perp,Phlda1,Phlda3,Plscr3,Ppp1r13l,Prkcd,Psen1,Purb,Rassf7,Rfk,Rhob,Rnf41,S100a14,Sema6a,Sgms1,Sh3glb1,Shb,Shisa5,Stk25,Ticam1,Tmbim4,Tmem173,Tnfaip3,Tnfrsf10b,Tnfrsf12a,Tnfrsf1a,Tnfsf15,Trp63,Unc5b				
GO:0048471, perinuclear region of cytoplasm	4.6916	0	1.8551	51
Anxa2,Anxa6,App,Atn1,Camk2d,Capn6,Car4,Cav2,Ccng1,Cdh1,Clic1,Clu,Csf1,Ctgf,Ctsb,Cyfip2,Dab2,Dst,Egfr,Ehd1,Ehd2,Ehd4,Fas,Flna,Galnt3,Gsn,Hspa1a,Hyal2,Inadl,Itm2c,Kalrn,Lamb1,Lamc2,Lmna,Mapre3,Mvp,Myo5b,Ndrg1,Nox4,Nppb,Pam,Pla2g4a,Plec,Prdx5,Psen1,Ptges,S100a14,S100a4,S100a6,Tmem173,Wwc1				
GO:0048704, embryonic skeletal system morphogenesis	4.6841	0	3.4111	12
Dscaml1,Gli3,Hoxa3,Hoxb4,Hoxc9,Hoxd9,Hspg2,Mdfi,Satb2,Six2,Tbx15,Wnt9a				
GO:0051146, striated muscle cell differentiation	4.6691	0	5.9039	5
Akt1,Bnip2,Cdh2,Cdon,Rb1				
GO:0017017, MAP kinase tyrosine/serine/threonine phosphatase activity	4.6691	0	5.9039	5
Dusp1,Dusp10,Dusp14,Dusp5,Dusp8				
GO:0046697, decidualization	4.6691	0	5.9039	5
Ctsb,Ctsl,Junb,Pla2g4a,Ptgs2				
GO:0005578, proteinaceous extracellular matrix	4.5716	0	2.0664	35
Adamts1,Adamts7,Anxa2,Bmp1,Ccdc80,Col16a1,Col2a1,Col4a2,Cpz,Ctgf,Dcn,Ecm1,Efemp1,Efemp2,Gpc4,Hpse,Hspg2,Lama4,Lama5,Lamb1,Lamb2,Lamc2,Lgals1,Lgals3,Npnt,Podn,Pxdn,Tgfbr3,Timp3,Vcan,Wnt10a,Wnt3,Wnt4,Wnt7a,Wnt9a				
GO:0045121, membrane raft	4.5713	0	2.2772	27

Adcy6,Arid3a,Bsg,Cav2,Cd24a,Dag1,Efnb1,Egfr,Fas,Flot1,Flot2,Furin,Grip1,Hpse,Hyal2,Itgb1,Lrrk2, Mal,Myo1c,Pdzk1,Prnp,Prss8,Psen1,Sdc4,Sorbs1,Thy1,Tnfrsf1a				
GO:0005509, calcium ion binding	4.5451	0	1.721	62
Anxa1,Anxa2,Anxa3,Anxa4,Anxa5,Anxa6,Bmp1,Capn2,Cdh1,Cdh2,Cgref1,Dag1,Dgka,Dlk1,Dlk2,Dsc 2,Dst,Efemp1,Efemp2,Efhd2,Ehd1,Ehd2,Ehd4,Fkbp10,Fkbp7,Fkbp9,Galnt3,Gpd2,Heg1,Itpr2,Ltbp2,Mi cu1,Mmp14,Myl12b,Myl6,Myl9,Notch3,Npnt,Nucb1,Pam,Pcdh18,Pla2g4a,Plcl2,Pls3,Prrg3,Prrg4,Rcn2, Runx1,S100a10,S100a11,S100a14,S100a4,S100a6,S100a7a,S100g,Scube3,Sgce,Slc25a24,Sri,Syt1, Thbs1,Vcan				
GO:0010718, positive regulation of epithelial to mesenchymal transition	4.4996	0	4.4771	7
Dab2,Smad3,Tgfb1i1,Tgfb2,Tgfb3,Wwtr1,Zfp703				
GO:0019838, growth factor binding	4.4987	0	3.8375	9
Cyr61,Fgfbp1,Fgfbp3,Flt1,Ghr,Igfbp3,Igfbp6,Ltbp2,Osmr				
GO:0031175, neuron projection development	4.4728	0	2.6568	18
App,Areg,Bloc1s2,Btg2,Capzb,Cd24a,Ehd1,Fgfr1,Gdnf,Itgb1,Lamb1,Lamb2,Micall1,Myh10, Palld,Rab13,Rb1,Ryk				
GO:0045669, positive regulation of osteoblast differentiation	4.4545	0	3.2316	12
Cebpa,Clic1,Cyr61,Gli3,Gnas,Ilk,Msx2,Npnt,Pdlim7,Trp63,Wnt4,Zhx3				
GO:0002020, protease binding	4.4442	0	2.8782	15
Cast,Cstb,Ecm1,F3,Furin,Hspa1a,Itgb1,Mbp,Serpinb6a,Serpinb9,Serpine1,Sri,Tnfaip3,Tnfrsf1a,Ubc				
GO:0001968, fibronectin binding	4.4293	0	4.8474	6
Ccdc80,Ctgf,Igfbp3,Itgb1,Sdc4,Thbs1				
GO:0030111, regulation of Wnt signaling pathway	4.4288	0	5.4822	5
Dkk1,Mdfi,Mdfic,Ppap2b,Tax1bp3				
GO:0005160, transforming growth factor beta receptor binding	4.4288	0	5.4822	5
Gdnf,Smad3,Tgfb2,Tgfb3,Tgfbr3				
GO:0043410, positive regulation of MAPK cascade	4.3874	0	2.8426	15
Adrb2,Bnip2,Cdh2,Cdon,Fgfr1,Flt1,Igfbp3,Ilk,Itgav,Itgb1,Pdgfb,Prkcd,Ryk,Timp2,Wwc1				
GO:0006469, negative regulation of protein kinase activity	4.3813	0	3.1759	12
Akt1,Dbndd2,Epha1,Gadd45b,Gadd45g,Ilk,Itprip,Psen1,Rb1,Sh3bp5,Thy1,Wwtr1				
GO:0043406, positive regulation of MAP kinase activity	4.3629	0	3.4887	10
Cd24a,Edn1,Egfr,Fgfr1,Flt1,Ilk,Nox4,Pdgfb,Prkcd,Psen1				
GO:0030018, Z disc	4.2889	0	2.6989	16
Anxa5,Capzb,Cryab,Dst,Fhl2,Krt19,Krt8,Myl12a,Myl12b,Myl9,Pak1,Pdlim3,Psen1,Sri,Sync,Vcl				
GO:0016459, myosin complex	4.2836	0	3.2472	11
Cgn,Dynll2,Myh10,Myh9,Myl12b,Myl6,Myl9,Myo1c,Myo1d,Myo1e,Myo5b				

GO:0007050, cell cycle arrest	4.2764	0	2.7741	15
Cdkn1a,Cdkn1c,Cdkn2b,Cgref1,Cgrrf1,Dst,Gas1,Ilk,Irf6,Pmp22,Rassf1,Rb1,Smad3,Tgfb2,Thbs1				
GO:0005102, receptor binding	4.2488	0	1.8835	40
Acot2,Adcy6,App,Artn,Edn1,Efnb2,Egfr,Fzd1,Gstk1,H2-D1,H2-K1,H2-Q6,H2-Q7,H2-T10,H2-T9,Hilpda,Itgav,Itgb1,Itgb5,Lama4,Lama5,Ldlrap1,Msn,Myo1c,Nrip1,Nudt7,Nxph3,Pfn1,Prdx5,Ren1,Serpine1,Serpine2,Slc39a1,Tgfb2,Tnfsf9,Wnt10a,Wnt3,Wnt4,Wnt7a,Wnt9a				
GO:0005198, structural molecule activity	4.2413	0	2.329	22
Anxa1,Bves,Cldn4,Cldn6,Cldn7,Dsp,Epb4.1l3,Evpl,Krt18,Krt19,Krt20,Krt42,Krt7,Krt8,Krt81,Krt85,Lmna,Nefl,Ocln,Sync,Tln1,Vcl				
GO:0009925, basal plasma membrane	4.2409	0	3.8375	8
Cldn4,Itga6,Itgb4,Ldlrap1,Myo1c,Plec,Slco3a1,Tacstd2				
GO:0034446, substrate adhesion-dependent cell spreading	4.2409	0	3.8375	8
Axl,Bves,Epha1,Fndc3b,Ilk,Lama5,Lamb1,Parva				
GO:0007411, axon guidance	4.2336	0	2.4716	19
Artn,Efna5,Efnb1,Flot1,Flrt3,Gap43,Gas1,Gata3,Gli3,Lamb2,Lhx2,Myh10,Ryk,Scn1b,Sema3c,Sema6a,Tgfb2,Unc5b,Wnt3				
GO:0004190, aspartic-type endopeptidase activity	4.2197	0	4.1327	7
Bace1,Ctsd,H13,Nrip3,Psen1,Ren1,Ren2				
GO:0030667, secretory granule membrane	4.2105	0	5.1167	5
Car4,Msn,Pam,Rab27b,Vgf				
GO:0006693, prostaglandin metabolic process	4.2105	0	5.1167	5
Hpgd,Pdpn,Ptges,Ptgs2,Tnfrsf1a				
GO:0071347, cellular response to interleukin-1	4.2105	0	5.1167	5
Ankrd1,Hyal1,Hyal2,Rbmx,Rela				
GO:0030036, actin cytoskeleton organization	4.1913	0	2.3985	20
Amot,Arhgef5,Bcar1,Capzb,Csrp1,Daam1,Fam101b,Flna,Iqsec1,Myh10,Pak1,Pak3,Parva,Pdgfb,Pdlim7,Pfn1,Pfn2,Rhou,Rock2,Tmsb4x				
GO:0007179, transforming growth factor beta receptor signaling pathway	4.171	0	3.0197	12
Bmp8a,Bmp8b,Cited1,Dusp22,Fos,Gdnf,Hpgd,Jun,Smad3,Tgfb2,Tgfb3,Tgfbr3				
GO:0043234, protein complex	4.1375	0	1.7364	50
Acta1,Anxa1,Anxa2,Anxa6,Camk2d,Cav2,Cdh2,Cdx2,Cebpa,Crb3,Dpysl2,Flna,Flt3,Gata2,Gchfr,Gsn,Hspa1a,Id2,Ilk,Inadl,Iqgap1,Mdm2,Myh9,Myl12a,Myl12b,Myo1e,Myo5b,Nr3c1,Pak1,Pard6b,Pard6g,Parva,Psen1,Ptgs2,Rela,Sdc1,Serpinb6a,Sh3glb1,Smad3,Ssbp3,Stxbp2,Tes,Tnfrsf1a,Tpm1,Trp63,Tuba1a,Tuba8,Tubb6,Vcl,Wwc1				
GO:0030173, integral to Golgi membrane	4.1303	0	3.7213	8
Acer3,Qsox1,Sgms1,Sgms2,St3gal1,St3gal5,St3gal6,St8sia4				
GO:0008285, negative regulation of cell proliferation	4.1274	0	1.865	39

Adm,Btg2,Cav2,Cdkn1a,Cdkn2b,Cebpa,Ddr1,Flt3,Gata3,Gli3,Hyal1,Ifitm3,Igfbp3,Irf6,Itgb1,Jun,Kiss1r,Msx2,Nox4,Pkp2,Pmp22,Podn,Pparg,Ptges,Ptgs2,Ptpn14,Rb1,Rnf41,Serpine2,Smad3,Tax1bp3,Tes,Tfap2a,Tgfb2,Tgfb3,Timp2,Tmem127,Wnt9a,Wt1				
GO:0008284, positive regulation of cell proliferation	4.1199	0	1.7849	45
Acer3,Adm,Adrb2,Areg,Atf3,Bcl2l1,Bloc1s2,Cd47,Cd81,Cdx2,Clu,Csf1,Ctgf,Cxcl12,Edn1,Egfr,Epha1,Ets1,Fgfbp1,Fgfr1,Gas1,Hilpda,Hoxa3,Id2,Ilk,Irs2,Itgav,Itgb1,Jun,Nkx2-5,Osmr,Pax3,Pbx1,Pdgfb,Pdgfc,Pla2g4a,Ptgs2,Rela,S1pr1,Scx,Ssbp3,Tgfb2,Wnt7a,Wwtr1,Zfp703				
GO:0016328, lateral plasma membrane	4.1011	0	3.4538	9
Bves,Cdh1,Cldn4,Cldn7,Iqgap1,Myh9,Myo1c,Rab13,Tacstd2				
GO:0043408, regulation of MAPK cascade	4.098	0	4.3858	6
Bmp8b,Cd24a,Id1,Ren1,Rnf41,Timp2				
GO:0005912, adherens junction	4.0901	0	3.9797	7
Cdh1,Cdh2,Itgb1,Mpp7,Myo1e,Pkp2,Vcl				
GO:0010951, negative regulation of endopeptidase activity	4.0412	0.0001	2.5584	16
Akt1,App,Cast,Cstb,Furin,Lxn,Serpinb6a,Serpinb6c,Serpinb9,Serpinb9e,Serpinb9f,Serpinb9g,Serpine1,Serpine2,Serpinh1,Timp2				
GO:0045944, positive regulation of transcription from RNA polymerase II promoter	4.0411	0.0001	1.541	77
Adrb2,Akt1,Ankrd1,App,Arid3a,Arntl2,Atxn1,Bloc1s2,Btg2,Cdon,Cebpa,Creb3,Csrnp1,Cyr61,Ddn,Dlx2,Egfr,Egr1,Egr2,Elf1,Esrra,Ets1,Fgfr1,Fos,Gata2,Gata3,Gdnf,Gli3,Hmgn3,Hoxb4,Hoxc10,Hoxd9,Hyal2,Ikzf2,Itga6,Jun,Klf6,Lhx2,Mef2d,Meis2,Mkl1,Nfatc2,Nkx2-2,Nkx2-5,Nr3c1,Nr4a1,Nr4a2,Nrip1,Pax3,Pbx1,Pfn1,Pid1,Pitx2,Pparg,Rb1,Rbmx,Rela,Runx1,S1pr1,Satb2,Scx,Six2,Smad3,Sox6,Ssbp3,Tead3,Tead4,Tfap2a,Tfap2e,Tgfb3,Tmem173,Tnfrsf1a,Trp63,Wnt7a,Wt1,Wwtr1,Zfp750				
GO:0043066, negative regulation of apoptotic process	4.0305	0.0001	1.7521	46
Akt1,Axl,Bcl2l1,Btg2,Ccng1,Cdkn1a,Clu,Cryab,Cyr61,Dab2,Dusp1,Egfr,Egr1,Egr2,Ercc5,Fas,Fhl2,Gas1,Gli3,Hipk3,Id1,Ilk,Jun,Krt18,Mdm2,Msx2,Nkx2-5,Plaur,Plk2,Prdx5,Prnp,Psen1,Rb1cc1,Rela,Scx,Serpinb9,Sh3rf1,Smad3,Tfap2a,Tgfb2,Thbs1,Tmbim4,Tnfrsf1a,Trp63,Wnt7a,Wt1				



Appendix 7 Alignment of mouse Pou5f1/Oct4 mRNA with genomic region containing Pou5f1 gene. The alignment show five extron structure. The corresponded exon is plotted against the POUV protein domain (on the right). The letters are amino acid, the color of each amino acid is corresponded to the POUV protein region (NTD, N-terminal domain; POUs, POU specific domain;linker; POUhd, POU homeodomain; CTD, C-terminal domain) Underlying the amino acid sequences are the exon structures as shown on the left. This figure is used with figure 3.19 and 3.20 to understand the extron-intron structure/protein coding region of lamprey POU5F1/3