MOLECULAR EVOLUTION OF THE MAMMALIAN EPIBLAST

LIM LENG HIONG
(BSc (Hon), University of Alberta)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

I would like to thank my advisor Dr. Paul Robson for his guidance during my PhD programme, fellow PhD student Luo Wenlong and postdoc Dr. Andrew Hutchins for their advice and encouragement through these years, and other members of the Robson Group, especially research assistant Woon Chow Thai, who have provided help and materials.

Specifically, I thank Woon Chow Thai for performing the Illumina BeadArray experiment and instructing me in using GeneSpring for expression analysis, Dr. Andrew Hutchins and Dr. Chu Lee Thean for developing the Excel template to analyze BioMark Realtime PCR data, Tahira Bee Allapitchay for adapting Yamanaka's iPS protocol and instructing me on the experimental technique and virus safety procedures, and finally Dr. Eric Lam Chen Sok for providing me with the Sox2-EGFP knock-in mice.

I am also very grateful to my parents Lim Beng Cheng and Ong Chong Mooi, as well as my siblings Lim Hwee San, Lim Hwee Leng and Lim Leng Joon for their constant support and understanding.

Publication List

Rodda D.J., Chew J.L., Lim L.H., Loh Y.H, Wang B., Ng H.H. and Robson P. (2005) Transcriptional Regulation of Nanog by OCT4 and SOX2. *J Biol Chem* **280** : 24731-24737

Table of contents

Summary

The mammalian pluripotent cell is a transitory cell type that lasts for only a day during *in vivo* development, but can be cultured *in vitro* to form embryonic stem (ES) cells which exhibit long-term self-renewal. This unique potential may have evolved in early mammals and is likely to have co-evolved with the process of placental formation. My thesis work focused on identifying the origins of this cell type at the molecular level.

Mutations that alter developmental genetic regulatory networks are thought to be an important mechanism in evolution, thus I have focused my studies primarily on a single transcription factor essential to the pluripotent cell regulatory network, namely Oct4. From screening genomic BAC libraries and database searches, I have uncovered new sequence information pertaining to Oct4, which is encoded by the *Pou5f1* gene.

Notably, I identified a *Pou5f1* homolog in platypus that is syntenic to eutherian *Pou5f1*. Additional sequence information from non-mammal vertebrates indicates that the origin of the genomic location of mammalian *Pou5f1* predates the base of mammalian evolution, and thus the presence of the gene itself is not a eutherian-specific change. However, from a more detailed sequence analysis I found 12 amino acid positions within the Oct4 DNA binding domain (DBD) to be completely conserved within all eutherians but differing in platypus, opossum, and kangaroo. Experiments focused on identifying eutherian-specific gene regulation mediated through the Oct4 DBD have been done. Oct4 DBDs of mouse, human, elephant and platypus have been fused with a strong repressor (EnR) and a strong activator (VP16) of transcription and these transfected into ES cells to study alterations in

gene expression. In addition, full-length Oct4 chimeras containing the DBDs of mouse, elephant and platypus have been constructed and tested for their ability to induce pluripotency using the induced pluripotent stem cell (iPS) experimental system.

In sum, I show that there are only subtle cell-level phenotypic differences between eutherian and platypus Oct4 DBDs, strongly suggesting that the pluripotent capability of Oct4 already exists prior to the appearance of eutherian mammals. Current results point towards the possibility that the eutherian-specific functions of the Oct4 protein did not arise from the emergence of a newly evolved ability to induce or maintain pluripotency, but may have occurred due to changes in its pre-existing pluripotent capability.

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Historical Background

When Charles Darwin first published *On the Origins of Species* in 1859, he proposed that species were not fixed, but gradually evolve over geological timescales via the process of natural selection, thus establishing the foundation for evolutionary biology. However, right at the beginning there were two significant weaknesses in his theory of evolution (Wilkins 2002).

One of them was the lack of a detailed mechanism for inheritance, which would later be addressed in the early 1900s when Gregor Mendel's work on pea plants was rediscovered. Also missing was the precise relationship between embryonic development and the development of morphological differences which result in the diversification of species, an area of investigation that remains hotly debated today.

From the beginning, Darwin was already aware of the importance of embryological data to the development of evolutionary theory, although he had very limited evidence available to him at that time (Darwin 1859).

In Chapter 13 of the first edition, he concluded that: "Thus, as it seems to me, the leading facts in embryology, which are second in importance to none in natural history, are explained on the principle of slight modifications not appearing, in the many descendants from some one ancient progenitor, at a very early period in the life of each, though

perhaps caused at the earliest, and being inherited at a corresponding not early period. Embryology rises greatly in interest, when we thus look at the embryo as a picture, more or less obscured, of the common parent-form of each great class of animals."

As English poet William Wordsworth once wrote, "The Child is father of the Man". To understand the detailed mechanism of biological evolution, understanding embryonic development is indispensable, because the phenotypic divergence of adult organisms must be mediated via the developmental process.

I should also emphasize that natural selection does not wait until an adult animal is fully formed before it begins to act. The opportunity for internal and environmental factors to shape an organism starts right from the beginning of the developmental process, and thus transitory embryonic characteristics are at least of equal importance to the terminally differentiated characteristics of adult forms.

Despite Darwin's early appreciation of the key role of embryology to evolution, the rediscovery of Mendelian genetics caused the two fields to drift further and further apart (Wilkins 2002). At that time, evolutionary biologists believed that evolution proceeded via a series of small, virtually imperceptible steps, also known as phyletic gradualism, whereas Mendelian geneticists believe that evolution proceeded through discrete "jumps", also known as saltationism or mutationism.

One vocal Mendelian was William Bateson, who lamented that: “By suggesting that the steps through which an adaptive mechanism arises are indefinite and insensible, all further trouble is spared. While it could be said that species arise by an insensible and imperceptible process of variation, there was clearly no use in tiring ourselves by trying to perceive that process. This labor-saving counsel found great favor.” (Orr 2005). Since embryologists can only study developmental changes that are large enough to be robustly observable, they shared very little common ground with evolutionary biologists.

This schism only worsened with the advent of the modern evolutionary synthesis in the 1930s by Fisher, Dobzhansky, Haldane and others. The new synthesis maintained that natural selection is the chief driving force behind evolution and emphasized the importance of phyletic gradualism. Ronald Fisher demonstrated using his geometric model of adaptation that mutations of infinitesimal size have a 50% probability of being beneficial, whereas larger mutations have a lower probability of being beneficial (Orr 2005). Such an interpretation effectively renders all developmental variations investigated by embryologists and developmental biologists irrelevant to the evolutionary process.

What Fisher and other prominent evolutionary biologists did not realize at that time was that the smallest mutations may not necessarily play any role in adaptive evolution - they needed to be large enough in order to escape accidental loss (Orr 2005). About 50 years later, when Motoo Kimura proposed the Neutral Theory of Molecular Evolution, he observed that the vast majority of individual mutations at the DNA and amino acid levels

had no effect at the organism level due to the redundancy of the genetic code (Kimura 1983). In addition, molecular-level mutations were predominantly fixed in a population via neutral substitution rather than natural selection, and the substitution rate is so uniform that it formed the basis of our current molecular clock dating technique.

The prevailing view on the centrality of natural selection to evolution was further criticized when palaeontologist Stephen Jay Gould proposed a thought experiment where he argued that life on Earth would look very different if we could turn back the clock and replay the "tape of Life" (Gould 1989) - due to unpredictable historical contingencies along the way. This was immediately countered by Simon Conway Morris, who argued that natural selection would constrain organisms to a limited number of adaptive options, and he used some striking examples of convergent evolution to support his stand. Of course, it is impossible to test either of these views at the planetary scale, but a recent study has investigated this by "replaying" the evolutionary process on frozen batches of bacteria (Blount et al. 2008), and they show that the appearance of a key phenotypic feature could be impossible or at least very delayed, without the random appearance of some previous enabling mutations. Results so far suggest that no matter how powerful natural selection is in the evolutionary process, the genetic history of the organism also plays an important role and cannot be simply dismissed out of hand.

These challenges to the neo-Darwinian orthodoxy promoted a new view of mutations, not merely as a non-descript and passive substrate for the environment act upon, but as the genetic source of evolutionary novelty. With the emphasis in the evolutionary biology

community slowly drifting towards internal factors and perceptible mutations, the sort of formative changes studied by developmental biologists became relevant once again, opening up the possibility of investigations into the detailed genetic causes of biological evolution.

1.2 Role of Genetic Regulation in Evolution

One important question about the role of internal factors to the evolutionary process is the type of mutations that are involved. Do all mutations contribute equally, or are some mutations more likely to result in significant phenotypic difference at the whole-organism level?

In a classic paper thirty four years ago, Marie-Claire King and Allan Wilson observed that despite substantial differences in the anatomy and behavior of chimpanzees versus human beings, their protein sequences are nearly identical, at least in their limited number of sequences they studied. They concluded that there was far more variability in untranscribed DNA using a comparative DNA hybridization approach as this work predates the development of DNA sequencing technologies. They then postulate that regulation of gene expression may play the major role in organismal evolution (King and Wilson 1975).

Their model was based on very little evidence at that time, but soon developmental studies done initially on the fruit fly *Drosophila melanogaster* would lend support to their ideas. A class of DNA-binding genes involved in the regulation of developmental

patterns, later called Hox genes, was independently discovered by Walther Gehring's group (McGinnis et al. 1984) and Thomas Kaufman's group. Hox genes are transcription factors with hundreds of downstream targets, thus any mutational change that occurs to them has the potential for large phenotypic effects, particularly to the body form of the animal. This was shown to be correct when mutations in the region of *D. melanogaster* chromosome 3 containing the Antennapedia Gene Complex (ANT-C) resulted in abnormal head development of the fly embryo (Wakimoto et al. 1984). Later studies demonstrated a high degree of functional conservation of the Hox gene family, from the nematode worm *Caenorhabditis elegans* all the way to complex vertebrates such as mouse and human beings (Purugganan 1998).

The discovery of a highly conserved gene family that underlies the body plan formation of such morphologically diverse animals was unexpected; phyletic gradualism in conventional Darwinian theory would predict that their developmental mechanisms should also be widely diversified. This apparently paradoxic discovery sparked off the new field of evolutionary developmental biology (Wilkins 2002), and now that a specific class of mutations has been identified to produce organism-level effects, they are amenable to experimental study.

Since then, a number of research groups have been working out the role of gene regulation at other loci to the evolution of various model animals. Eric Davidson's group has studied the development of the sea urchin *Stronglyocentrotus purpuratus* comprehensively and has compiled a highly-detailed genetic network map (Davidson et

al. 2002). David Kingsley's group works on the stickleback fish *Gasterosteus aculeatus* and has recently uncovered regulatory changes to the skin pigmentation in the fish; strikingly regulatory region changes in the orthologous gene in humans appear to account for the rapid evolution of skin colour in people (Miller et al. 2007). Sean Carroll's group continues work on the Drosophila, focusing on the role of cis-regulatory sequences in the evolution of morphological changes, such as wing pigmentation patterns (Gompel et al. 2005).

Carroll strongly believes that morphological evolution occurs primarily via mutations in the cis-regulatory sequence of developmental gene loci and has recently proposed a new genetic theory regarding this (Carroll 2008). His views on cis-regulatory evolution are consistent with evidence from more complex vertebrates as well, such as limb development in mice (Sagai et al. 2005) and wing development in bats (Cretekos et al. 2008). However, due to the difficulty of isolating the effects of purely cis-element sequence changes, the overall importance of cis-regulatory changes relative to coding sequence changes remain controversial today. Opponents such as Jerry Coyne and Hopi Hoekstra point out that there is still insufficient evidence for Carroll's assertion (Pennisi 2008). Whichever the case, more experimental data that directly links cis-element changes to higher organizational level effects will be helpful to resolve this debate.

I should emphasize that all these previous works focuses predominantly on the terminally differentiated morphological features of adult organisms. A complete account of evolutionary novelty must include the elucidation of the developmental processes leading

to the appearance of such features. It would be very interesting to investigate if genetic regulation also plays an important role in the evolution of transitory structures during development, especially novel morphological features that are common only to a specific class of animals - for example placental mammals.

1.3 Early Mammalian Development as a Model

Placental mammals are unique in their development in that the early embryo does not include any nutritive yolk, thus its growth has to be supported by the mother via a placenta. The need for the placental precursors to develop prior to embryo implantation is thought to be one explanation of why eutherian body plan determination is delayed relative to other vertebrates. This difference can be clearly seen when eutherian early development is compared in detail to other vertebrate animals (Fig. 1).

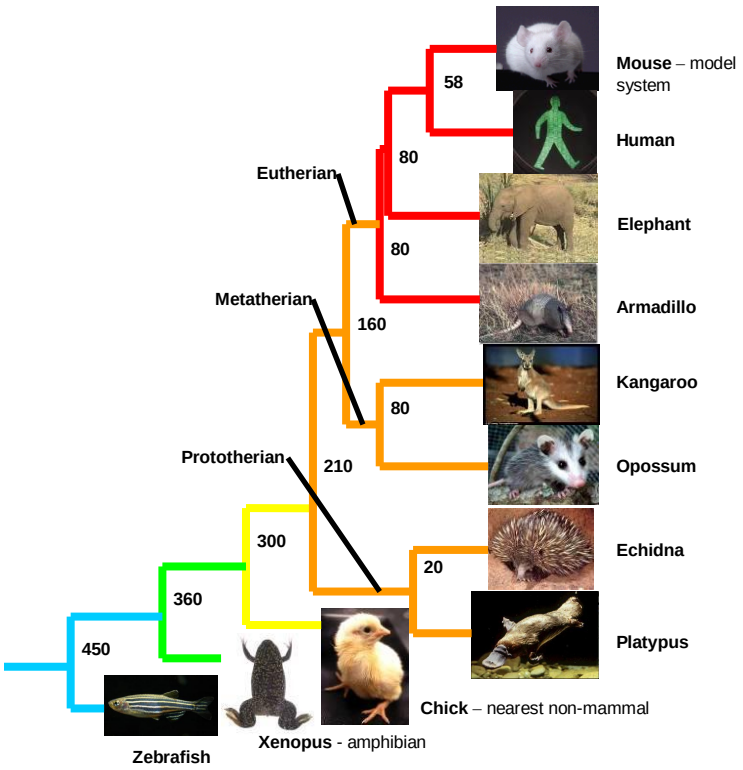


Figure 1. A phylogenetic tree of vertebrates relevant to my project. Vertebrate species in phylogenetic positions that can provide relevant sequence information to study the molecular evolution of the rounded epiblast cell type. Divergence times (Springer *et al.* 2003) shown in millions of years.

To start, in the frog *Xenopus laevis*, fertilization and embryo development occurs externally, so there is no implantation. Dorsoventral axis determining factors already exist in the oocyte at the vegetal pole, ready to migrate to a new location opposite to the sperm entry site after fertilization (Weaver and Kimelman 2004). This demonstrates that there is asymmetry very early in *Xenopus* development; after the first zygotic cell division, the two blastomeres are already different, and they are ready to develop further without delay.

In chick, fertilization occurs internally, but like in frog, there is no placental formation. Most of its embryonic development occurs externally in a hard-shelled egg. There is no

blastocyst, instead, their comparable blastula stage is a bilaminar blastoderm above the yolk, which contains the epiblast and the hypoblast. Development then proceeds without delay to gastrulation, which begins just 7 hours after fertilization (Hamburger and Hamilton 1951).

Monotremes (also called prototherians) such as the platypus nurse their young with their mammary glands and thus are considered mammals, but most of their development occurs externally, after the leathery-shelled eggs are laid. The early development of these animals is not well studied, however based on data obtained from a small number of specimens, early developmental stages resemble those of birds (Hughes and Hall 1998).

Metatherian embryonic development is also not well studied, as they are not common laboratory animals yet. Some metatherians appear to have a blastocyst stage similar to eutherians; however it lacks the inner cell mass (ICM). Instead, a region of the unilaminar blastocyst wall later becomes the epiblast that develops into the embryo proper. Moreover, since the metatherian blastocyst contains a substantial amount of yolk, preimplantation development is supported well into somitogenesis (Yousef and Selwood 1993), a much later stage compared to eutherians. Embryos are only implanted briefly before continuing development in the mother's pouch. In the North American opossum for example, implantation only occurs for the last three days of the 12.5 day gestation period, when its yolk sac placenta establishes a tenuous relationship with the uterine wall (Kumano *et al.* 2005). This suggests that metatherian early development has transitory features between non-placental and placental mammals.

Finally, all eutherian mammals have blastocysts, well developed placentas and sustained implantation in the uterus. In contrast to the frog, there is experimental evidence to show that the eutherian body plan, in particular the anterior-posterior axis, is not determined until the early egg cylinder stage at about E5.5 (Mesnard *et al.* 2004).

A summary of key features mentioned above in vertebrate early development is shown in Table 1.

	Fish / Amphibian	Chick / Prototherian	Metatherian	Eutherian
Fertilization	External	Internal	Internal	Internal
Implantation	No	No	Late, transient	Early, sustained
Placenta	No	No	Small	Well developed
Source of nutrients	Yolk	Yolk	Yolk	Mother, via placenta
Gastrulation onset	5.3 hours (zebrafish)	7 hours (chick)	7 days (opossum)	6.5 days (mouse)

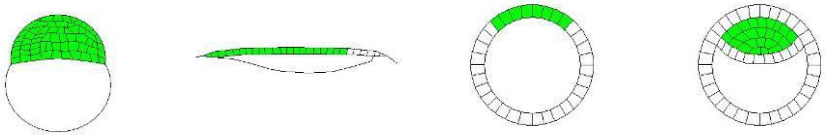
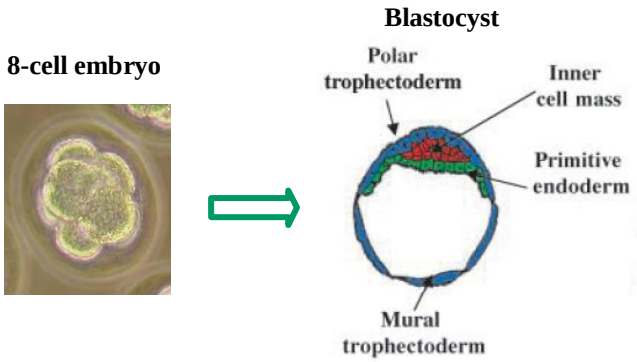


Table 1. Summary of key features in vertebrates early development.
The blastula-stage early embryo of various animals shown as schematics below the table. Green denotes cell population that will develop into embryo proper.

Since eutherian mammals have similar early development, I have selected the mouse as a prototypic eutherian to be used as my experimental model species. Mouse preimplantation development has been studied in detail. After fertilization, the 1-celled zygote is formed, dividing into the two-cell stage at E1.5 (Embryonic day) when the activation of the zygotic genome begins. The embryo then continues division until E3.5, when it becomes a blastocyst, the most relevant stage to my project. After that the blastocyst hatches from its zone pellucida, and on E4.5 it implants into the uterus. Next, at E5.5 it becomes the egg cylinder stage. Gastrulation occurs at E6.5 resulting in the formation of the three definitive germs layers – endoderm, mesoderm and ectoderm. As the primitive streak forms, the node appears on the epiblast, and the anterior-posterior axis of the embryo becomes apparent. The embryo then continues further growth and development supported by nutrition from the mother.



Adapted from Tam and Rossant, *Development* 2003

Figure 2. A schematic of the eutherian blastocyst.

The mouse blastocyst (Fig. 2) forms at the 32-cell stage and once fully expanded contains three distinct cell types. It contains a cluster of cells called the inner cell mass (ICM) of

about 20 cells, made up of two cell types, the rounded epiblast (RoE) and primitive endoderm (PrE) cells. The rounded epiblast is my terminology and I use it to distinguish this cell from the epithelialized epiblast of the egg cylinder stage, which is a slightly later and transcriptomically distinct pluripotent cell population. The ICM is contained within the trophectoderm (TE), the third cell type of the blastocyst. The TE is a functional epithelium that generates the fluid-filled cavity of the blastocyst called the blastocoel. Notably the blastocyst does not contain any yolk. The RoE is pluripotent and thus can give rise to all cell types in the embryo proper. The trophectoderm on the other hand, gives rise to placental tissue. Thus, it is a distinctly mammalian cell type that first appears in the blastocyst, leading to the development of the placenta.

In addition to the TE, I argue that the RoE is also a mammalian-specific (possibly eutherian-specific) cell type. In non-mammalian embryos, patterning occurs early in development, often before the blastula stages. This is different from the mouse, where embryonic stem (ES) cells can be derived from the RoE cells of a donor blastocyst, and when injected into the cavity of a recipient blastocyst, these cells can contribute to all cell types of the embryo proper, demonstating *in vivo* pluripotency (Evans and Kaufman 1981). These lines of evidence strongly support the view that RoE cells are of equivalent developmental potential, and that eutherian patterning is delayed compared to other animals, due to a need to set up placental precursors first. A prime example of this is the armadillo, where a single ICM in a single blastocyst normally results in quadruplets (Enders 2002). In addition, its blastocyst delays implantation for about 3.5 months in the wild. Delayed implantation (embryonic diapause) is common among mammals - almost

100 mammal species undergo diapause (Renfree and Shaw 2000), including the mouse where its blastocyst can remain in diapause for up to 30 days (Rinkenberger *et al.* 1997), demonstrating its ability to maintain its developmental potential over a long period of time. Since there is no direct equivalent of the RoE in metatherians or non-mammalian vertebrates, the RoE is uniquely eutherian, likely co-evolving with the TE and placental formation.

The focus of my thesis is on identifying the molecular changes that have led to the evolution of the RoE. The most interesting molecular changes are those that are common within all eutherians but different to all other vertebrates. Not only is this an interesting evolutionary question, but it is also relevant to ES cell biology. All these are strong reasons why I concentrated on the RoE cell type for my thesis.

So, what are the genetic changes that result in the evolution of the RoE? As mentioned earlier, King and Wilson proposed that gene regulation may have a key role in organismal evolution. It is now well accepted that alterations in the genetic regulatory architecture are central features of the evolutionary process (Davidson 2001). Thus, examining the transcriptional regulation of a developmental feature is very informative because some important transcription factors are at the upstream position of their respective gene networks. This allows them to regulate the expression profile of a number of target genes, amplifying small sequence changes into large and observable effects. As I argued that the RoE is likely to be a novel, eutherian-specific cell type in the early embryo, it thus represents an interesting model system to investigate the importance of

gene regulation in the evolutionary process. This is why my interest is in studying the molecular changes leading to the RoE genetic regulatory network.

1.4 Oct4-Sox2-Nanog Regulatory Network

In the RoE, though there are likely many other transcription factors involved in the RoE phenotype I am restricting my investigations to three well-characterized ones: Oct4 (encoded by the *Pou5f1* gene), Sox2 and Nanog. Each of these three genes, examined independently, play an important role in the normal development of a mouse. Oct4 null embryos have the earliest phenotype - they do not develop a RoE, and are peri-implantation lethal (Nichols *et al.* 1998). Sox2 knockouts fail to maintain an epiblast and arrest development before the egg cylinder stage (Avilion *et al.* 2003). Nanog deficient embryos do develop an epiblast but this was observed to differentiate immediately into primitive endoderm, resulting in death at around implantation (Mitsui *et al.* 2003, Chambers *et al.* 2003), however a recent study has shown that Nanog-negative blastocysts have substantially fewer ICM cells and fail to develop a hypoblast, indicating that it is developmental failure, rather than differentiation, that impedes Nanog-negative cells from progressing to full pluripotency (Silva *et al.* 2009)

When examined together, these three genes interact as crucial components of the transcriptional circuitry in the RoE (Fig.3). Oct4 and Sox2 proteins bind together to form a complex that recognizes and binds to the composite oct-sox element in the enhancer regions of a number of downstream targets. Some of these targets discovered so far include *Nanog*, work which I was involved in (Rodda *et al.* 2005) and others (Kuroda *et*

al. 2005), in addition to *Pou5f1* (Chew *et al.* 2005) and *Sox2* (Tomioka *et al.* 2002) themselves in an auto-regulatory loop. Nanog has also been shown to be in its own auto-regulatory loop (Loh *et al.* 2006).

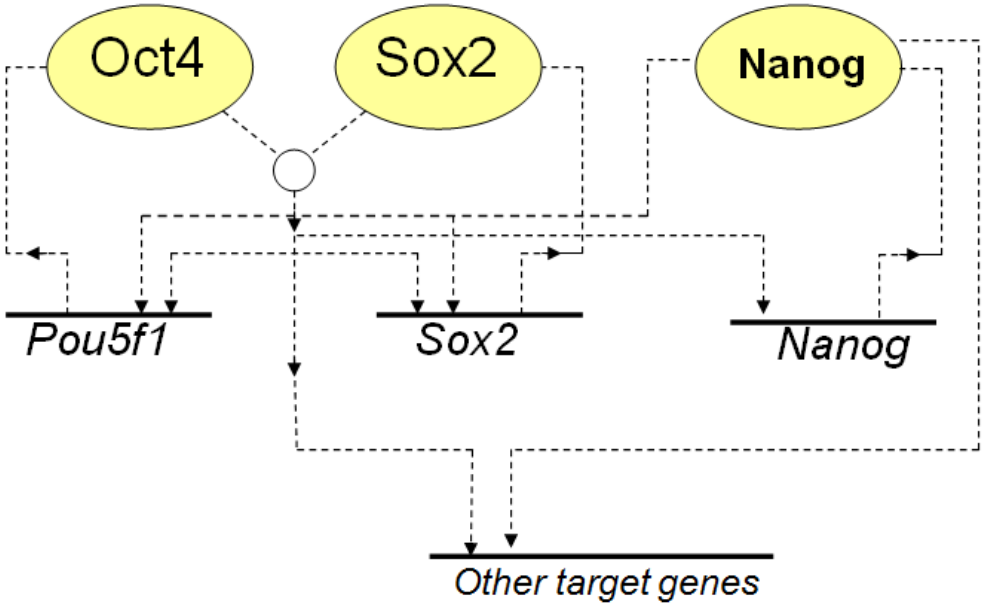


Figure 3. Diagram of the Oct4-Sox2-Nanog regulatory circuit.

Sox2 expression and function is not restricted to the RoE, indeed Sox2 is known to be essential to neuronal development. In this tissue it is known to partner with other POU class transcription factors such as Oct1 or Brn-1/2 (Miyagi *et al.* 2006). In fact, the structures of Oct1-Sox2-DNA ternary complexes have been solved (Remenyi *et al.* 2003, Williams Jr. *et al.* 2004). Both Oct1 and Sox2 use part of their DNA binding domain to interact with each other. The data emphasized the importance of this Oct-Sox protein-protein interface, when bound to the oct-sox element, to the activity of the whole complex. Using molecular modeling, knowledge gained from mutation studies on Oct1 can be extended to Oct4.

1.5 EC and ES Cell Culture System

To investigate cell-level effects, embryonal carcinoma and ES cell systems are used. Historically, embryonal carcinoma (EC) cells were the first pluripotent cell type to be isolated and used for long-term culture (Martin and Evans 1974). Derived from embryonic germ cell tumours called teratocarcinomas, when EC cells are injected into a mouse blastocyst, they can be regulated by the recipient environment and contribute to the somatic tissues of the chimeric mouse (Brinster 1974). EC cells are easy to grow, proliferate quickly and indefinitely (Martin and Evans 1974) without the need for feeder cells. However, they have their limitations since they have an abnormal chromosome complement and rarely contribute to the germ line (Bradley et al. 1998), weakening the potential of EC cells for studying embryo development and gene function.

ES cells, on the other hand, are usually obtained from the inner cell mass of a 3.5 day mouse blastocyst (Evans and Kaufman 1981) and cultured on a layer of inactivated mouse embryonic fibroblast cells. They can also be isolated from a disaggregated 16-20 cell morula, or microdissected from the epiblast of a 4.5 day embryo. Like EC cells, ES cells also can differentiate into all three embryonic germ layers when injected into mice (Bradley et al. 1984). However, ES cells have an added advantage of higher germline transmission efficiency and normal chromosome complement, thus making them a useful tool for genetic studies. Moreover it is the closest *in-vitro* equivalent of the RoE, sharing many morphological features and molecular markers with the endogenous cell type.

1.6 iPS Cell Culture System

The advent of the induced pluripotent stem cell (iPS) system provides an excellent tool for the direct investigation of the molecular factors that are crucial for pluripotency (Takahashi and Yamanaka 2006).

Mouse embryonic or adult fibroblast cells are infected with retroviral vectors which contain four key pluripotent factors, Oct4, Sox2, c-Myc and Klf4. The overexpression of these proteins reprograms the fibroblasts into iPS cells which have similar morphology and proliferation ability as ES cells. With the iPS culture system, versions of the pluripotent factors, such as Oct4, can be modified at the sequence level to resemble their homolog in other species to find out if they can also induce pluripotency just like mouse Oct4.

In this replacement approach, the Oct4 ortholog that fails to induce pluripotency would come from the species whose ancestors diverged from eutherian mammals prior to the evolution of pluripotent functions in the Oct4 protein.

1.7 Project Strategy

The first step is to identify significant changes in protein coding and *cis*-regulatory sequences that have occurred in at least some regions of *Pou5f1*, *Sox2*, and *Nanog* in the proto-eutherian mammal. I hypothesize that some of these molecular changes contributed to the uniqueness of the eutherian mammal preimplantation embryo. The goal of my thesis is to characterize some of the more salient molecular changes that have occurred in *Pou5f1*, *Sox2* and *Nanog* and some of their *cis*-regulatory targets that were essential in the evolution of the eutherian mammal RoE population of cells.

I begin my investigation of the transcriptional network in the RoE by performing sequence analysis of both the protein coding sequence and the cis elements of *Sox2*, *Pou5f1* and *Nanog*. The goal is to identify eutherian-specific elements that may be functionally important in the context of the pluripotent cell. Sequences are drawn from a number of vertebrate species in relevant phylogenetic positions, to allow common eutherian sequences to become apparent, while minimizing noise from possible species-specific sequences. Many eutherian-specific changes are likely be found, so only some of these with the most striking differences will be functionalized. To investigate the importance of these elements, a number of mutation and chimeric constructs are to be made, using a predominantly loss-of-function strategy. The effects of these modifications are then evaluated using the EC, ES and iPS cell culture system described earlier.

Chapter 2: Obtaining Sequence Data

2.1 Overview

To determine the selection of animal species where sequences should be obtained, it is helpful to know the early evolutionary history of mammals. The earliest known mammaliaform in the fossil record is the *Hadrocodium wui* which dates back to the Early Jurassic period approximately 195 million years ago (Luo et al. 2001). Fossil specimens with anatomical features that identify them as ancestral forms of prototherian, metatherian or eutherian mammals start appearing around 124.6 million years ago (Fig.4).

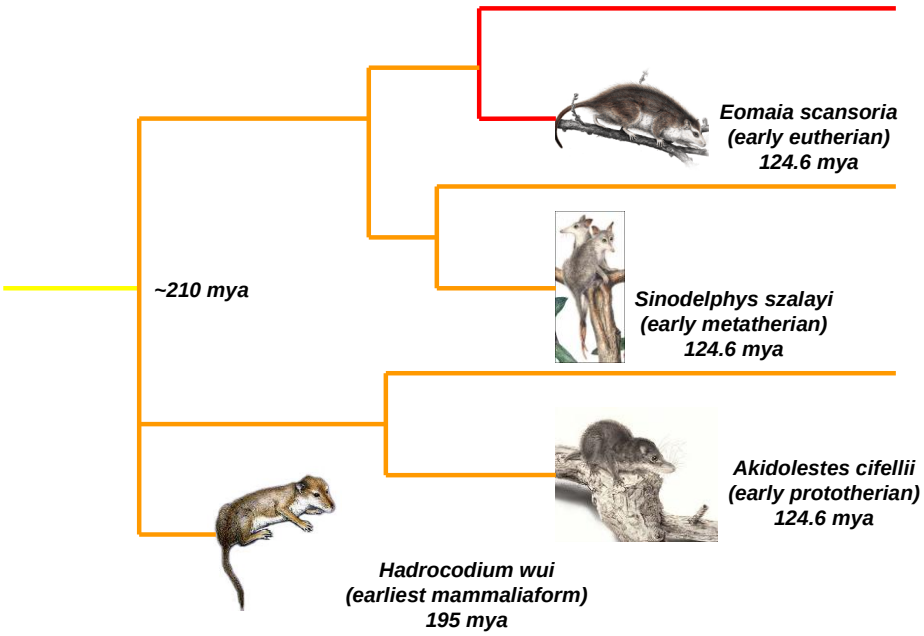


Figure 4. Fossil Record of Early Mammals.

This data, together with molecular clock estimates, suggest that the base of mammalian radiation occurred around 210 million years ago. Of course, there is currently no way of

obtaining sequences from these fossil specimens - this information would have to be obtained from modern vertebrate species. Since my model organism is the mouse (*Mus musculus*), as a general guide any mammal species that diverged from their last common ancestor with the mouse less than 124.6 million years ago would be categorized as in-group organisms, whereas other vertebrate species that diverged more than 210 million years ago would be categorized as out-group organisms.

Species	Category	Target	Project Status	BAC library
Mouse	Eutherian	Assembled	Complete	
Rat	Eutherian	Assembled	Complete	
Human	Eutherian	Assembled	Complete	
Dog	Eutherian	Draft assembly 8X	Complete	
Cow	Eutherian	Draft assembly 6X	Complete	
Elephant	Eutherian	Low coverage <2X	Incomplete	CHORI
Armadillo	Eutherian	Low coverage <2X	Incomplete	CHORI
Opossum	Metatherian	Draft assembly 7X	Incomplete	CHORI
Kangaroo	Metatherian	Low coverage <2X	Incomplete	AGI
Echidna	Prototherian	Not in pipeline to be sequenced		AGI
Platypus	Prototherian	Draft assembly 6X	Incomplete	CUGI
Chick	Bird	Draft assembly 6X	Complete	
Xenopus t.	Amphibian	Draft assembly 8X	Complete	
Zebrafish	Fish	Draft assembly 7X	Complete	

Table 2. Availability of Sequence Information.
(Sources - http://www.genome.gov/10002154 and http://www.ensembl.org) Target figures denote extent of genome coverage. CHORI = Children's Hospital Oakland Research Institute, AGI = Arizona Genomics Institute, CUGI = Clemson University Genomics Institute

Table 2 represents the status of various genome projects at the start of my project in 2004. In this table, genome projects in black were complete and at least had draft assemblies, so that sequences can be obtained by searching online databases. Where the sequences were not complete I performed a cross-species BLAST against their trace files and assemble them using VectorNTI (Invitrogen).

For the species indicated in red, there was limited online information, so I screened BAC genomic libraries of these species by hybridization and performed de novo sequencing of BAC clones that I pulled out. These species are in key phylogenetic positions with respect to the base of mammalian radiation, and I have selected two species each of distant eutherian, metatherian and prototherian mammals, so that there will be enough sequence information to reduce noise from species-specific sequence changes. The kangaroo (*Macropus eugenii*), for example, is 80 million years diverged from the opossum (*Monodelphis domestica*), so common sequences between these two species are more likely to be metatherian-specific. Similarly, the elephant and armadillo are the most distantly-related eutherians to the mouse. Using this strategy, more sequence information provides greater confidence to identify eutherian-specific sequences.

2.2 Materials and Methods

As mentioned earlier, if there was a genome sequencing project in progress for an animal species then sequence data is directly obtained via database searches, primarily from these four online sources:

1. Ensembl (www.ensembl.org) - European Bioinformatics Institute and the Wellcome Trust Sanger Institute.
2. VISTA (http://pipeline.lbl.gov/cgi-bin/gateway2) - Genomics Division of Lawrence Berkeley National Laboratory.
3. NCBI (http://www.ncbi.nlm.nih.gov/) - National Center for Biotechnology Information.

4. UCSC (http://genome.ucsc.edu/) – University of California, Santa Cruz, Genome Bioinformatics.

If the assembly of the sequences in the genome project was not complete, then I performed a cross-species BLAST using trace files obtained from the Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi?) and assembled currently available trace files using the contig assembly tool in the VectorNTI programme.

Where trace file information was sparse, I have purchased BAC libraries from these three sources:

1. CHORI (http://bacpac.chori.org/) – BACPAC Resource Center, Children's Hospital Oakland Research Institute.
2. AGI (http://www2.genome.arizona.edu/welcome) - Arizona Genomics Institute.
3. CUGI (https://www.genome.clemson.edu/) - Clemson University Genomics Institute.

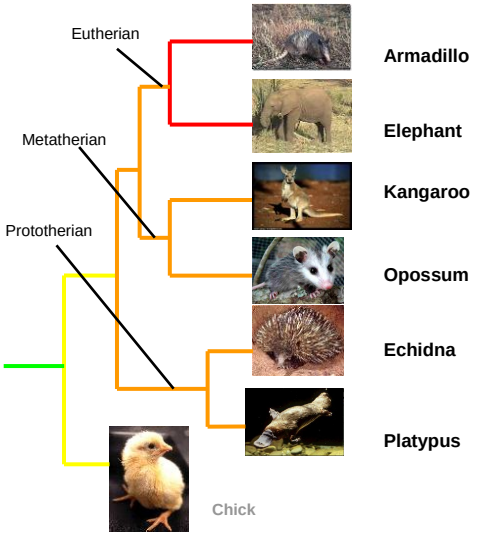


Figure 5. Screening BAC libraries for key mammalian species

This phylogenetic tree illustrates the relative positions of mammalian species where BAC screening was necessary (Fig.5). Southern hybridization was used to obtain additional sequence information for elephant, armadillo, kangaroo, opossum and platypus. Currently there is insufficient trace file information for the echidna to do BAC library screening.

BAC libraries of elephant, armadillo, opossum, kangaroo and platypus were obtained. Each library has 6 to 13 high density nylon filters, containing 18,432 clones spotted in duplicate per filter, which were screened by southern blot using oligo probes that were end-labeled with radioactive ^{32}P ATP. I designed these oligo probes (~30bp) with limited *Pou5f1* or *Nanog* sequence information from trace files, from a unique region of the gene such as the first 30bp of the coding sequence (Table 3).

Species	**Gene**	**Location**	**Sequence**
Elephant	Pou5f1	Exon 1	ATGGCGGGACACCTGGCTGCCGACTTTGCC
Armadillo	Pou5f1	Exon 1	ATGGCAGGACACCTGGCTCCGGACTTTGCC
Opossum	Pou5f1	Exon 5	TCACCCCGGGAGGATTTTGAGGCAGCTGGC
Kangaroo	Pou5f1	Exon 5	TCACCTCGAGAAGATTTTGAGGCAGCTGGT
Platypus	Tcf19	Exon 1	ATGCTGCCCTGCTTCCAGCTGCTGCGCATG
Elephant	Nanog	Exon 1	ATGAGTGTGGATCTAGCTTCTCCCCAAAGC
Armadillo	Nanog	Exon 1	ATGAGTGTGGATCTAGCTTCTCCCCAAAGT
Opossum	Nanog	Exon 2	CAGAACAAGCCCAAGACCCATCAGGGAAAA
Kangaroo	Nanog	Exon 2	AACAAGCCCAAGATCCATCAGGGAAAAGAA
Platypus	Slc2a3	Exon 6	CAGGACATCCAGGAGATGAAGGAGGAGAGT

Table 3. Sequence of the oligo probes used for BAC screening.

Platypus library screening is more challenging since there were no trace files in mapping to a putative *Pou5f1* or *Nanog* at that time. Instead, a probe designed to *Tcf19*, a neighboring gene just 2kb away from all currently known mammalian *Pou5f1*, was used. Similarly, a probe to *Slc2a3*, a neighboring gene to Nanog, was used.

Potential positive clones were visualized as bright spots on autoradiographs, or on storage phosphor screens which were then read by the Typhoon phosphorimager (GE Healthcare). Radiochemical levels were optimized in order to read the spots clearly without overexposing the filter (Table 4).

	For X-ray film	For phosphor screen
Pack size	250 μCi Gamma ^{32}P ATP (10 μCi per μl)	250 μCi Gamma ^{32}P ATP (10 μCi per μl)
Volume used	2.5 μl per filter	1.0 μl per filter
Radioactivity of labeled probe	2.0 x 10^6 cpm/μl	Estimated ~ 1 x 10^6 cpm/μl
Radioactivity after hybridization	30000 cpm at 1 cm distance	10000 cpm at 1 cm distance
Optimized exposure time	1 hour for 30000 cpm 3 hour for 10000 cpm 15 hours for 2000 cpm	1 hour for 10000 cpm
Optimized exposure radioactivity	1.8 x 10^6 counts in total	600000 counts in total

Table 4. Optimized radiochemical levels for autoradiographs and phosphor screens.

The BAC identity of these spots were decoded using a three-step protocol – this information was recorded into an Excel file (see Appendix A) and the BACs were purchased as agar stabs. Next, PCR screening was done using genomic primers. The entire workflow in screening BAC libraries is summarized in Figure 6, and details of the protocol can be seen in Appendix B.

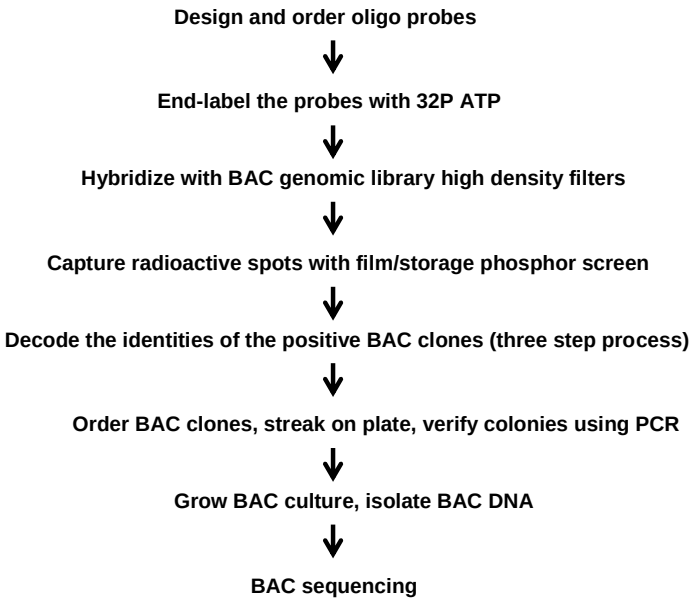


Figure 6. Summary of BAC screening workflow (see Appendix B for details).

The DNA was then isolated and purified using a BAC DNA preparation kit. This DNA can be used for sequencing or act as reagents for functional studies later. Finally relevant regions of those BACs were sequenced. All sequencing was done using capillary sequencing runs via BAC-end sequencing and primer walking. The difficulty of this approach resulted in numerous failed reads but there was sufficient sequence obtained to identify gene-specific sequence as well as pseudogenes.

All the raw sequence information from online databases, trace file assemblies and BAC sequencing reads were converted to VectorNTI files for compilation and analysis.

2.3 Results and Discussion

A total of 2 authentic Nanog clones were verified (elephant and opossum) and the rest were pseudogenes (armadillo) with no intronic sequence, or failed reads.

A total of 3 authentic *Pou5f1* clones were verified from elephant, opossum and platypus in addition to a number of pseudogenes (armadillo, kangaroo). The platypus BAC clone was first pulled out with an oligo to the *Pou5f1* neighbouring gene *Tcf19* thus sequencing of the BAC first verified the presence of *Tcf19* in this clone. When primers to the *Pou5f1* gene were used to amplify the same clone, the PCR yielded a fragment of the appropriate size. Subsequent BAC sequencing was able to read most of exon 4, intron 4 and exon 5 of platypus *Pou5f1*. This indicates the platypus *Pou5f1* is in close proximity to *Tcf19*, lying within the same BAC construct, therefore in the same genomic context (ie. syntenic) as eutherian mammal *Pou5f1* genes.

This discovery of platypus *Pou5f1* is intriguing as prior to this a syntenically positioned *Pou5f1* had not been found in the chick (Soodeen-Karamath and Gibbins 2001), suggesting that the location of the *Pou5f1* gene might have been a uniquely eutherian novelty. Finding it in the prototherian platypus thus rules out this possibility, and as the platypus does not have a blastocyst stage, *Pou5f1* is not specific to this eutherian embryonic feature.

However, this discovery opened the possibility that changes within the platypus Oct4 protein, rather than the existence of the gene itself, could account for the differences

between platypus and eutherian embryo development, which will be investigated in detail in Chapter 4.

Chapter 3: Sequence Data Analysis

3.1 Overview

The purpose of sequence analysis is to align and compare all the relevant sequence information in order to identify significant eutherian-specific sequence changes that are likely to have a large phenotypic effect on early embryo development.

In the simplest scenario, the mere appearance of a gene in a novel genomic context may be a major factor. This is not the case for *Sox2*, since it is a gene that has existed for a long time in vertebrate evolutionary history. Its coding sequence is highly conserved from fish to mouse (Table 5). To verify if there are direct orthologs to mouse *Sox2*, the synteny of surrounding genes, especially *Fxr1*, is examined. Here you can see that it has been in the same genomic context since the fish (Fig.7).

	Mouse	Rat	Human	Dog	Cow	Opossum	Frog Xt	Puffer fish Tr	Zebrafish
Mouse	100	100	98	97	98	84	88	83	87

Table 5. Sox2 protein coding sequence identity (% of amino acids)

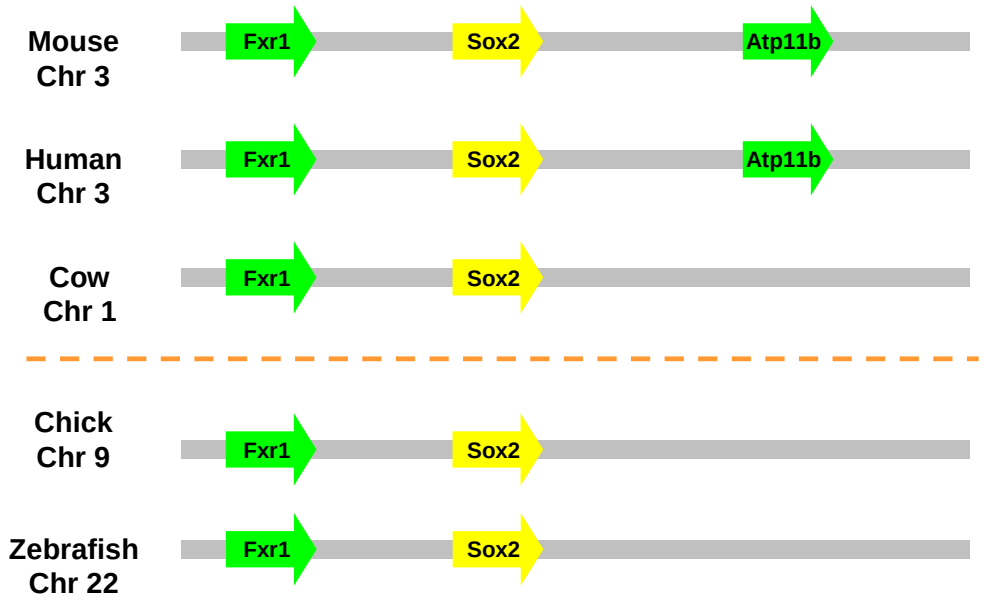


Figure 7. Sox2 gene synteny map.
Orange dash denotes the boundary between mammals and non-mammals.

By comparison, *Nanog* appears to be a newer gene and its amino acid sequence is certainly less conserved over evolutionary time (Table 6). Based on multi-species data, the gene seems to have first appeared in its eutherian genomic context in the chick, since the equivalent region in the frog does not have the *Nanog* gene (Fig. 8). This means that the appearance of *Nanog* already predates the base of mammalian radiation, so the gene itself is not a eutherian-specific change.

	Mouse	Rat	Human	Dog	Cow	Opossum	Chick
Mouse	100	84	68	64	66	45	20

Table 6. Nanog protein coding sequence identity (% of amino acids)

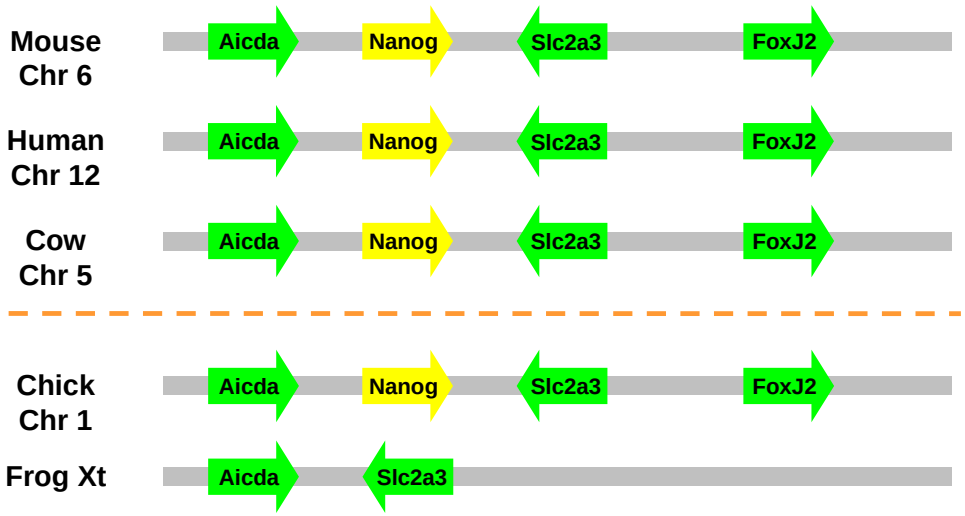


Figure 8. Nanog gene synteny map.

Pou5f1 also appears to be a newer gene and is not highly conserved over evolutionary time (Table 7). Although there are non-mammalian homologs of this gene, they share only limited sequence identity and seem to be at a different genomic context. Moreover at the beginning of this project a chick homolog had not yet been found (Soodeen-Karamath and Gibbins 2001). Thus I postulated that *Pou5f1* could have appeared around the base of mammalian radiation, which would be a significant discovery considering the vital importance of Oct4 to pluripotent cell development (Fig 9.).

	Mouse (Pou5F1)	Human (Pou5F1)	Cow (Pou5F1)	Opossum (Pou5F1)	Chick (PouV)	Frog (PouV)	Puffer fish (Pou2)	Zebrafish (Pou2)
Mouse	100	87	86	Partial	46	31	27	26

Table 7. Oct4 protein coding sequence identity (% of amino acids)

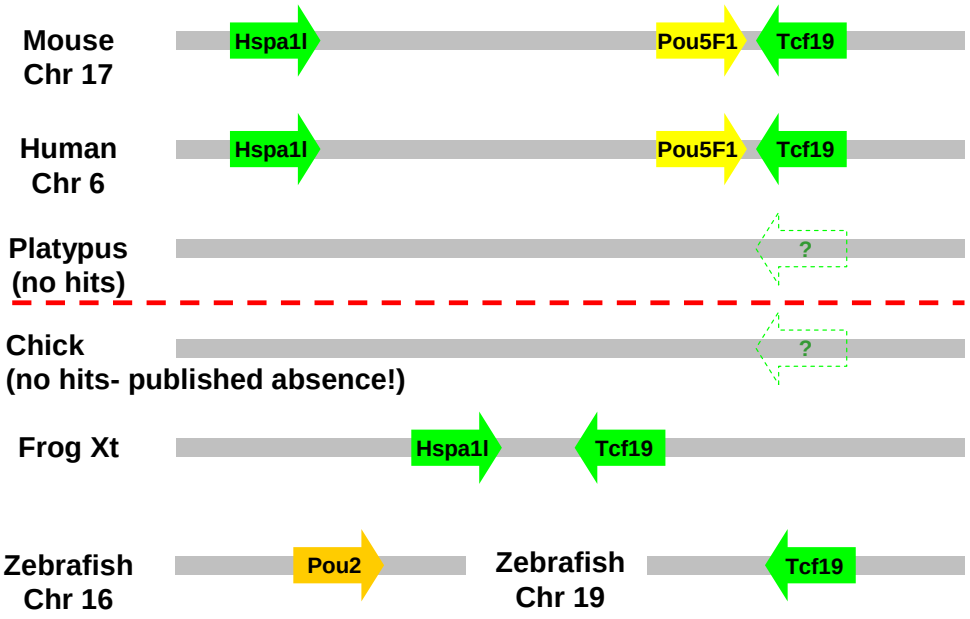


Figure 9. Initial *Pou5f1* gene synteny map (2004)

However, after the BAC sequencing effort and with the availability of sequences from more animal species, this postulate is no longer tenable.

Firstly, I was able to find platypus *Pou5f1* in close vicinity to *Tcf19*, indicating that the gene itself exists prior to the divergence with eutherian mammals. Next, a chick homolog of *Pou5f1*, called *PouV*, was found in 2007 and has a 46% sequence identity with mouse *Pou5f1* (Latvial et al. 2007). While the researchers provided evidence that chick PouV could support pluripotency in chick ES cells, *cPouV* is not located in the same genomic context as mammals - indeed it is in the vicinity of the *Fut7*, more similar to fish *Pou2*. PouV homologs that are located in this context are named *Pou5f2* for clarity (Fig 10).

Thus there still remained the possibility that the platypus *Pou5f1* is the first homolog that existed in the eutherian genomic context.

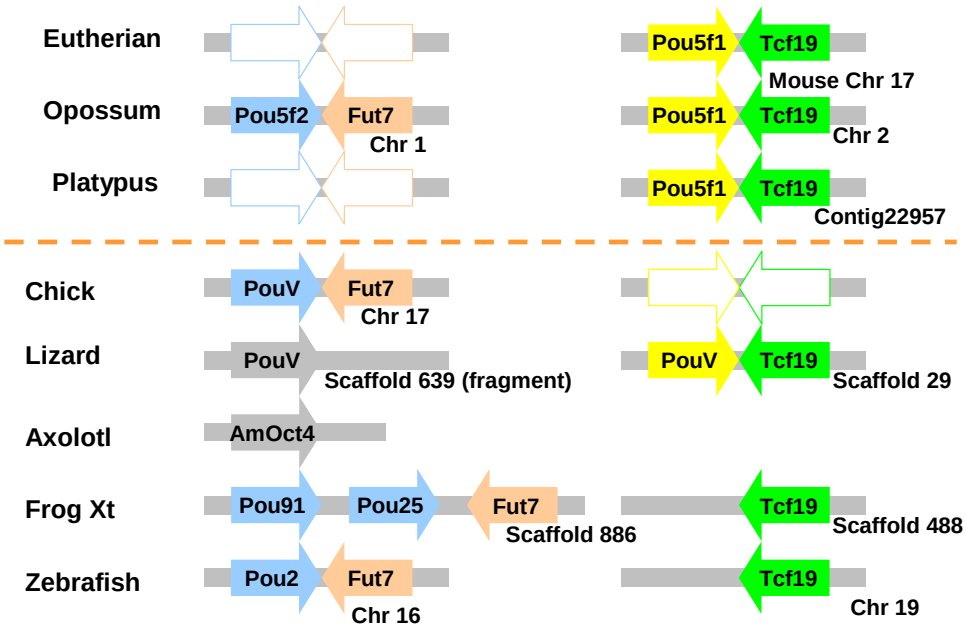


Figure 10. Latest *Pou5f1* gene synteny map (2009)

However, new sequence information from UCSC indicated the presence of lizard (*Anolis carolinensis*) homologs of *Pou5f1*. While the assembly was not yet complete, there was sufficient sequence to show that one such *PouV* homolog exists near *Tcf19*, placing it in the same genomic context as eutherian mammals and thus pushing back the appearance of *Pou5f1* long before the base of mammalian radiation (Fig. 10).

From this broad overview of the evolutionary history of *Pou5f1*, *Sox2* and *Nanog*, we can see that the appearance of these three genes were not coincident with the emergence of eutherian mammals. It is therefore necessary to perform more detailed multi-species analyses - of the DNA regulatory regions at the nucleotide level, and protein coding

sequences at the amino acid level - in order to uncover interesting eutherian-specific changes.

3.2 Materials and Methods

VectorNTI (Invitrogen) was used for most of the *in silico* sequence analysis, such as assembly and alignments.

Online tools such as ExPASy Proteomic Tools (http://www.expasy.ch/tools/) were also used for protein analysis.

RasMOL (ver 2.6) was used for visualizing protein structures.

3.3 Results of cis-element Analysis

As mentioned in the introduction, Oct4 and Sox2 bind synergistically to form a complex that binds to the composite oct-sox element in the enhancer regions of target genes. The element is a core component of the pluripotent transcriptional network; it consists of 15 base pairs of which seven of them are bound by Sox2 and the adjacent eight, the octamer sequence, are bound by Oct4. This element has been identified in a number of target genes, for example in the cis-regulatory regions of *Fgf4* (Yuan et al. 1995) and *Sox2* (Tomioka et al. 2002) itself.

More recently, using the chromatin immunoprecipitation (ChIP) and paired-end ditag sequencing strategy, Loh and colleagues produced a genome-wide map of Oct4 and Nanog binding sites in mouse ES cells (Loh et al. 2006). They were able to distill a sox-oct consensus binding logo, which serves as a general guide of the prototypic sox-oct binding site (Fig .11).

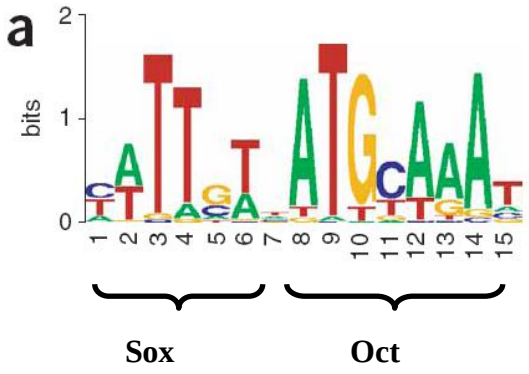


Figure 11. Oct-Sox Consensus Binding Logo (Loh *et al.* 2006)

Next, I examined the sequence conservation of the sox-oct element in the *Sox2* regulatory region bound by the Sox2/Oct4 protein complex, this molecular interaction considered to contribute to *Sox2* auto-regulation. This element is found in a highly conserved region 1.2 kb downstream of the *Sox2* mRNA's 3' UTR. I used sequence in this region for multi-species BLAST analysis in order to find the equivalent region in other mammal and non-mammal vertebrate species. As shown, the sox-oct element is quite well conserved with 11 of the 15 positions invariant (shown in bold print) from chick to mouse (Fig. 12). Since it is already present in the opossum and the chick, this finding precludes the possibility that the simple presence of the element is a eutherian-specific change. Focusing on the nucleotide level changes, the chick sox-oct element is the most divergent,

differing from the eutherian consensus by three base pairs. To investigate if these changes will result in an impairment of transcription factor binding, promoter assays were performed. This is discussed in detail in Chapter 4.

Sox2
Binding site is ~1200bp downstream of Sox2 3'UTR.
Tomioka et al. 2002 *NAR*

```
                          Sox2    Oct4
Mouse        CTCGGGCAGCCATTGTGATGCATATA-GGATTATT
Rat          CTCGGCCAGCCATTGTGATGCATATA-GGATTATT
Human        CCTGGCCAGCCATTGTAATGCATATACGGATTATT
Dog          CCTGGCCAGCCATTGTAATGCATATACGGATTATT
Elephant     CCCGGCCAGCCATTGTAATGCATATACGGATTATT
Opossum      GCTGCCCGGCTTTTGTAATGCATATA-GGATTATT
Chick        GCTGTGCGGCGTTTGTAATGCATCTGGGGATTATT
Frog                     No sox-oct site
```

Figure 12. Alignment of sox-oct binding site in Sox2

In contrast, the sox-oct binding site of *Nanog* is not found in the chick or the opossum. It is conserved only in eutherians and has remained unchanged for over 250 million years of cumulative evolution (Fig. 13). This strongly suggests the functional importance of this sequence. Experiments done by myself and others from our lab have shown that the sox-oct element is important to drive Nanog expression in ES cells (Rodda *et al*. (2005); see Chapter 4). This appears to be a striking example of eutherian-specific cis-evolutionary change.

Nanog
Binding site is ~181 bp upstream of TSS.
Rodda et al. 2005 *JBC*

```
                          Sox2   Oct4
Mouse        CCACCATGGACATTGTAATGCAAAAGAAGCTGTAA
Rat          CCACCAAGGACATTGTAATGCAAAAGAAGCTGTAA
Human        TCACCAAGGCCATTGTAATGCAAAAGTAGCTGCAG
Cow          TCACCAAGGCCATTGTAATGCAAAAGAGAGTTGCA
Elephant     TCATCAAGTCCATTGTAATGCAAAAGTTCCTGAAA
Opossum                No sox-oct site
Chick                  No sox-oct site
```

Figure 13. Alignment of sox-oct binding site in Nanog

Like *Sox2*, *Pou5f1* is also auto-regulated by the Sox2/Oct4 complex. The enhancer element is highly conserved and there are only three nucleotide differences from elephant to mouse (Fig. 14). The opossum *cis*-region sequence is not available as the trace files do not extend far enough in the 5' direction; similarly there is no data for platypus as the primer walking effort was unable to yield the first exon of the gene. There appears to be no chick equivalent of this conserved region, which is understandable since chick *PouV* is not located in the same genome context as its eutherian counterparts.

Oct4
Binding site is ~1992 bp upstream of TSS.
Chew et al. 2005 *MCB*

```
                         Sox2    Oct4
Mouse        TATCATGCACCTTTGTTATGCATCTGCCGTCTGCC
Human        ATCACGGCACCTTTGTCATGCATCTCTCTGCTGTC
Dog          ATCACGGCACCTTTGTCATGCATCTATCTGCTGTC
Elephant     ATCACAGCACTTTTGTCATGCACCTATCTGCTGTC
Chick                  Not found
```

Figure 14. Alignment of sox-oct binding site in *Pou5f1*

3.4 Results of Coding Sequence Analysis

Sox2 has existed for a long time in vertebrate evolutionary history, present in ancient lineages such as the fish. It is highly conserved from fish to mouse, the yellow blocks showing identical sequence (Fig .15). In fact in its 79 amino acid DNA binding domain, the HMG box, only two amino acids differ from fish to mouse. There is very limited diversification in the N and C-terminal ends as well – most of these changes do not conform to any phylogenetic pattern and are likely to be neutral substitutions. As such, I can conclude that functionally important eutherian-specific amino acid changes do not exist in the Sox2 coding sequence.

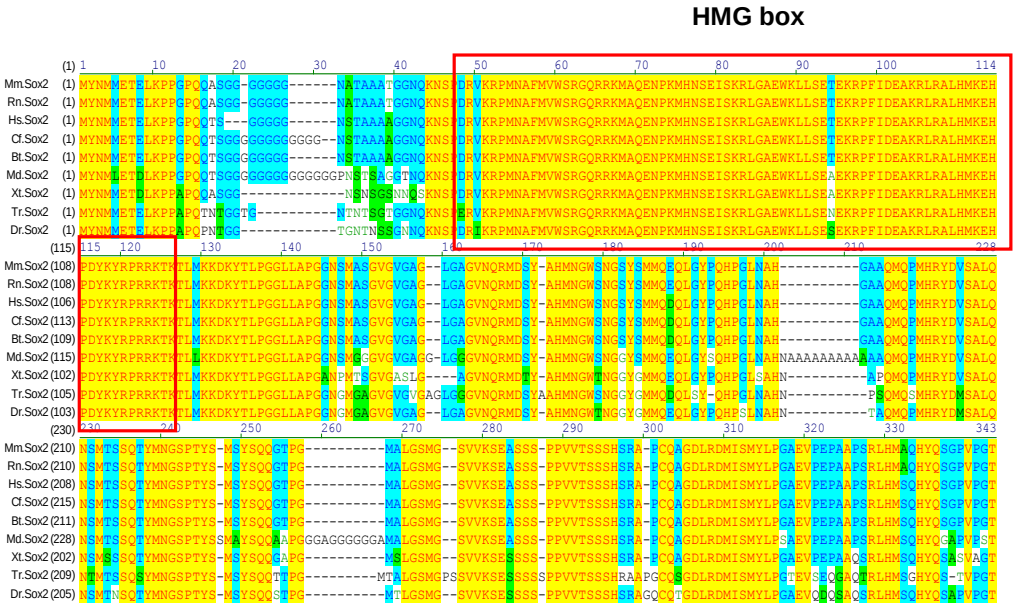


Figure 15. Sox2 protein alignment

Yellow denotes identical bases, blue and green are similar bases. White sections are dissimilar and dash sections cannot be aligned (such as insertions and deletions). Species abbreviations for this alignment: Dr = *Danio rerio* (Zebrafish), Tr = *Takifugu rubripes* (Fugu), Xt = *Xenopus tropicalis* (Frog), Md = *Monodelphis domestica* (Opossum), Bt = *Bos taurus* (Cow), Cf = *Canis familiaris* (Dog), Hs = *Homo sapiens* (Human), Rn = *Rattus norvegicus* (Rat) and Mm = *Mus musculus* (Mouse).

Nanog, on the other hand, appears to be a newer gene and is more variable. Only the DNA-binding homeodomain is quite conserved, but even within this region there is variability among eutherian species. There is a region where interesting changes have occurred – the tryptophan repeat region (WR), which is highlighted in the red box (Fig. 16).

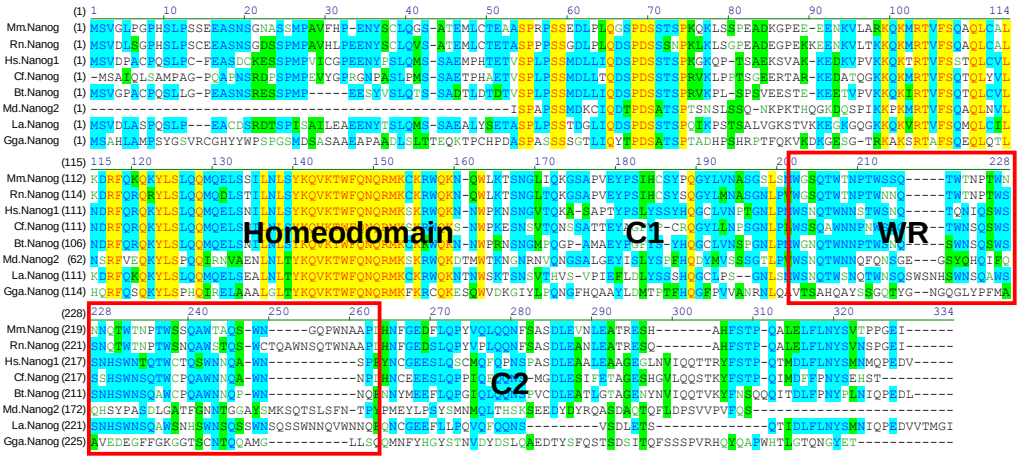


Figure 16. Nanog protein alignment

Yellow denotes identical bases, blue and green are similar bases. White sections are dissimilar and dash sections cannot be aligned (such as insertions and deletions). Species abbreviations for this alignment: Gga = *Gallus gallus* (Chicken), La = *Loxodonta africana* (Elephant), Md = *Monodelphis domestica* (Opossum), Bt = *Bos taurus* (Cow), Cf = *Canis familiaris* (Dog), Hs = *Homo sapiens* (Human), Rn = *Rattus norvegicus* (Rat) and Mm = *Mus musculus* (Mouse).

The tryptophan repeat region has previously been shown to be important for the transactivation ability of Nanog (Pan and Pei 2005). The mechanism of its activity is not yet known. There are at least 8 tryptophan repeats in eutherian mammals, but only two in the opossum and none in the chick (Table 8). This suggests that the appearance of these repeats occur at around the base of mammalian radiation (Fig. 17).

Species	Mouse	Rat	Human	Dog	Cow	Elephant	Opossum	Chick
No. of W repeats	10	11	8	9	9	12	2	0

Table 8. Number of Tryptophan repeats in Nanog transactivation domain

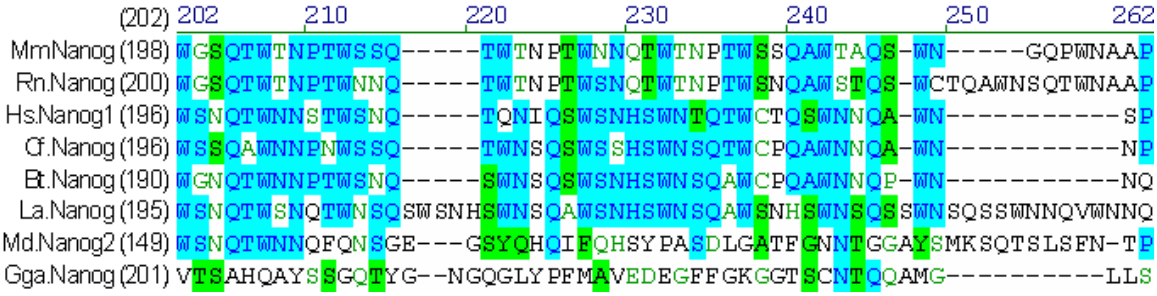


Figure 17. Detailed alignment of the Nanog transactivation domain
Blue and green are similar bases. White sections are dissimilar and dash sections cannot be aligned (such as insertions and deletions). Species abbreviations for this alignment: Gga = *Gallus gallus* (Chicken), Md = *Monodelphis domestica* (Opossum), La = *Loxodonta africana* (Elephant), Bt = *Bos taurus* (Cow), Cf = *Canis familiaris* (Dog), Hs = *Homo sapiens* (Human), Rn = *Rattus norvegicus* (Rat) and Mm = *Mus musculus* (Mouse).

Like Nanog, Oct4 is also not highly conserved outside of its DNA binding domain, which is made up of two parts: the POU domain and the homeodomain. This is a functionally important part of the protein, especially the POU domain which is crucial for protein interactions with Sox2 and is highlighted in a red box (Fig. 18). More details about the POU domain will be discussed later.

Figure 18. Oct4 protein alignment

Yellow denotes identical bases, blue and green are similar bases. White sections are dissimilar and dash sections cannot be aligned (such as insertions and deletions). Species abbreviations for this alignment: Dr = *Danio rerio* (Zebrafish), Tr = *Takifugu rubripes* (Fugu), Bt = *Bos taurus* (Cow), Cf = *Canis familiaris* (Dog), Hs = *Homo sapiens* (Human), Rn = *Rattus norvegicus* (Rat) and Mm = *Mus musculus* (Mouse).

This initial protein alignment of Oct4 suggested there were eutherian-specific residues within the DNA binding domain. To determine if this were true I did a more comprehensive alignment of this region with multiple mammalian sequences: 15 eutherian sequences, two metatherian species, and the prototherian platypus. Strikingly, this revealed 12 amino acid positions that are apparently under eutherian-specific selection – these are marked by small arrows (Fig .19) – they are invariant in all the eutherian species but differ in the non-eutherian mammals. Notably, there is a great similarity between platypus and metatherian amino acid sequences in this region, highlighted within the red box.

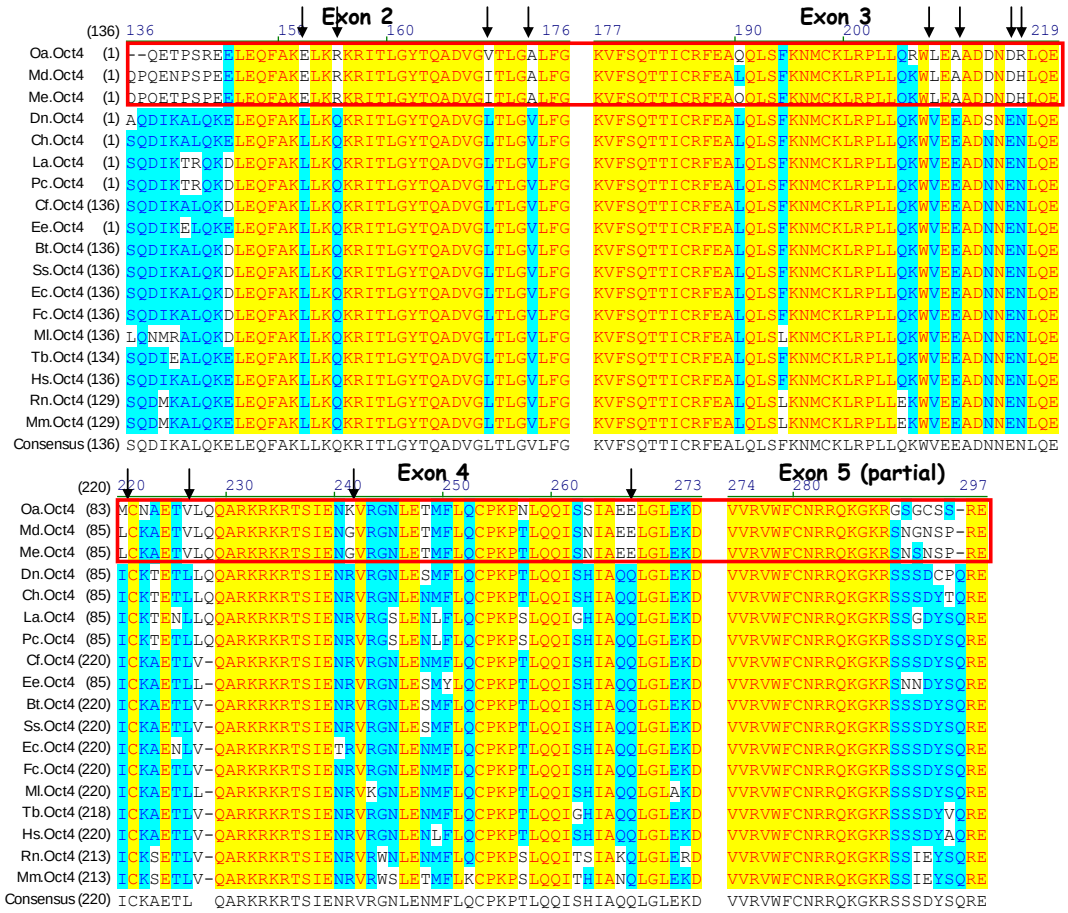
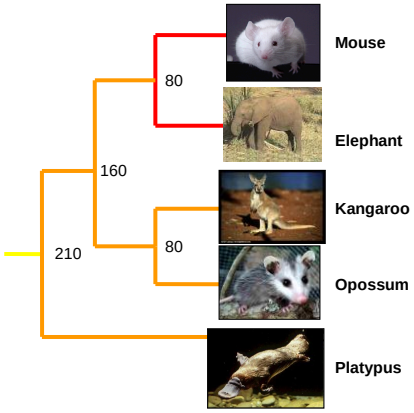


Figure 19. Detail alignment of Oct4 DNA binding domain

Yellow denotes identical bases, blue and green are similar bases. White sections are dissimilar and dash sections cannot be aligned (such as insertions and deletions). Red box highlights non-eutherian mammals. Arrows point to eutherian-specific changes. Species abbreviations: Oa = *Ornithorhynchus anatinus* (Platypus), Md = *Monodelphis domestica* (Opossum), Me = *Macropus eugenii* (Kangaroo), Dn = *Dasypus novemcinctus* (Armadillo), La = *Loxodonta africana* (Elephant), Cf = *Canis familiaris* (Dog), Bt = *Bos taurus* (Cow), Ss = *Sus scrofa* (Pig), Ec = *Equus caballus* (Horse), Fc = *Felis catus* (Cat), Hs = *Homo sapiens* (Human), Rn = *Rattus norvegicus* (Rat) and Mm = *Mus musculus* (Mouse).

From this identity table, it appears that the entire Oct4 DBD is under eutherian-specific selection pressure (Fig. 20). Eutherians (red numerals) are highly similar to each other, whereas non-eutherians (green numerals) are highly similar to each other. If changes in this region were neutral the metatherian sequences should be more similar to the

eutherians than to the prototherians as they last shared a common ancestor more recently. The changes that led to the 12 eutherian-specific amino acids presumably appeared in the eutherian line sometime between 160 and 80 million years ago at the base of eutherian mammal evolution. Since Oct4 is essential for the formation of the inner cell mass, the appearance of these changes correlates with the emergence of the inner cell mass. The location of these changes, within the DNA binding domain, suggests they may affect Oct4 DNA binding specificity and its interaction with other protein partners in the pluripotency transcriptional network. I was intrigued with the possibility that these eutherian-specific features of Oct4 coding sequence may play some eutherian-specific, possibly pluripotency-related function.



	Mouse	Human	Elephant	Kangaroo	Opossum	Platypus
Mouse		91	88	75	76	74
Human			92	78	79	76
Elephant				77	78	76
Kangaroo					98	91
Opossum						90
Platypus						

Figure 20. Sequence identity of the Oct4 DBD
Red figures = comparisons among eutherians, green = comparison among non-eutherians.

One obvious interacting partner that may be influenced by these eutherian-specific changes in the Oct4 DNA binding domain is Sox2. In 2003, Reményi and colleagues solved the crystal structure of the Oct1-Sox2-DNA ternary complex (Reményi et al.

2003). I used their structure to see where these eutherian-specific changes were positioned with respect to Sox2 and its interaction with DNA. By doing a sequence alignment of Oct1 and Oct4, I mapped the eutherian-specific amino acid positions onto this crystal structure, and then displayed it using RasMOL. Here Oct1 is shown in yellow, Sox2 in orange and DNA in blue, while eutherian-specific amino acid positions are highlighted in red (Fig. 21). Some of these may be important for DNA binding specific, and some of these may affect Oct-Sox protein-protein interaction.

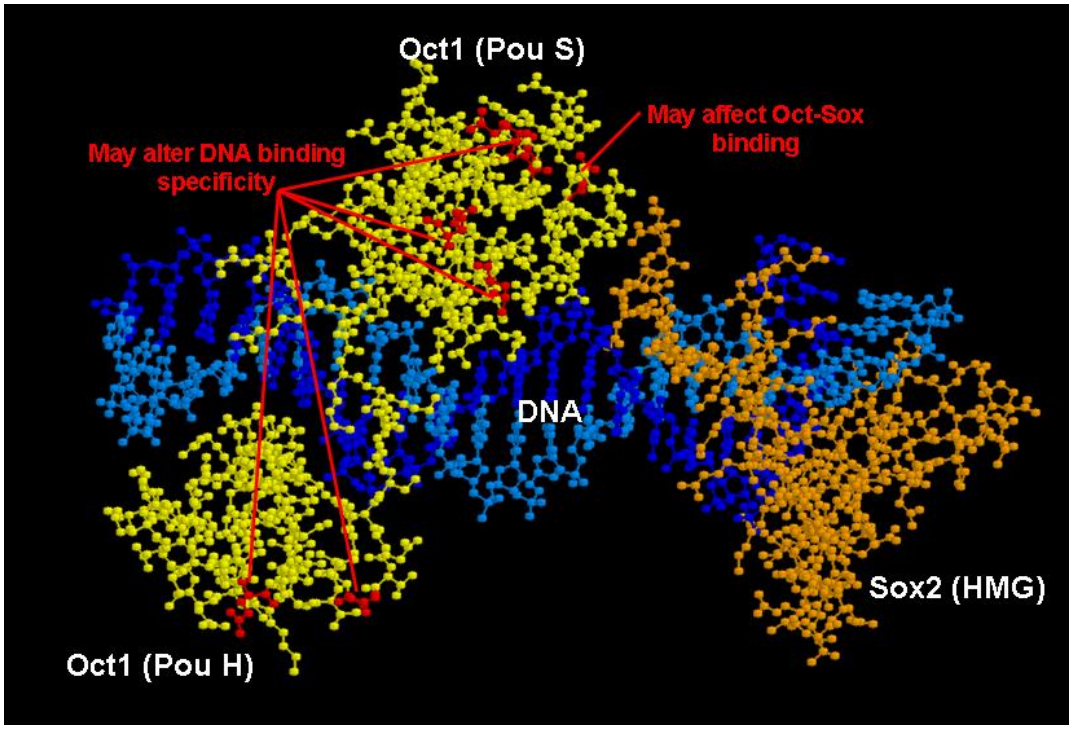


Figure 21. Eutherian-specific changes in Oct4 mapped onto Oct1 crystal structure Oct1 is shown in yellow, Sox2 in orange and DNA in blue. Red denotes amino acid positions mapped from Oct4.

As mentioned earlier, the POU domain of Oct4 is crucial because it is postulated to have protein-protein interactions with the HMG box of Sox2 that allow them to bind together and form a protein complex. A solution structure has been solved for Sox2 and Oct1, which was applied to Oct4 using homology modeling (Williams et al. 2004). It can be

immediately seen that the region of Sox2 HMG involved in this interaction is completely identical from fish to mouse. For Oct4, I have filled in the blanks with sequence information of the opossum and the platypus from my BAC sequencing efforts (Fig. 22).

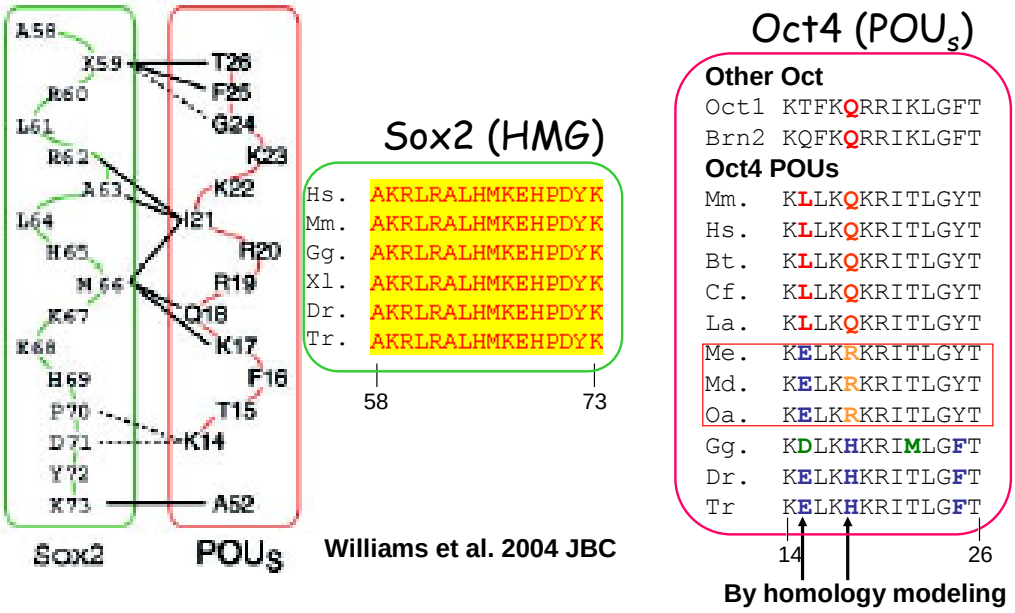


Figure 22. Comparison of amino acid variation in the Sox-Oct interface region Red box highlights non-eutherian mammals. Arrows point to key differences between fish and mouse. Species abbreviations: Tr = *Takifugu rubripes* (Fugu), Dr = *Danio rerio* (Zebrafish), Xl = *Xenopus laevis* (Frog), Gg = *Gallus gallus* (Chicken), Oa = *Ornithorhynchus anatinus* (Platypus), Md = *Monodelphis domestica* (Opossum), Me = *Macropus eugenii* (Kangaroo), La = *Loxodonta africana* (Elephant), Cf = *Canis familiaris* (Dog), Bt = *Bos taurus* (Cow), Hs = *Homo sapiens* (Human) and Mm = *Mus musculus* (Mouse).

Here, there are two key amino acid differences between fish and mouse which may be crucial in Sox2-Oct4 binding – a glutamate (acidic) to lysine (basic) change and a histidine (basic) to glutamine (polar) – indicated with small arrows. The eutherian-specific glutamine 18 residue of Oct4 is also conserved in mouse Oct1 and Brn2, two other octamer proteins that are known to interact with Sox2. By zooming in on the

structure, you can see that glutamine 18 (highlighted in red) is positioned at the Oct-Sox interface. Thus, among all the other eutherian-specific changes, an amino acid substitution in this position appears to have the greatest potential to affect protein binding (Fig. 23).

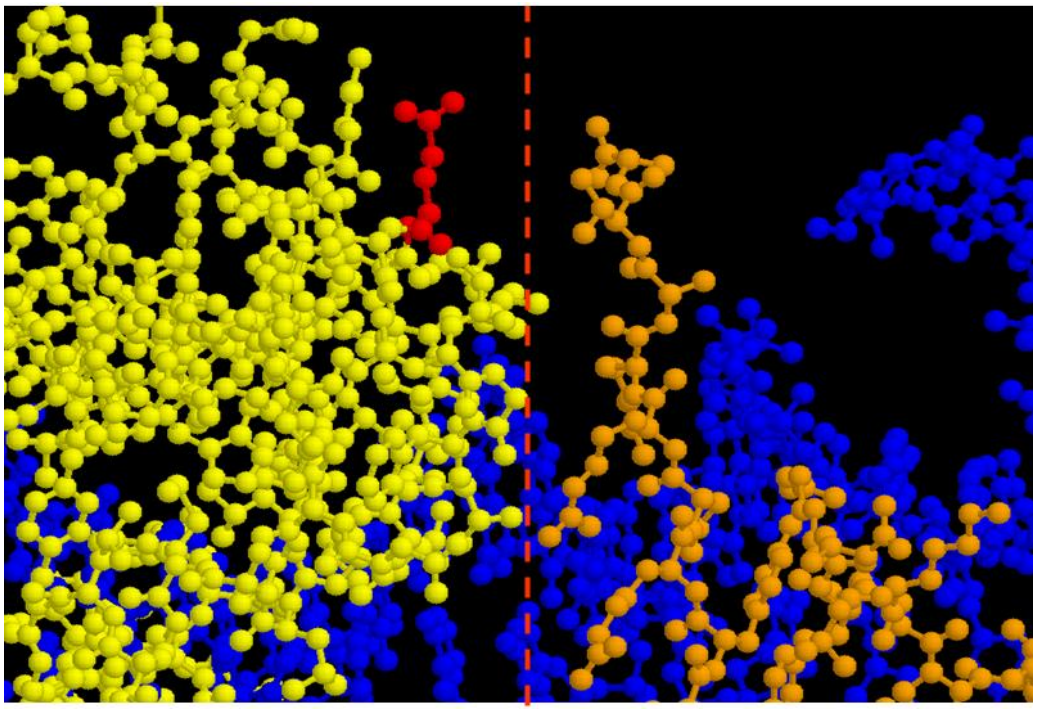


Figure 23. Position of Glutamine 18 is near the Oct-Sox interface
Oct1 is shown in yellow, Sox2 in orange and DNA in blue. Glutamine 18 position shown in red. The dashed line approximates the interface between Oct and Sox proteins.

With a number of eutherian-specific changes found, the next step was to prioritize and select some of the most salient changes that have a good chance of playing a direct and significant role in pluripotency. These changes were then functionalized using mouse sequences as the raw material and then using molecular techniques to revert them into non-eutherian sequences in order to reveal any interesting cell-level phenotype through a loss-of-function approach. The details of these experiments are discussed in the next chapter.

Chapter 4: Functionalization At Cell Level

4.1 Overview

The pluripotent transcriptional network is likely to involve a large number of genes. Within the Oct-Sox-Nanog core, Nanog alone already has dozens of direct protein partners (Wang et al. 2006). Moreover all three transcription factors have pleiotropic effects – both Oct4 and Nanog are also involved in germ cell development, whereas Sox2 is also involved in neuronal differentiation. As such it is beyond the scope of this project to functionalize every interesting eutherian-specific change in this network. In order to avoid chancing upon changes that are not involved in pluripotency, the focus is to select some of the most promising ones that lie at the heart of the network – more precisely, molecular changes that allow Sox2 and Oct4 to work together.

These include both cis-element changes that allow the Sox-Oct complex to target a specific gene, and coding sequence changes that allow the formation of the Sox-Oct complex in the first place. Based on sequence analysis the following three directions were considered the highest priority to be pursued:

1. Promoter assays of Sox-Oct element changes.
2. Oct4 DBD-activator/repressor fusion protein experiments.
3. Oct4 Full-length chimera induced pluripotency experiments.

There were a few other experiments that also have a good potential to yield measurable results but are not as high in priority or are plagued by technical difficulties.

4.2 Sox-oct Element Materials and Methods

The Sox-Oct element identified in the *Sox2*, *Nanog* and *Pou5f1* genes are conserved, but with a few base pair differences when compared across a number of eutherian and non-eutherian species. These specific base differences may affect its function as the binding site for the Sox2-Oct4 heterodimer. To investigate this, the promoter regions of mouse *Sox2*, *Nanog* (Fig. 24) and *Pou5f1* are subcloned into a pGL3 Basic luciferase reporter (Promega). The *Nanog* promoter is obtained by PCR from genomic DNA, the *Pou5f1* promoter obtained by long PCR (Roche Expand) from a BAC clone provided by Dr. Thomas Lufkin's lab, and the *Sox2* promoter is from Dr. Ng Huck Hui's lab.

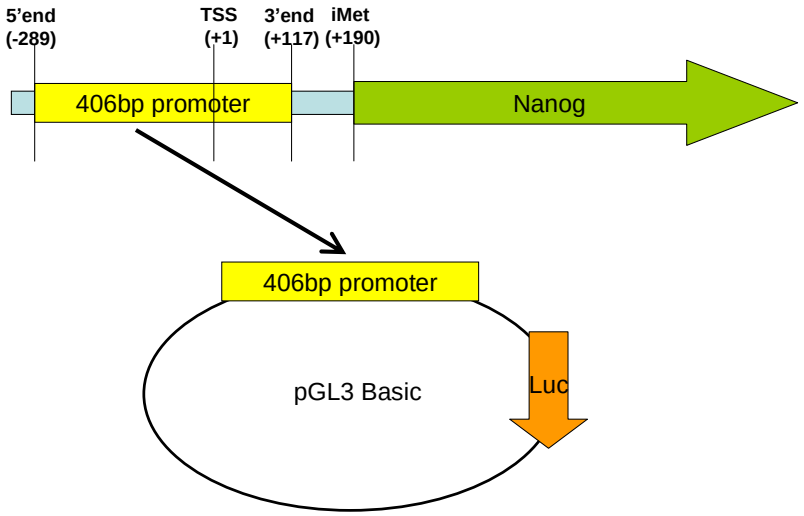


Figure 24. Nanog promoter subcloning

Next the Sox-Oct elements are modified by mutagenesis (Clontech Transformer / Stratagene QuikChange II) to resemble homologous regions in a number of species (Fig. 25-27). Luciferase assays are performed in the F9 EC culture system, to see if there is any significant reduction in reporter activity after these changes.

```
              Sox2
Mm.           CATTGTGATGCATAT
Hs.           CATTGTAATGCATAT
Md.           TTTTGTAATGCATAT
Gga.          GTTTGTAATGCATCT
Abolish       CAGGTTGATTTGTAT
              ⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣
                sox      oct
```

Figure 25. Point mutations on the Sox2 sox-oct element
Each colour denotes one successive round of site-directed mutagenesis. Green denotes a drastic change to abolish sox-oct binding to the element, serving as a negative control. Species abbreviations: Gga = *Gallus gallus* (Chicken), Md = *Monodelphis domestica* (Opossum), Hs = *Homo sapiens* (Human), Mm = *Mus musculus* (Mouse).

```
              Nanog
Mm.           CATTGTAATGCAAAA
Rn.           CATTGTAATGCAAAA
Hs.           CATTGTAATGCAAAA
Bt.           CATTGTAATGCAAAA
La.           CATTGTAATGCAAAA
Abolish       CAGGTTAATTTGAAA
              ⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣⌣
                sox      oct
```

Figure 26. Point mutations on the Nanog sox-oct element
Species abbreviations: La = *Loxodonta africana* (Elephant), Bt = *Bos taurus* (Cow), Hs = *Homo sapiens* (Human), Rn = *Rattus norvegicus* (Rat), Mm = *Mus musculus* (Mouse).

```
                Oct4
Mm.             CTTTGTTATGCATCT
Hs.             CTTTGTCATGCATCT
La.             TTTTGTCATGCACCT
Abolish         CTGGTTTATTTGTCT
                 sox      oct
```

Figure 27. Point mutations on the *Pou5f1* sox-oct element
Each colour denotes one successive round of site-directed mutagenesis. Green denotes a drastic change to abolish sox-oct binding to the element, serving as a negative control. Species abbreviations: La = *Loxodonta africana* (Elephant), Hs = *Homo sapiens* (Human), Mm = *Mus musculus* (Mouse).

4.3 Sox-oct Element Results and Discussion

The mutation constructs are then transfected into F9 teratocarcinoma cells, co-transfected with the renilla luciferase control plasmid (Promega Dual Luciferase Assay). Luciferase levels were read using the Centro LB960 luminometer (Berthold Technologies), normalised and then expressed as a percentage of wild-type levels. These assays were performed in triplicate.

In *Sox2*, the point mutations to change the mouse element to the chick sequence produced no significant reduction in reporter activity, suggesting that the element is functionally active in chick (Fig. 28). This was initially a surprise when the result was obtained because there was a reported absence of chick Oct4, thus I postulated that the sox-oct element could serve as a binding site for other protein complexes such as Brn1-Sox2. With the discovery of chick Oct4 (Lavial et al. 2007) it is more likely that this element is already used for Oct4-Sox2 binding in the autoregulation of *Sox2*.

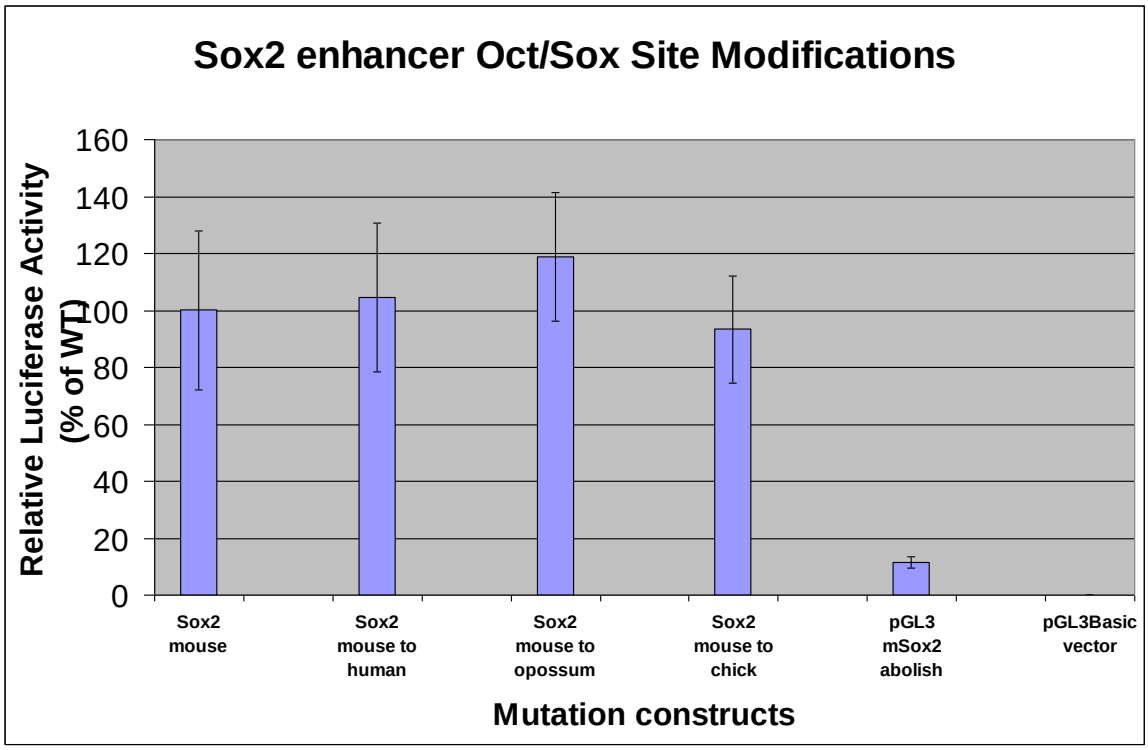


Figure 28. Sox2 promoter assay results.

Nanog promoter assays demonstrate the strong effect of point mutations on the activity of the Oct-Sox site. A three nucleotide change to either the Oct or Sox element is enough to reduce the luciferase activity to less than 20% of wildtype levels. A 6 base pair mutation to both reduces the luciferase activity further to less than 10% (Fig. 29). This provides strong evidence for the functional importance of this sox-oct element, which resides in the proximal promoter, to drive pluripotent expression of *Nanog*. The presumption then, is that this element is required to drive pluripotent expression of *Nanog* in all eutherian mammals as its location in the proximal promoter and its sequence is conserved in all eutherian mammals analyzed including the most distal, the elephant. It was then intriguing that there was no evidence for this sox-oct element in the metatherian and prototherian mammals nor in the chick *Nanog*.

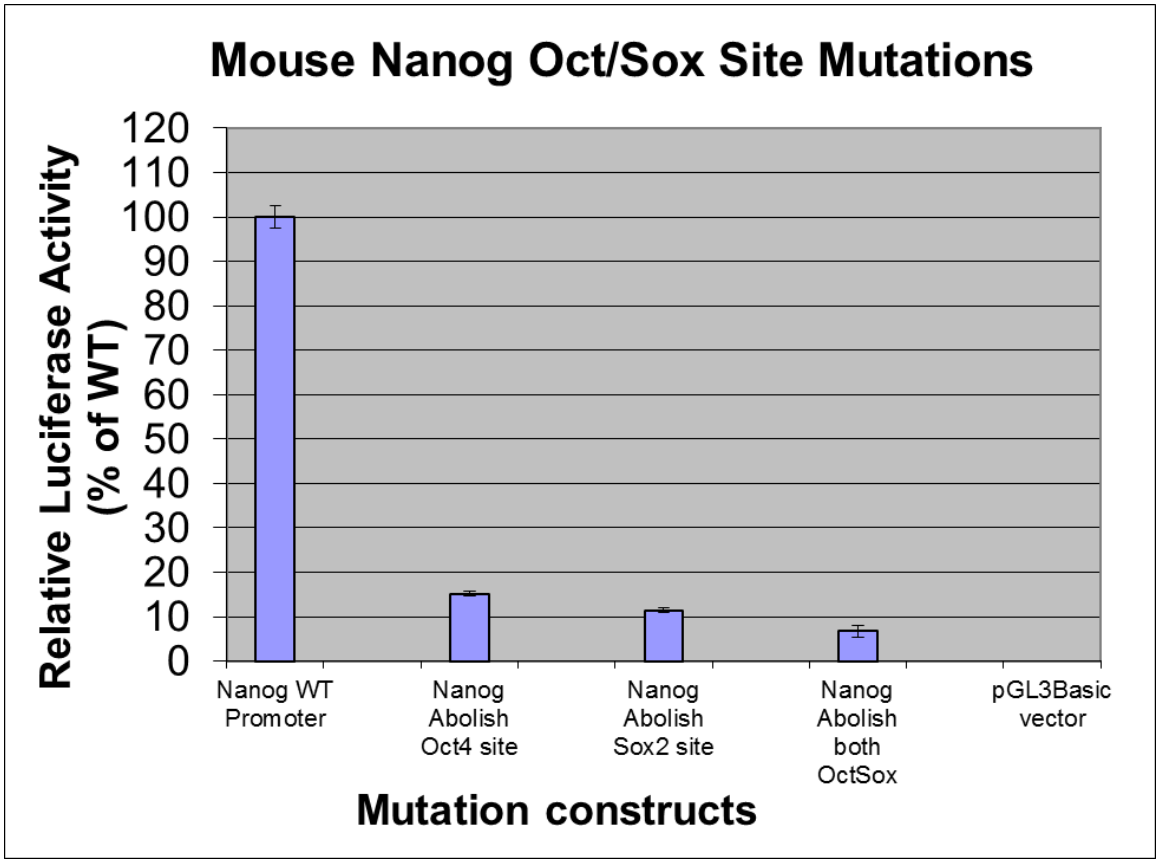


Figure 29. Nanog promoter assay results.

Luciferase assays in F9 EC cells also indicated a requirement for the sox-oct element within the *Pou5f1* enhancer to drive maximal expression, a six base pair mutation resulting in the disruption of both the octamer and sox elements reduced promoter activity to less than 40% of its wildtype levels (Fig. 30). That said, this was not as great a drop as that seen for the similarly designed *Nanog* promoter perhaps suggesting the sox-oct element plays a greater role in the expression of *Nanog* then it does of *Pou5f1*. With this considerations it is interesting to note that the sox-oct element within the *Pou5f1* enhancer was less conserved between eutherians than was the *Nanog* element (compare Figures 26 & 27). There was no available sequence information on the cis-regions of the opossum and platypus, and thus no out-group sequences to compare with.

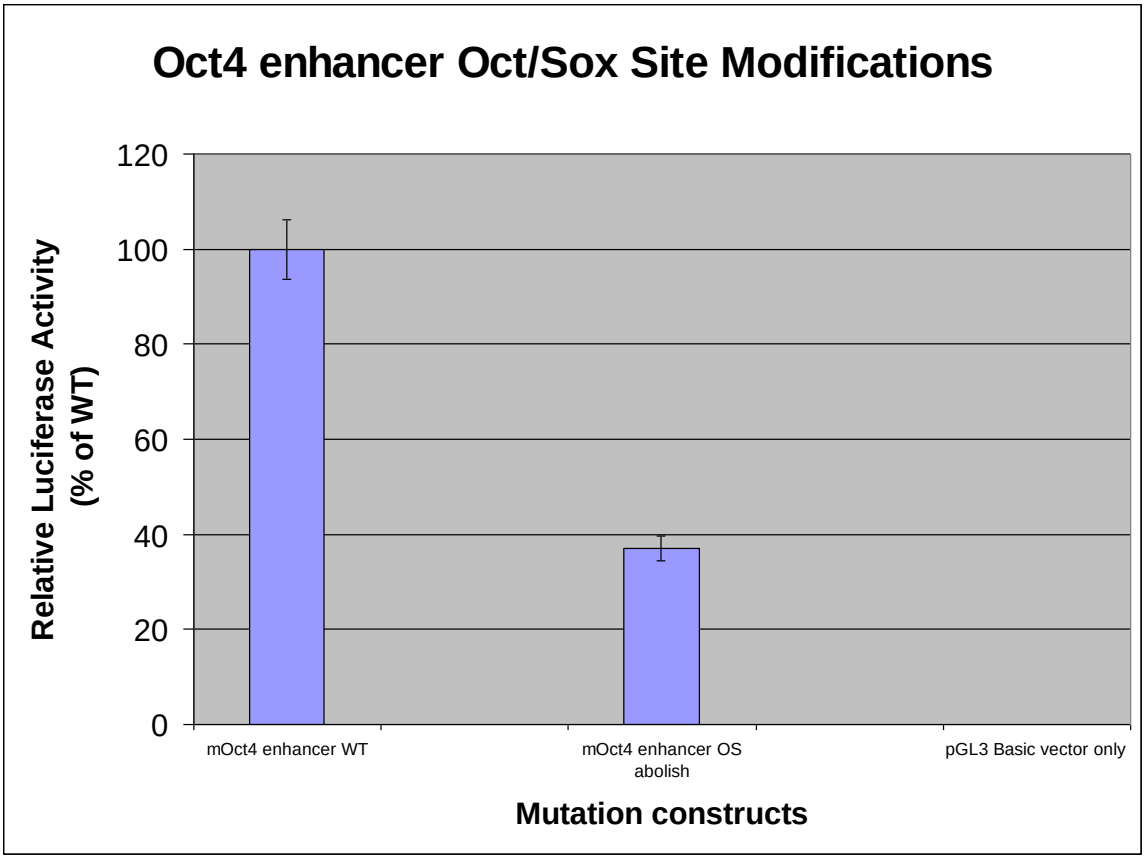


Figure 30. *Pou5f1* promoter assay results.

4.4 VP16/EnR Fusion Materials and Methods

As I detected 12 amino acid residues within the DNA binding domain of Oct4 to be eutherian-specific I next sought to investigate the importance of these changes on the ability of Oct4 to bind downstream targets. I chose first to use a strategy which involved fusing the Oct4 DNA binding domain of various relevant species to either a strong activator or repressor of transcription. The system used to test this is the VP16 activator/ EnR repressor system which is used extensively in developmental biology (Carsona et al. 2004). The principle behind this is that genes normally repressed by a transcription factor of interest would remain repressed with the EnR fusion protein but activated by the VP16

fusion protein. Thus, this is a brute force method to elicit the strongest possible response. A number of expression constructs from relevant mammal species have been made, transfected into ES cells and the changes in global gene expression studied using real time PCR and Illumina BeadArray analyses.

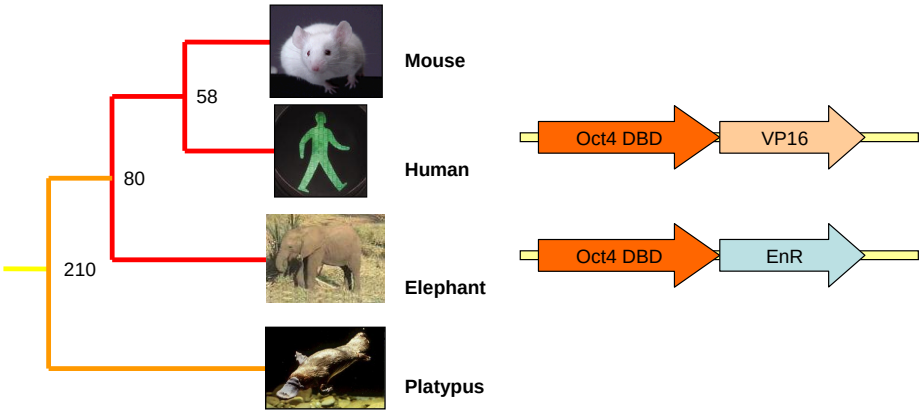


Figure 31. Oct4 DNA binding domain constructs

The strategy was to study four species: mouse, human, elephant and platypus (Fig. 31). Mouse and human data is useful for optimizing the protocol and checking expected Oct4 downstream targets. Elephant data is important because it is the most distant eutherian to the mouse. I postulated that platypus Oct4 DBD should target different genes compared to eutherian mammals, potentially through different binding partners or DNA recognition sites, as a result of having different amino acid residues at the 12 positions previously identified to be eutherian-specific.

The data can also be used to generate more interesting results. By comparing the pan-mammalian DBD targets with eutherian-specific targets, I can check if we have binding

site data for these genes, and find out how the platypus Oct4 binding elements differ from eutherian-specific elements (Fig. 32).

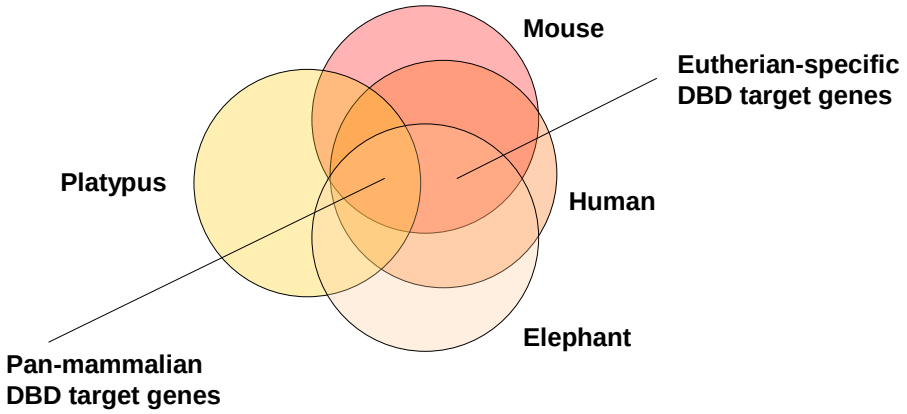


Figure 32. Discover eutherian-specific functions of Oct4

The preparatory work for this set of experiments can be divided into two main parts: (1) cloning the Oct4 DBD constructs and (2) optimizing the ES cell culture conditions.

1. The mammalian Oct4 DBD spans 4 exons (2-5) and thus making this part of the construct should necessitate a multistep PCR cloning strategy (Fig. 33) as I did not have access to mRNA (or cDNA) of elephant and platypus material, only genomic DNA. However, in the interest of time and to avoid PCR errors due to the numerous PCR steps, in the case of Platypus and Elephant Oct4 DBD, the whole DBD was synthesized *de novo* (Codon Devices) flanked with suitable restriction sites on both ends.

For the mouse and human Oct4 DBD, the process is more straightforward since the fragment can simply be obtained by PCR in one step from the cDNA stocks available in the lab.

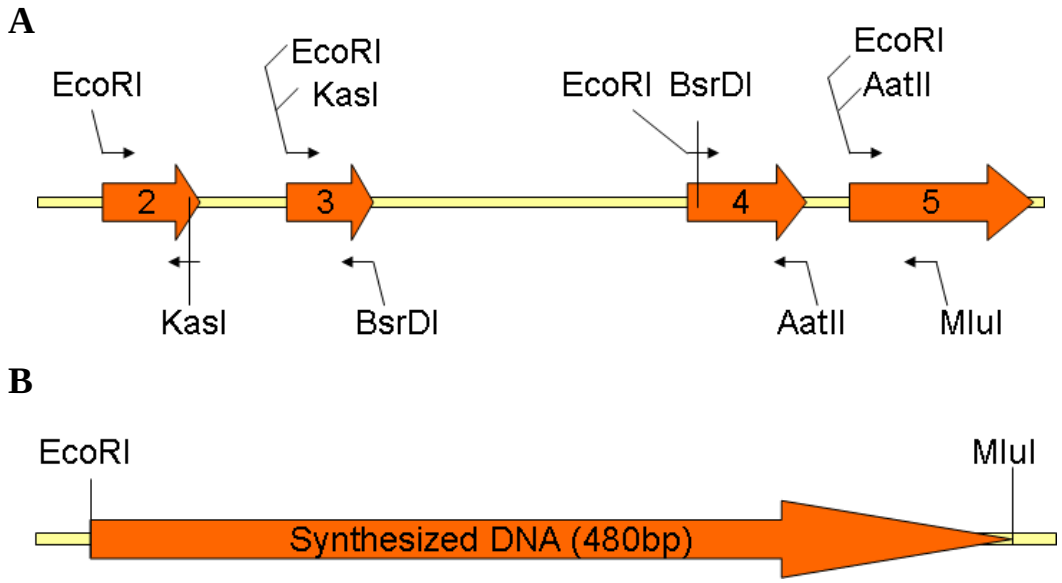


Figure 33. Cloning strategy for Platypus Oct4 DBD
(A) Initial strategy involving multistep PCR
(B) Actual strategy using synthesized DNA

The VP16 and EnR fragments were separately obtained by PCR from their respective plasmids and cloned together with the Oct4 DBD fragments into the CAG-pIRES-EGFP expression vector. A short, two amino acid long linker (MluI = Thr-Arg) connects the DBD to the VP16 or EnR to form a fusion protein (Fig. 34).

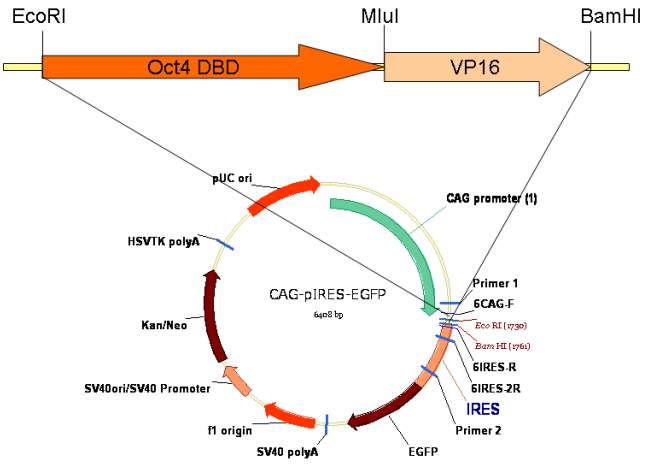


Figure 34. Mammalian Oct4 DBD VP16 expression construct

This process was repeated in the same way for all four species (Fig. 35) and the constructs were verified by DNA sequencing.

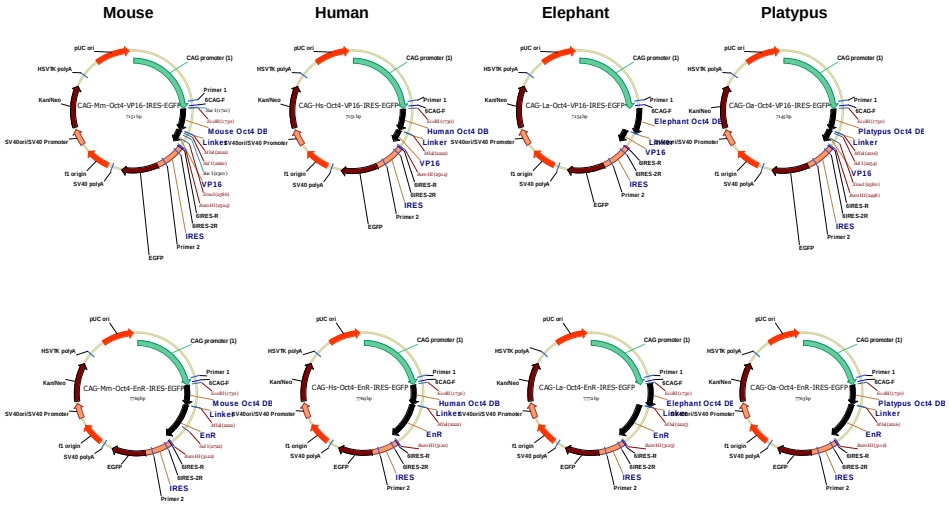


Figure 35. Eight constructs made for the Oct4 DBD fusion experiments

2. The completed constructs were then transfected (Lipofectamine 2000) using a suspension transfection protocol in E14 ES cells to optimize the cell culture conditions in

a 6-well plate. By tweaking the conditions to minimize cell death, I was able to optimize a reliable set of conditions for subsequent experiments (Table 9).

ES cell type	E14 passage 31+
Medium	Standard ES medium
Volume of Lipofectamine	10 ul
Amount of DNA	4 ug
Volume of cells seeded	500 ul (1.6 X of protocol)
Time point for cell harvest	24h

Table 9. Optimized E14 culture conditions

In the first set of experiments, the transfected cells were examined at 24h post-transfection for robust proliferation and GFP expression (Fig. 36). Next, the cells were collected and protein was extracted for Western blot verification.

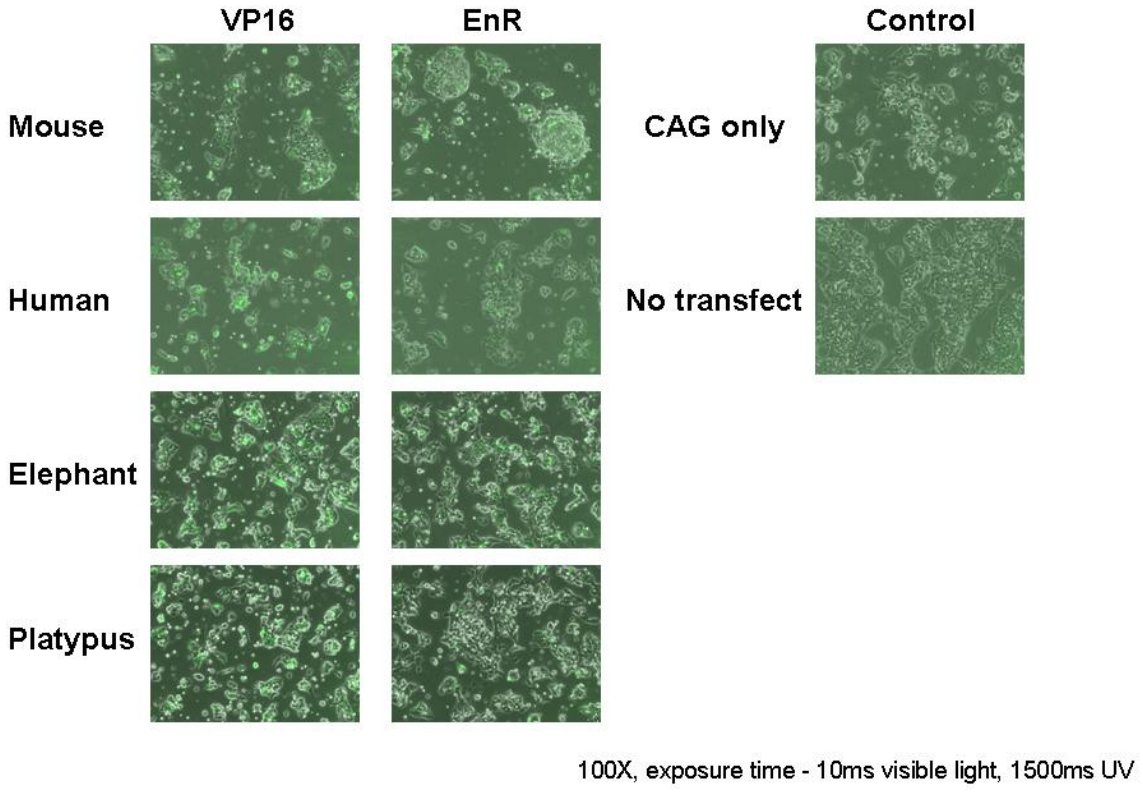


Figure 36. E14 (p33) transfections at the 24h time point

To check that the Oct4 DBD fusion proteins were expressed in the ES cells in their entirety, Western blot analysis was performed using antibodies to VP16 (total DBD-VP16 size of 29 kDa) and EnR (total DBD-EnR size of 51 kDa). The results indicate fusion proteins of correct size (Fig. 37 and 38).

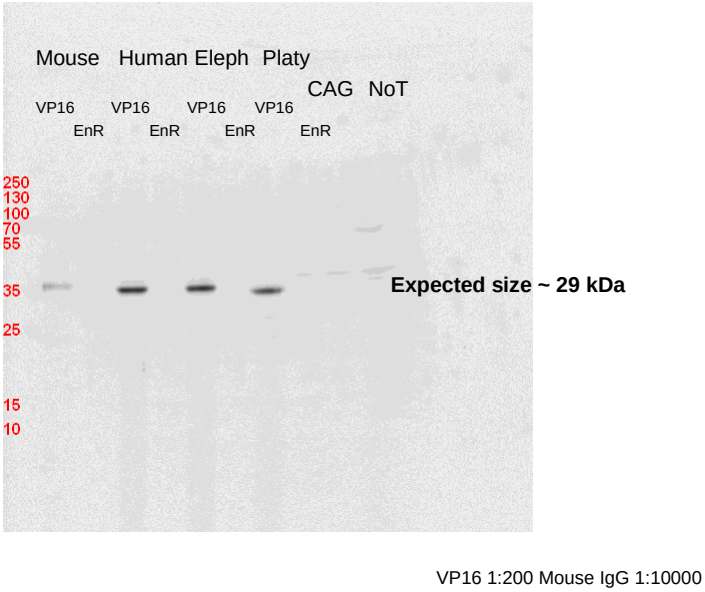


Figure 37. Western blot verification using VP16 antibody
CAG = CAG vector only, NoT = No transfection control

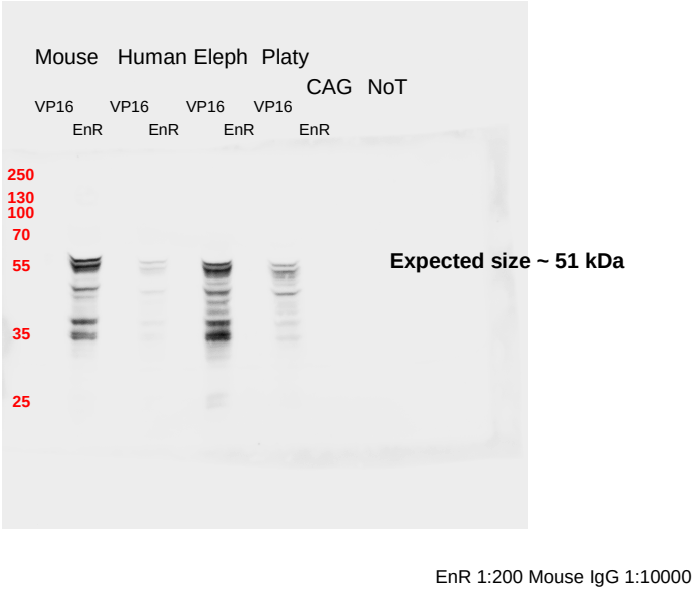


Figure 38. Western blot verification using EnR antibody
CAG = CAG vector only, NoT = No transfection control

4.5 VP16/EnR Fusion Results and Discussion

Next, an experiment was done in triplicate and RNA was extracted from the transfected ES cells to be used for real time PCR analysis. Details of the real time PCR (BioMark) protocol are in Appendix C.

This real time PCR format can analyze the expression of up to 48 genes in a single run, thus a selection of 48 real time probes of genes relevant to pluripotency and early embryo development were analyzed (Table 10).

Gene name	Function
Actb	Normalization control
Ascl1 (Mash1)	Neural development
Bmp4	Bone and muscle development
Cdh1	Cell-cell adhesion, tumour suppressor
Cdx2	Placental development
Dll1	Haematopoiesis
EGFP	Fluorescent marker
Elavl3	Neural development
Eomes	Trophoblast development
Esrrb	Placental development
Fbox15	Pluripotency marker
Fgf4	Cell proliferation, oct-sox target, bone development
Fgfr2	Fgf receptor, bone development
Gadd45g	Placental marker, tumour suppressor
Gata3	Mesoderm differentiation, T lymphocyte development
Gata6	Endoderm differentiation
Gdf3-exon1	Mesendoderm development
Hand1	Trophoblast development, heart development
Hes6	Neuronal differentiation
Irx3	Neural development
Klf2	Pluripotency maintenance
Klf4	Pluripotency maintenance
Klf5	Pluripotency maintenance
Lfng	Mesoderm development, Notch signaling
Mfng	Mesoderm development, Notch signaling
Nanog	Pluripotency, germ cell development
Nes	Neural development
Nrarp	Blood vessel formation, Notch signaling
Pax6	Eye development, neural development
Pecam	Inner cell mass marker, endothelial marker
Pou3f1	Neural development
Pou3f2 (Brn2)	Neural development
Pou5f1	Pluripotency, germ cell development
Rax	Eye development
Rest	Neural development
Rhbdl3	Membrane protein, signal transduction
Sall4	Pluripotency
Sox11	Neural development
Sox15	Placental marker, skeletal muscle regeneration
Sox17	Endoderm formation
Sox2	Pluripotency, neural development, tumourigenic
Sox21	Neural development, hair formation
Sox3	Neural development
Sox4	Apoptosis
Sox7	Endoderm formation
Tubb3	Microtubule component, control
Utf1	Pluripotency marker
Zfp42 (Rex1)	Pluripotency marker

Table 10. Real time PCR probes and some of the gene functions

The real time PCR raw data is first displayed as a heat map (Fig. 39) and then the data is processed to show gene expression differences as fold change on a bar chart. Genes that have increased expression level in response to the Oct4 DBD VP16 fusion protein and corresponding decreased expression in response to the Oct4 DBD EnR fusion protein are considered to be strong direct targets of Oct4 (Fig. 40).

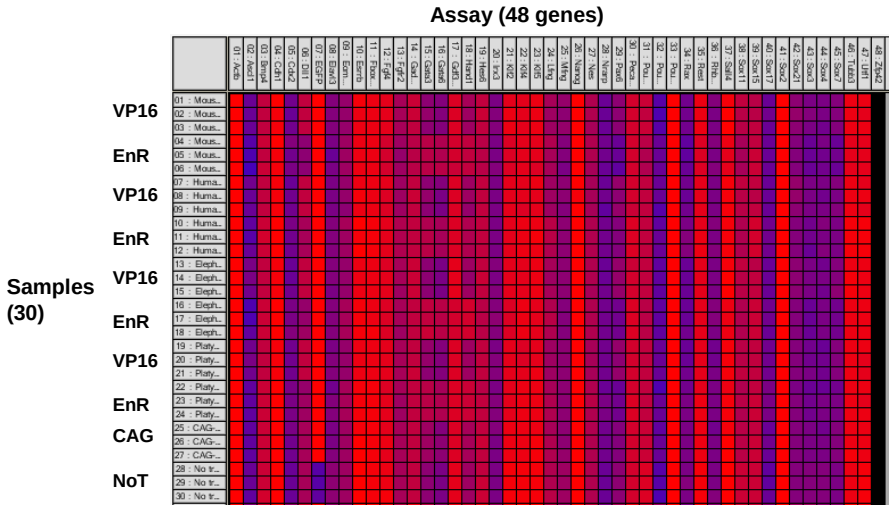


Figure 39. BioMark Real Time PCR – Raw Data (Heat Map)
Red = high expression, Blue = low expression

Meanwhile, genes with the converse response are considered to be strong indirect targets, while genes that display unidirectional response to both VP16 and EnR are likely to be abnormally regulated due to the aggressive treatment in this experimental strategy. Genes with abnormal regulation responses are not such interesting candidates compared to the strong direct or indirect targets.

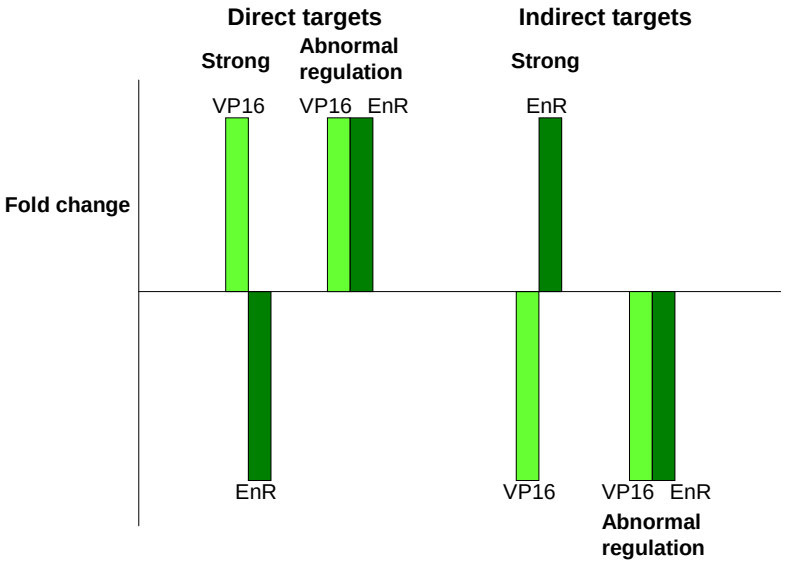


Figure 40. How to interprete the real time PCR results

In the first set of experiments, the gene expression response of known Oct4 targets involved in pluripotency are shown in Figure 41. As can be clearly seen, the levels of Oct4 itself appear to be very high for both Mouse Oct4 VP16 and EnR fusion experiments, because the real time PCR probes are designed to Mouse Oct4 and cannot distinguish between endogenous Oct4 and the transfected Oct4 fusions proteins. For the other species, Fgf4 and Sox2 respond normally while other Oct4 targets display abnormal regulation. Contrary to expectations, there is no qualitative difference in response between the platypus Oct4 DBD fusions and the eutherian Oct4 DBD fusions.

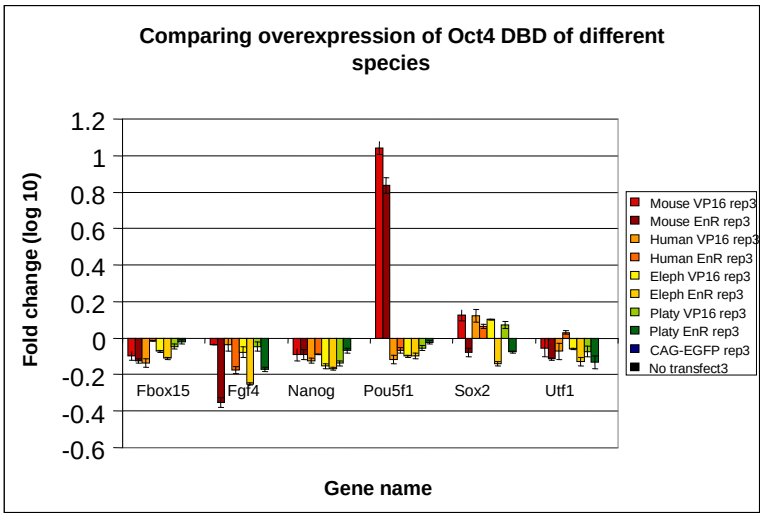


Figure 41. Real time PCR results of pluripotency-related genes

The strongest responses are not from genes involved in pluripotency, but those involved in other aspects of early embryo development. Mash1, Dll1 and Pax6 present as strong direct targets of all the Oct4 DBD fusions, while Gata3 and Gata6 are clearly strong indirect targets (Fig. 42). Again, there is no distinct difference in response between platypus and the eutherians; indeed, for Dll1 and Pax6 the quantitative difference is greater between the human and the mouse, compared to between the platypus and the mouse. The direction of the response is practically identical for all the species in this experiment.

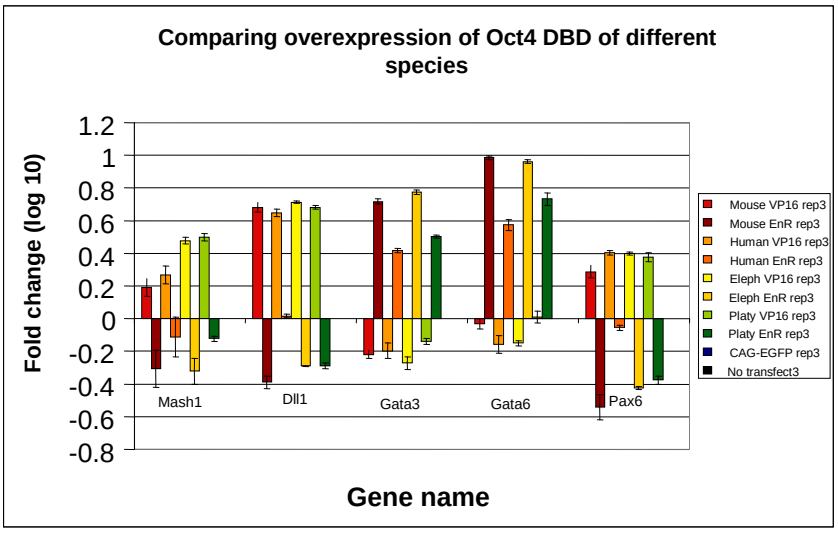


Figure 42. Real time PCR results of genes with strongest response

Other results also indicate the similarity of response direction and magnitude (Fig. 43) between the platypus and the eutherians. An unexpected but consistent finding is the difference in response between the mouse and the human, exemplified here by Cdx2 and Hand1.

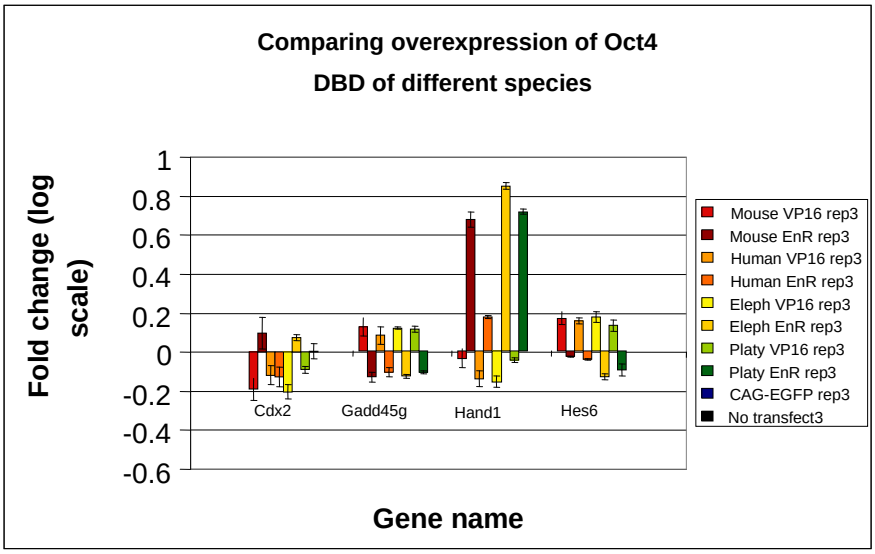


Figure 43. Real time PCR results of other genes with normal response

Due to the fact that some pluripotency related genes displayed abnormal regulation, I postulated this could be due to excessively high levels of Oct4 competing for binding sites, since endogenous Oct4 expression remained intact in this set of experiments.

To address this, a second set of experiments was done where an Oct4 RNAi vector was co-transfected with the fusion protein vectors in order to knockdown the endogenous Oct4 expression. The results show a normal response for Nanog, suggesting that Nanog may be abnormal regulated when Oct4 levels are too high (Fig. 44), while other genes now appear to be abnormally regulated.

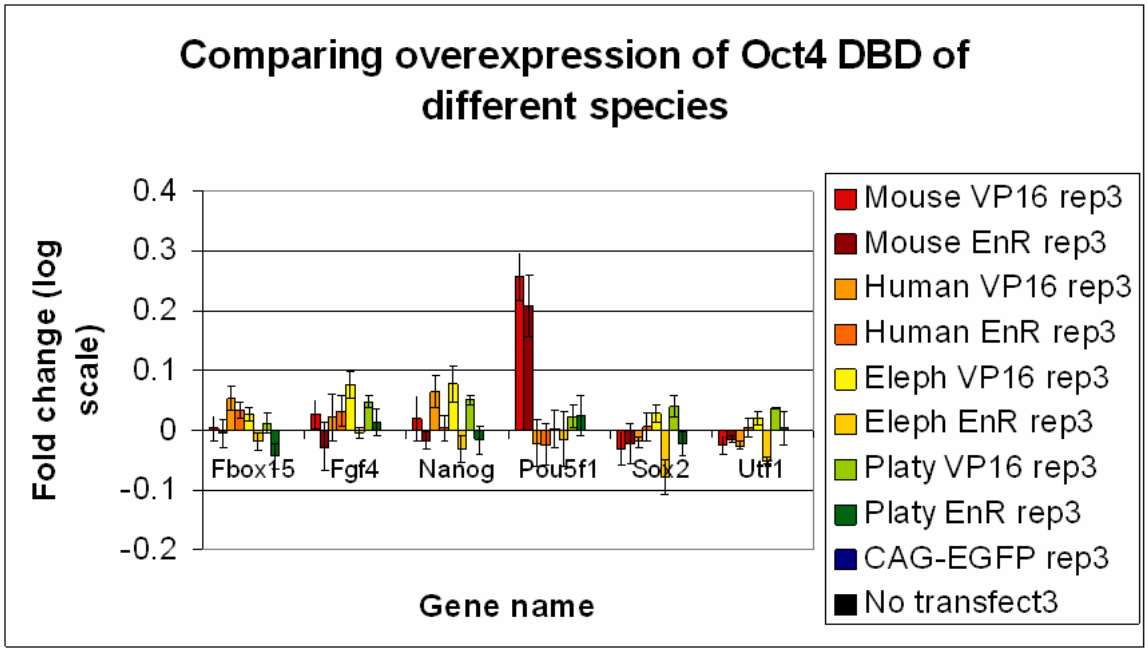


Figure 44. Real time PCR of pluripotency genes with Oct4 RNAi co-transfection.

When the result for all 48 genes was analyzed, once again there was no qualitative difference between the response to platypus and eutherian fusions, just like in the first experiment. Thus there is a strong possibility that interesting gene expression changes lie outside this selection of 48 genes.

To broaden the search for potential targets differentially regulated by platypus Oct4 DBD compared to the eutherian Oct4 DBDs, I utilized mouse Illumina BeadArrays to achieve a more global readout of gene expression changes resulting from the above over-expression experiments. An initial cut-off of two fold change or less from normalized levels yielded only a handful of genes that have expression level differences, suggesting that the responses are highly similar for all four Oct4 DBD fusions.

In order to avoid missing any subtle change, a lower cut-off of 1.5 fold or less was used, and the complete list of these genes is available in Appendix D. This list of several hundred genes was then manually analyzed to shortlist those genes that have directionally different responses between the platypus and the eutherian group. Only six genes fulfill this condition, and none have been previously implicated in playing a role in pluripotency (Table 11). Three of these genes have unknown function, whereas two (Nlrp3 and Irf1) play a role in the immune response. It is interesting to speculate, as they are downregulated in eutherians in contrast to the platypus, that the down-regulation of these two immune response genes is functionally related to the requirement for the maternal immune system to be suppressed upon eutherian embryo implantation..

Gene name	Function	Expression Level Change
2010002N04Rik (Nid67)	Small membrane protein (unknown function)	Upregulated in eutherians, unchanged in platypus
BC055811 (Igsf21)	Immunoglobin, extracellular (unknown function)	Downregulated in eutherians, unchanged in platypus
Nlrp3 (Cias1)	Apoptosis, inflammatory response, NALP3 inflammasome complex	Downregulated in eutherians, unchanged in platypus
Fbxw5	Ubiquitin cycle (unknown function)	Downregulated in eutherians, unchanged in platypus
Herpud1	Unfolded protein response, stress response	Downregulated in eutherians, unchanged in platypus
Irf1	Transcription factor, inflammatory response, tumour suppression	Downregulated in eutherians, unchanged in platypus

Table 11. Genes with the greatest gene expression difference between Platypus and the eutherian group

4.6 Oct4 Full-length Chimera iPS Materials and Methods

The use of strong modulators like VP16 and EnR produce Oct4 expression levels that far exceed the levels found in an endogenous setting. To create a closer approximation of endogenous conditions, full-length mouse Oct4 chimeras containing the elephant and platypus DBD were constructed (Fig. 45) to be tested in the iPS cell culture system to find out if the platypus chimera will lack the ability to reprogramme mouse fibroblasts into iPS cells, in contrast to mouse Oct4. The elephant was chosen for in-group comparison because it is the most distantly related eutherian relative to the mouse.

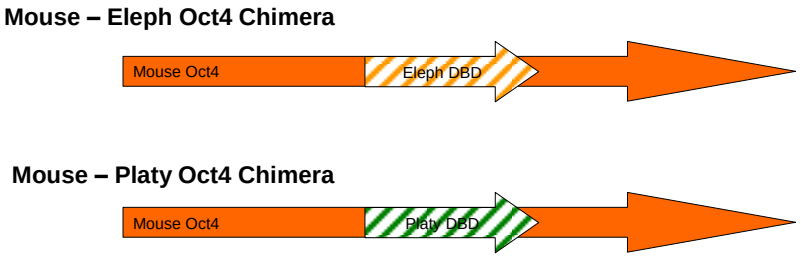


Figure 45. Full-length mouse Oct4 chimeras containing elephant or platypus DBD

The initial plan was to construct the entire chimera at one go using a fusion PCR strategy (Fig. 46). The mouse Oct4 sections (1&3) were obtained by PCR from a full-length mouse Oct4 expression vector, while the elephant and platypus sections (2) were obtained by PCR from the Oct4 DBD fusion constructs made in the previous VP16/EnR experiments.

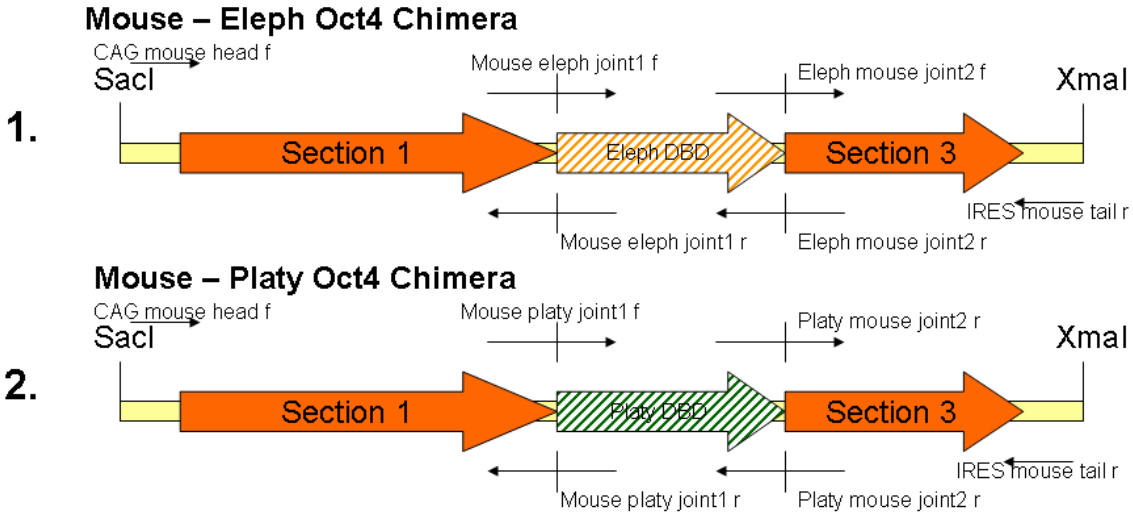


Figure 46. Fusion PCR strategy for construction of Oct4 chimeras

Unfortunately, a PCR product fused from 3 sections could not be obtained in one step. A two-step fusion strategy also did not work as only 2 sections of any combination could be fused together in total (Table 12).

Attempt	Combinations tried	Does it work?
Single PCR	Mouse 1 Eleph 2 Platy 2 Mouse 3	Yes Yes Yes Yes
One-step fusion PCR	Mouse-Eleph 1+2+3 Mouse-Platy 1+2+3	No No
Two-step fusion PCR (a)	Mouse-Eleph 1+2 Mouse-Eleph 2+3 Mouse-Platy 1+2 Mouse-Platy 2+3	**Yes** **Yes** **Yes** **Yes**
(b)	Mouse-Eleph (1,2) + (2,3) Mouse-Eleph (1,2) + 3 Mouse-Eleph 1 + (2,3) Mouse-Platy (1,2) + (2,3) Mouse-Platy (1,2) + 3 Mouse-Platy 1 + (2,3)	No No No No No No

Table 12. Exhausting all fusion PCR permutations to produce Oct4 chimera

A new hybrid strategy was adopted that combined fusion PCR with two additional cloning steps (Fig. 47). These fragments were to be cloned into the pMXs-gw-Oct4 viral expression vector, which already contains full-length wild-type mouse Oct4 (Takahashi and Yamanaka 2006). Although the pMXs vector uses Gateway cloning technology, I decided to select restriction sites within the Oct4 coding sequence in order to keep the sequence between the coding sequence and Gateway clone sites identical, eliminating any functional variability that may result from the cloning process.

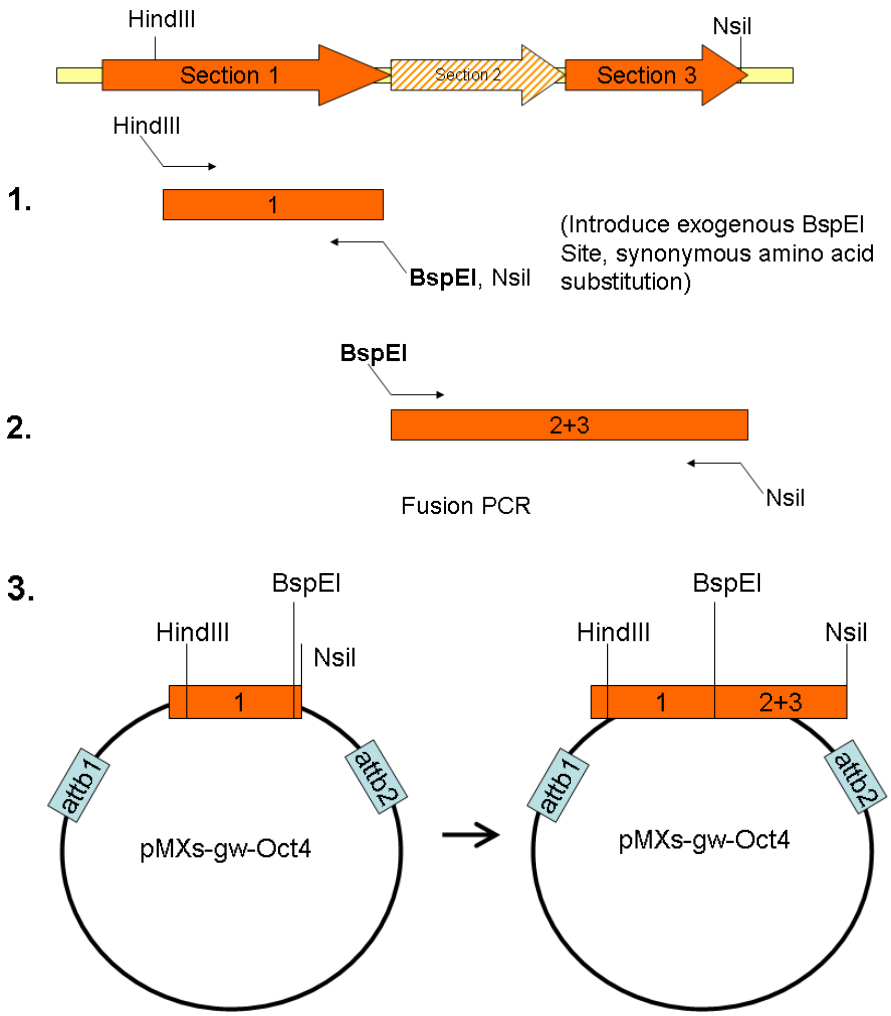


Figure 47. Hybrid PCR-cloning strategy and use of internal RE sites to avoid disturbing Gateway clone sites

The modified plasmids were sequence verified to be error-free. Then, they were transfected into Platinum-E cells in 10-cm dishes for retroviral production. In total seven plates of Plat-E cells containing different viral vectors were prepared: Yamanaka's mouse Oct4, Sox2, Klf4, c-Myc, and an empty pMX vector and my two chimeric platypus DBD-mouse Oct4 and elephant DBD-mouse Oct4 constructs. Finally the viruses were isolated, concentrated and used to infect BL6 embryonic fibroblast cells plated on 6-cm dishes. For details of the iPS protocol, please refer to Appendix E.

The first set of experiment consisted of six conditions including controls and was done in triplicate (Table 13). In the two experimental conditions, wild-type mouse Oct4 was replaced by either the platypus or elephant Oct4 chimera.

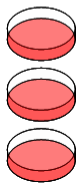
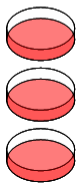
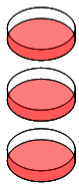
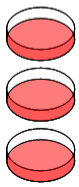
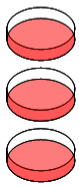
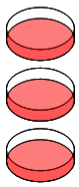
Plate Identity	A Positive control	B Platy	C Eleph	D No Oct4	E Empty vector	F No infect control
Constituents (In 1 ml portions)	Sox2 Klf4 c-Myc mOct4	Sox2 Klf4 c-Myc pOct4	Sox2 Klf4 c-Myc eOct4	Sox2 Klf4 c-Myc Empty vector	Empty vector Empty vector Empty vector Empty vector	FP medium FP medium FP medium FP medium
Triplicate (6 cm dish)						

Table 13. iPS experimental setup

4.7 Oct4 Full-length Chimera iPS Results and Discussion

On the 5th day of the protocol, the viral-laden media on the 6-cm dishes was aspirated away and replaced by fresh ES cell media. Each dish was closely monitored daily for the appearance of induced cell colonies. Some of the early colonies stopped growing a few days after they appeared, while others grew very rapidly. Curiously, well-defined colonies started to appear on the platypus plates (Fig. 48).

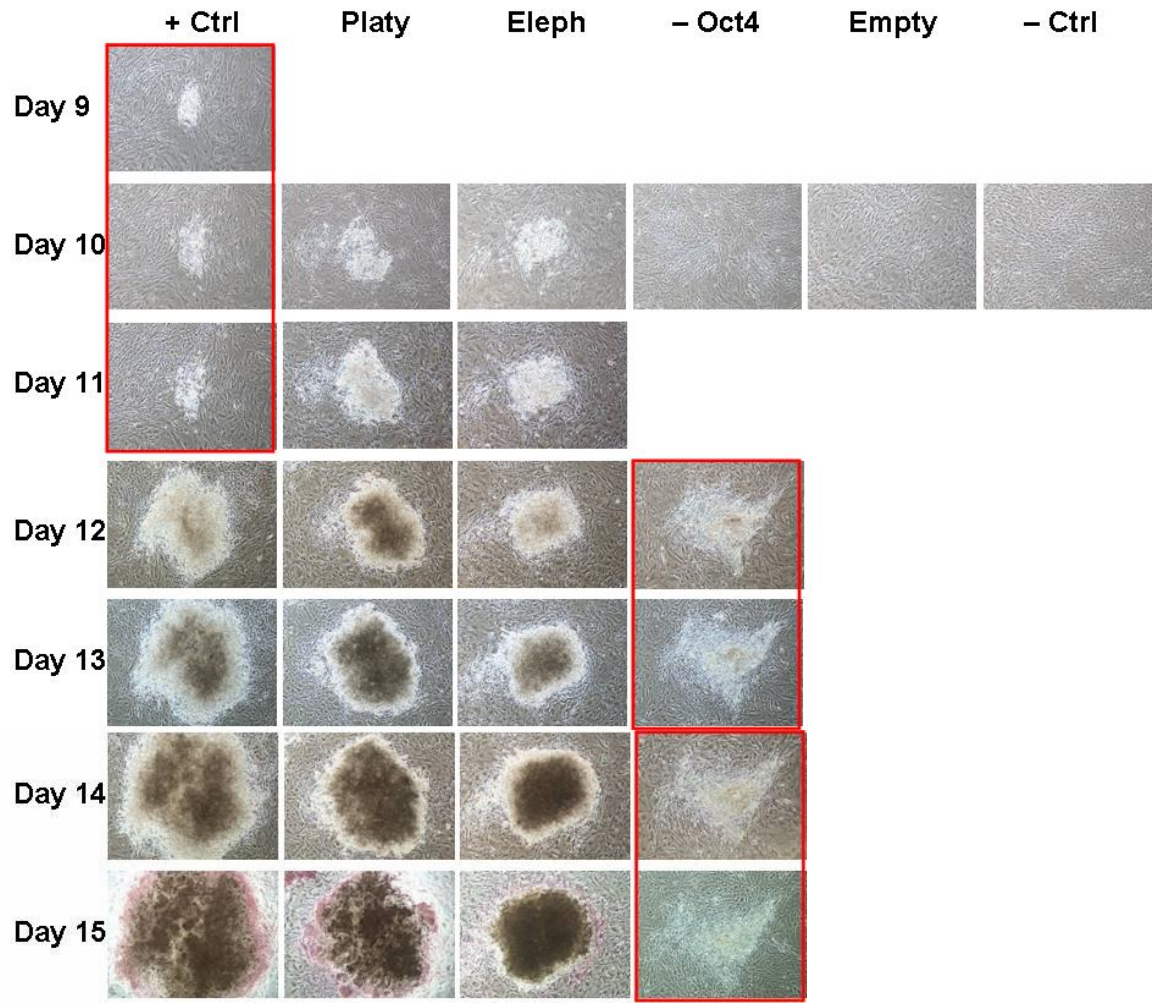


Figure 48. A selection of photos of induced colonies
"Day" denotes the number of days post-infection
Red boxes highlight colonies that failed to continue growing - alternate colonies (eg. Day 12 on +Ctrl plate) were then monitored
Purple colouration on Day 15 due to AP staining

On the 15th day post-infection, one set of plates was treated with alkaline phosphotase (AP) which stains for rapidly proliferating cells. Interestingly, the platypus experimental plate had the largest number of AP+ colonies, relative to the elephant and the mouse plates (Fig. 49). No AP+ colony was detected in the other three control plates.

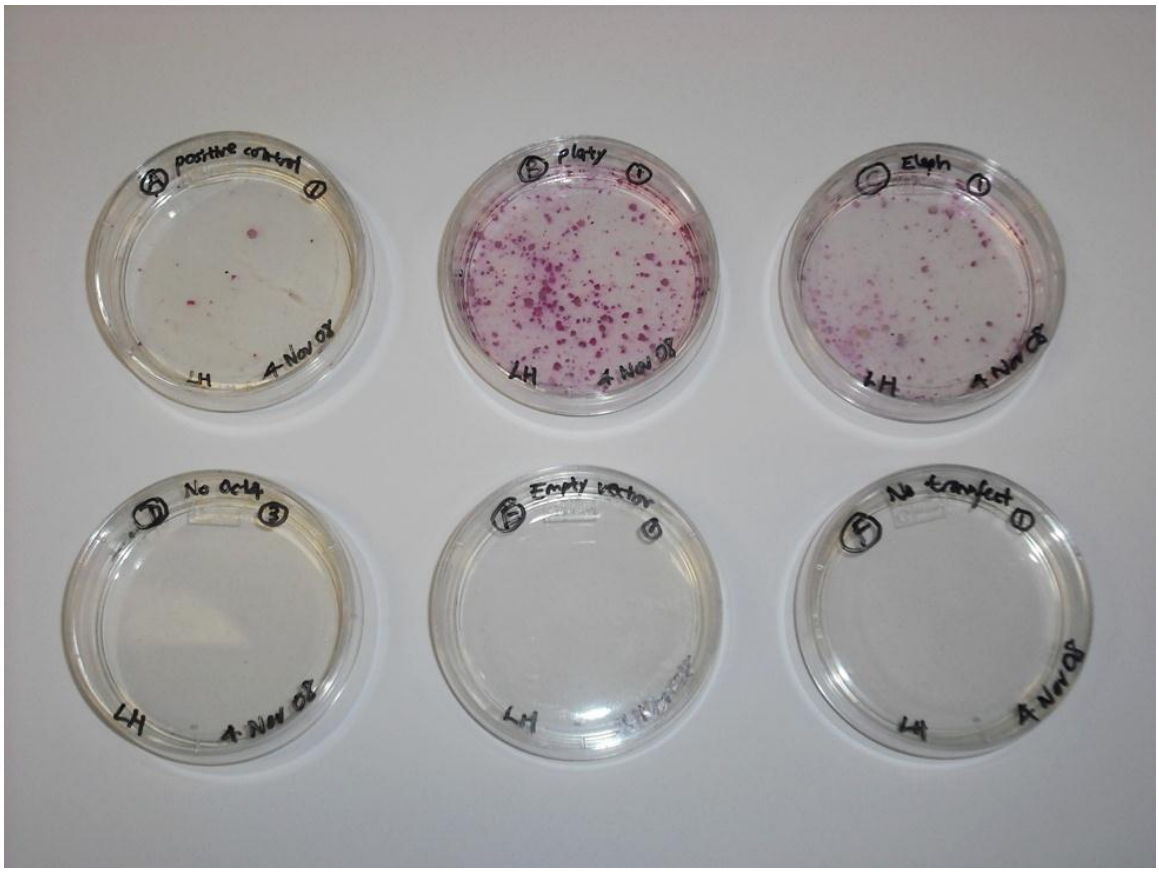


Figure 49. Alkaline phosphate staining on Day 15 post-infection

This result does not indicate that the platypus Oct4 chimera is the most effective at colony induction, since viral titres were not directly measured and so there may be variability in the viral infection efficiency. Nonetheless it is unexpected since the platypus Oct4 DBD does not have the eutherian-specific amino acid changes initially thought to be involved in the pluripotent function of Oct4, and thus should have no ability to induce any colony at all.

To find out if the induced colonies have long-term proliferation ability and maintain ES cell-like morphology, eight colonies (large and medium-sized) were picked for each of the mouse, elephant and platypus and seeded into a 24-well plate for growth monitoring.

Once a colony grows to confluency, it is then re-seeded into a 6-cm dish and stored for future analysis.

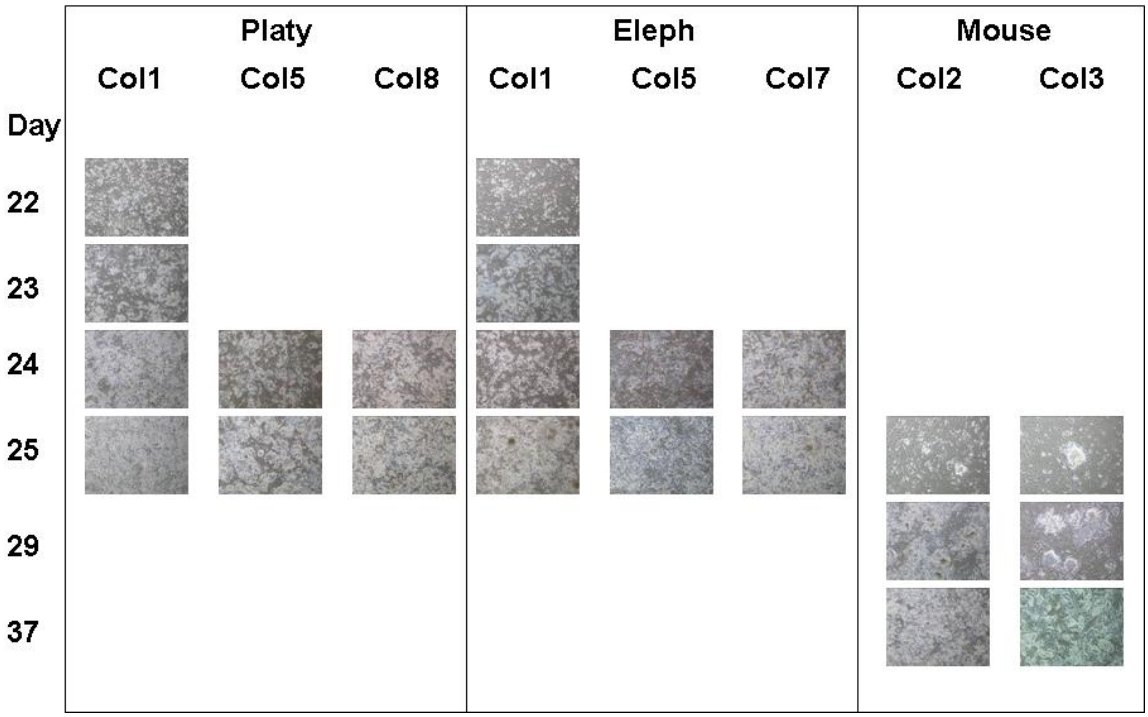


Figure 50. Monitoring the re-seeded iPS cells

In all three conditions, there were colonies that robustly proliferated after re-seeding (Fig. 50). However they did not uniformly display ES cell-like morphology with clearly defined colony edges. Some of the fast growing cell populations did not grow in colonies, or exhibited a flatter, EC-like morphology (Fig. 51). These variations could have been caused by incomplete induction of pluripotency in some of the plates, leading to partially reprogrammed iPS cells that are significantly different from ES cells.

Figure 51. Three main types of morphology

Thus, there is a possibility that the platypus Oct4 chimera could only induce cells to a partially reprogrammed state, where the other two eutherian Oct4 chimeras could fully reprogramme cells into true iPS cells. To further validate the endogenous activation of pluripotency genes, I derived fibroblast cells from mice containing an EGFP reporter knocked into the Sox2 locus (Ellis *et al.* 2004). In these mice, EGFP recapitulates endogenous Sox2 expression and this has been used to visualize and identify fully reprogrammed cells (Stadtfeld *et al.* 2008). These mice were available from Sohail Ahmed's lab (IMB, A*Star, Singapore).

I prepared adult fibroblasts from these mice by dissection to obtain lungs and a short section of their tails, which contain a large proportion of fibroblast cells (Fig. 52). These tissue samples are rinsed several times, finely minced and then plated onto T75 tissue culture flasks for continuous expansion until they reach sufficient cell numbers to be used

for the iPS experiments. Due to their faster doubling time and homogeneity of the cell population, tail fibroblasts were selected for this purpose.

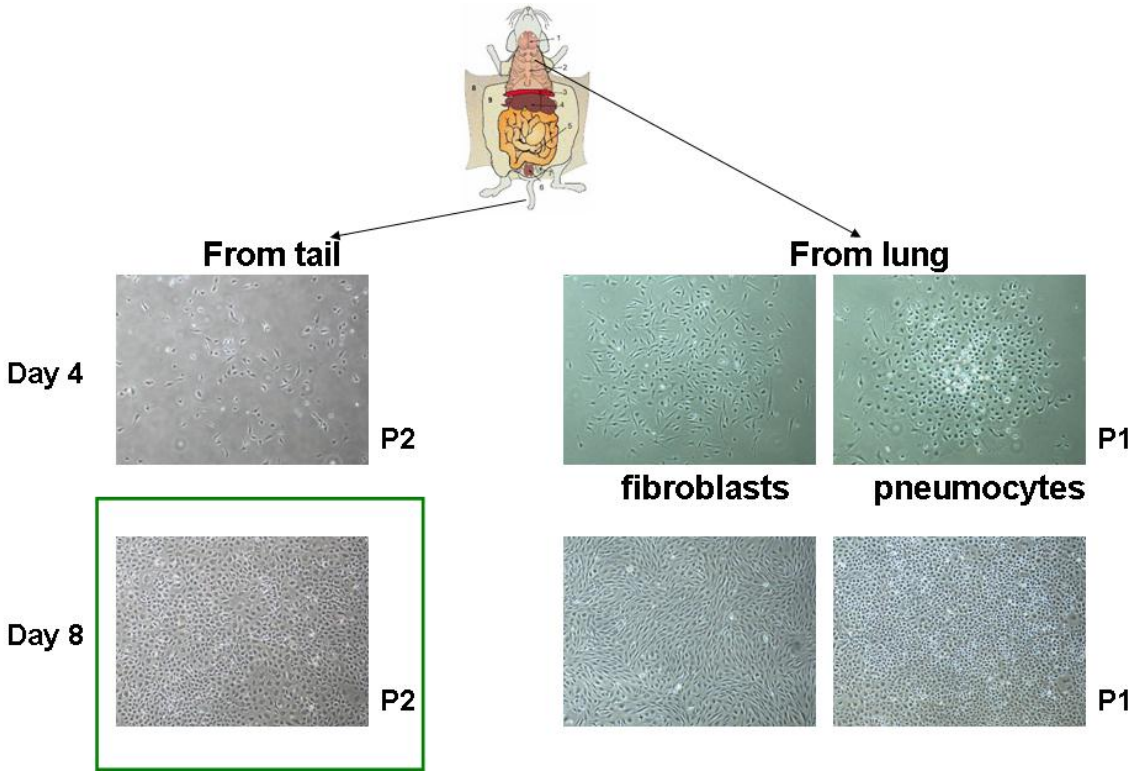


Figure 52. Primary culture of Sox2-EGFP fibroblast from adult mouse lung and tail
Green box indicates that tail fibroblasts were used for subsequent iPS experiments

The exact same iPS protocol was repeated using these Sox2-EGFP adult fibroblasts instead of BL6 embryonic fibroblasts. This time the colony induction process seemed to be slower, with distinct colonies appearing from Day 23 post-induction onwards, but they then proceeded to proliferate quickly as in the previous experiment.

Here again the platypus plate yielded a surprise when EGFP+ cells started to appear from Day 25, shining brightly within some of the induced colonies, similar to the eutherian plates (Fig. 53). This data indicates that the platypus Oct4 chimeric construct has the

capability to fully reprogramme, based on morphology and the induction of endogenous Sox2 expression, *adult* fibroblast cells into iPS cells further validating the my previous finding with *embryonic* fibroblasts.

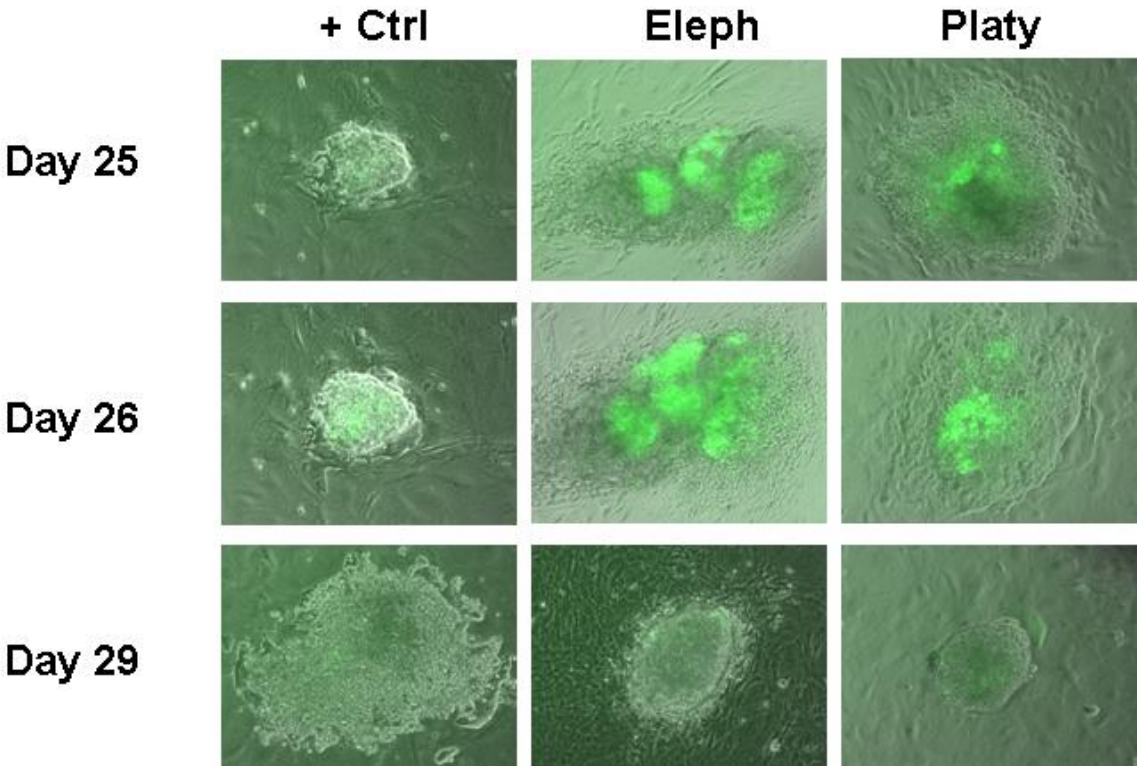


Figure 53. A selection of photos of Sox2-EGFP expressing colonies

In some of the colonies, the EGFP+ cells make up the majority of the cells, and the boundary of the colony is clearly demarcated under visible light and UV light (Fig. 54). Later, alkaline phosphate staining confirmed that many of these colonies are AP+ (Fig. 55).

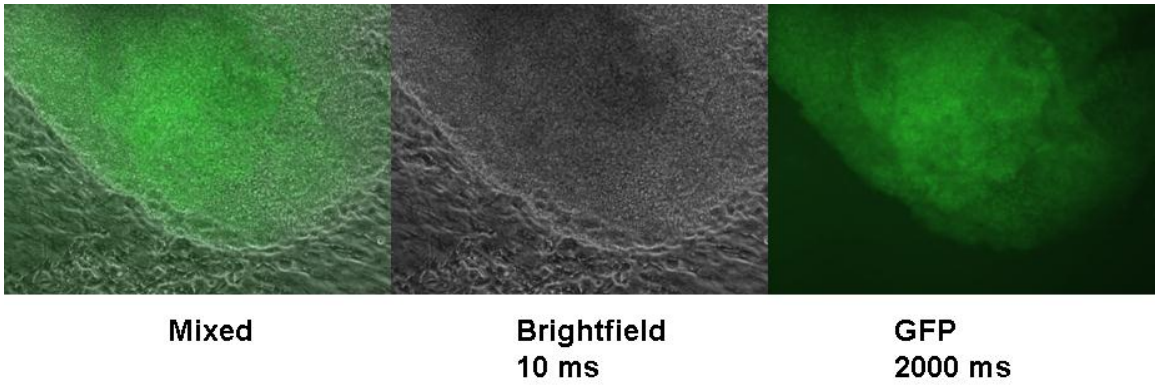


Figure 54. Close up of a Sox2-EGFP positive colony on the platypus plate, Day 31 post-infection

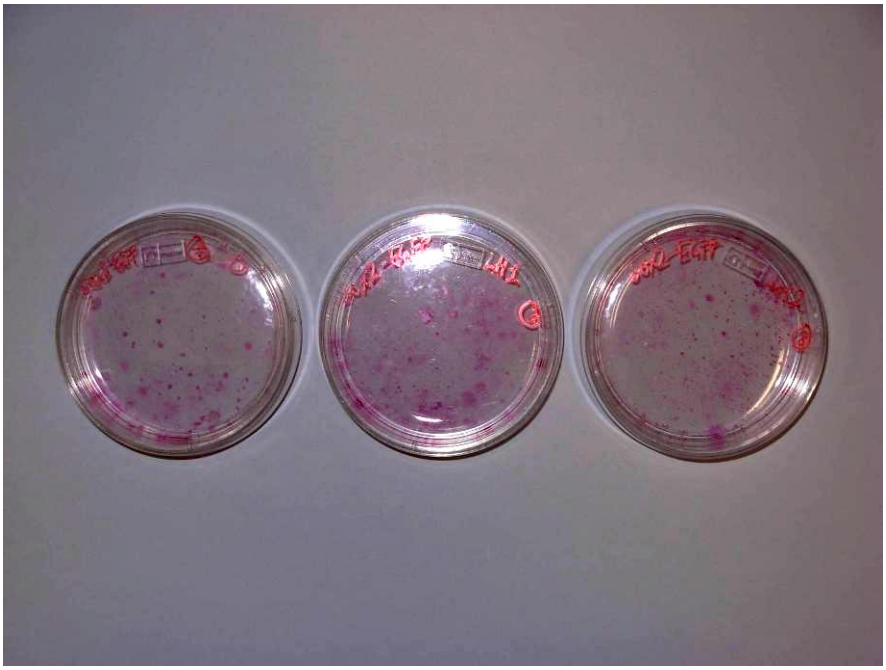


Figure 55. Alkaline phosphatase staining of Sox2-EGFP iPS plates
[+] = mouse Oct4, LH1 = elephant Oct4 chimera and LH2 = platypus Oct4 chimera

Although the platypus Oct4 chimera appeared to be capable of fully inducing pluripotency in the Sox2-EGFP fibroblasts, a possibility remains that the timing of the induction process might be delayed or slower relative to the eutherian Oct4.

To investigate this, all the experiment dishes in triplicate were examined daily and a colony scoring method was devised. Once the first green colonies begin to appear, all colonies are marked and the total number of EGFP+ and EGFP- colonies was counted daily for each dish. These figures are then averaged over the three replicates to minimize the variation of infection efficiency in each dish. The mean figures are then compiled and presented as an area chart to show the absolute growth in the number of colonies over time, and the relative growth of the green colonies vs non-green colonies (Fig. 56)

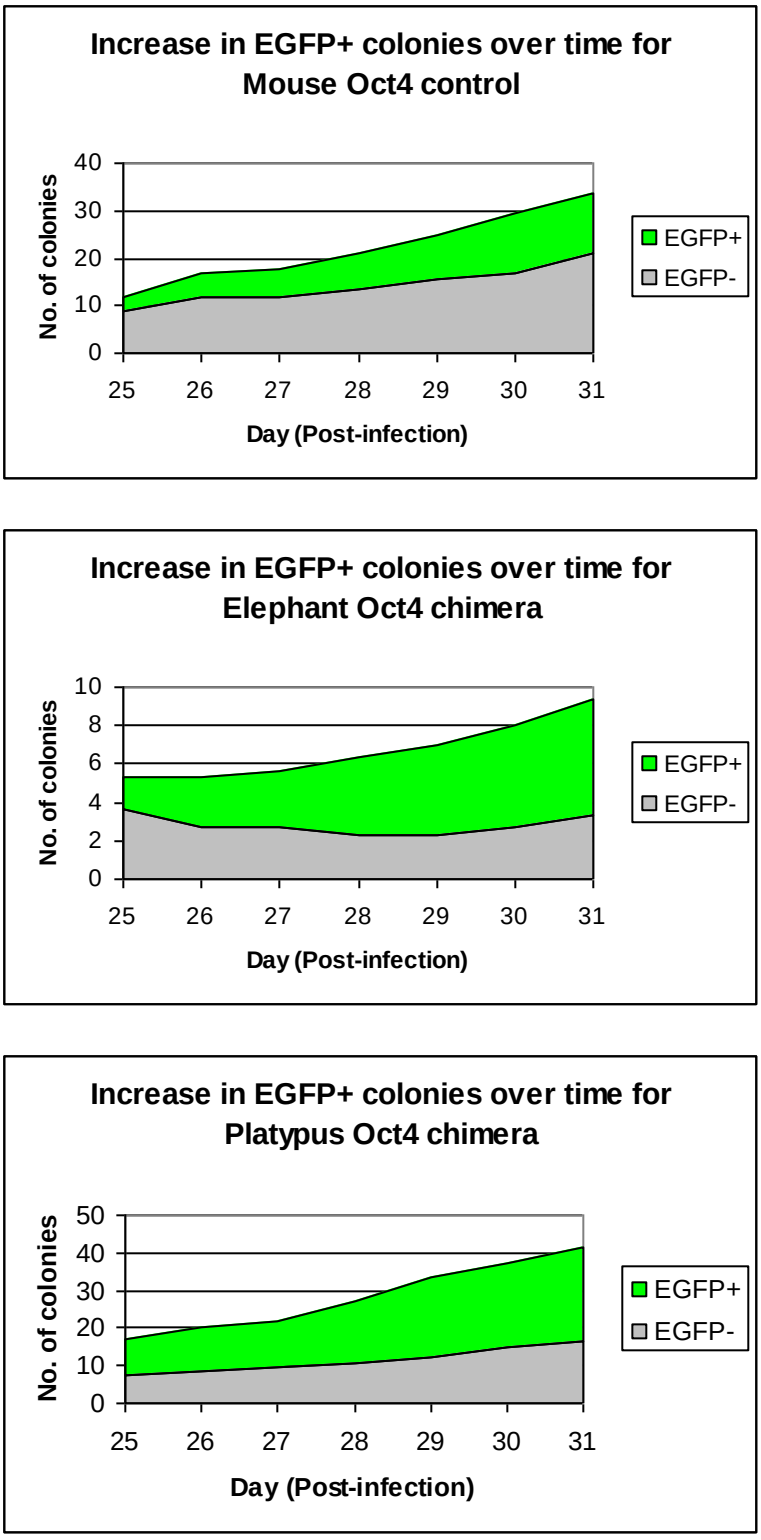


Figure 56. Increase in EGFP positive colonies over time

Once more the platypus had a surprise in store; instead of the delayed appearance of green colonies or a slower proliferation rate of the green colonies, the exact opposite occurred. Right from the beginning, the platypus dishes had proportionally more EGFP+ colonies and more in terms of absolute numbers than the mouse or the elephant. The elephant and the platypus had a higher rate of increase of EGFP+ colonies than the mouse. The overall rate of increase in the number of colonies, however, is fairly similar among all three species. While a single triplicate experiment is not enough to conclude that the platypus Oct4 chimera is better at inducing pluripotency than the eutherian Oct4, results so far strongly suggest that it is at least not deficient in this capability.

Finally, there was one last surprise. Of the 46 EGFP+ colonies on one of the platypus plates (LH2 plate1), six of them contain a differentiated cluster of cells that contain cardiomyocytes, despite the high LIF concentrations in the ES medium to maintain pluripotency. These EGFP+ cells beat spontaneously like a tiny heart at about 30 beats per minute and are only seen in that one plate, not on the plates of any other species (Fig. 57).

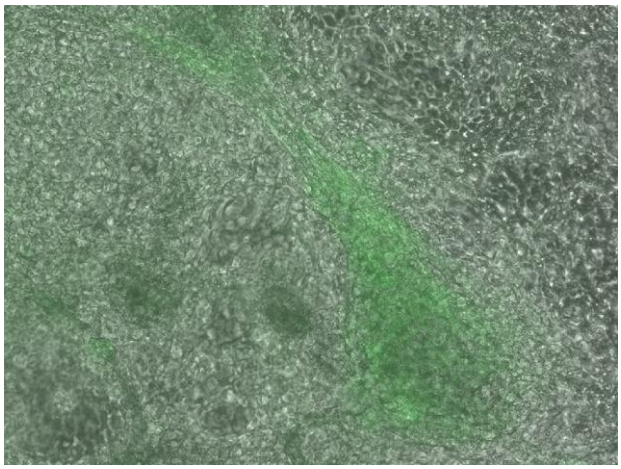


Figure 57. EGFP positive cardiomyocyte cluster in platypus dish

Chapter 5: Conclusion and Suggestions

5.1 Key Conclusions

Results from the functional studies so far do not show any major cell-level effects caused by the coding sequence difference between the platypus Oct4 DBD and that of eutherian mammals.

Subtle differences were revealed in the VP16/EnR real-time PCR experiments in the magnitude of downstream target responses, but no directionally different responses were seen. In the microarray experiment, the expression profile after the VP16/EnR treatment was highly similar between the platypus and the eutherians, and the handful of genes which did respond slightly differently are not known to be involved in the Oct4-Sox2-Nanog regulatory network, or pluripotency in general.

Likewise in the Oct4 chimera iPS experiments, the platypus Oct4 DBD chimera was fully capable of inducing both mouse embryonic and adult fibroblasts into iPS colonies. In fact, current results hint towards the possibility that the platypus Oct4 DBD may be more effective at inducing pluripotency than its eutherian counterparts, which seems counterintuitive. Either way, my data provides useful structure-function data with respect to Oct4 and its ability to reprogramme. Of the four original Yamanaka reprogramming factors Oct4 is the only one that was not replaceable by a homolog. Sox2 could be replaced by Sox1, 3, 18 and others, Klf4 with Klf2, 5 and others, while Myc is now known not to be an essential component. In contrast, Oct4 could not be replaced by Oct1

or Oct6 (Nakagawa *et al.* 2008). It will be interesting to see how far the reprogramming potential of Oct4 orthologs extends into the non-mammalian vertebrates.

During the early stages of this project, two assumptions were made. It was believed that (1) the early embryonic development in the platypus was in a discrete category separate from the eutherians, and that (2) significant amino acid changes in the Oct4 protein near the oct-sox interface is likely to have a significant effect on pluripotency. Based on these assumptions, a loss-of-function strategy was designed in anticipation of qualitatively different results between the platypus and the eutherians.

1. Halfway through my iPS experiments a research study was published that revealed the platypus has a simple placenta, formed from trophectoderm-like cells, that supports the embryo in the egg (Niwa *et al.* 2008). This observation suggests that platypus early development may not be as discretely different from eutherian development as first thought – the eutherian placenta may be an elaboration of a simple placenta in early mammals, rather than an outright evolutionary novelty. In that case, the platypus would not be a suitable out-group species. In addition, if this had been known earlier, an experimental strategy that is more sensitive to the quantitatively different results between the platypus and the eutherians would have to be employed.

The study also reported that the full-length platypus Pou5f1 is able to restore self-renewal in an ES cell line with its endogenous Pou5f1 expression conditionally repressed by tetracycline treatment (Niwa *et al.* 2008), while zebrafish Pou2 and opossum Pou2

homologs are unable to do so. Mouse ES cells maintained using platypus Pou5f1 were morphologically indistinguishable from untreated ES cells, expressing Sox2 and Nanog. This result contrasts with a similar complementation assay done using the chick PouV, where stem cell colonies were generated, but expressed low levels of Sox2 and Nanog, had limited capacity to be passaged and exhibited a differentiated morphology (Lavial *et al.* 2007). The researchers also did mutation experiments in the mouse Pou5f1 on amino acid positions that differ between Pou5f1 orthologs and Pou2 orthologs in an attempt to abolish its pluripotent function, but were unable to do so. Their findings are consistent with my experimental results that the platypus Oct4 DBD is capable of pluripotent function.

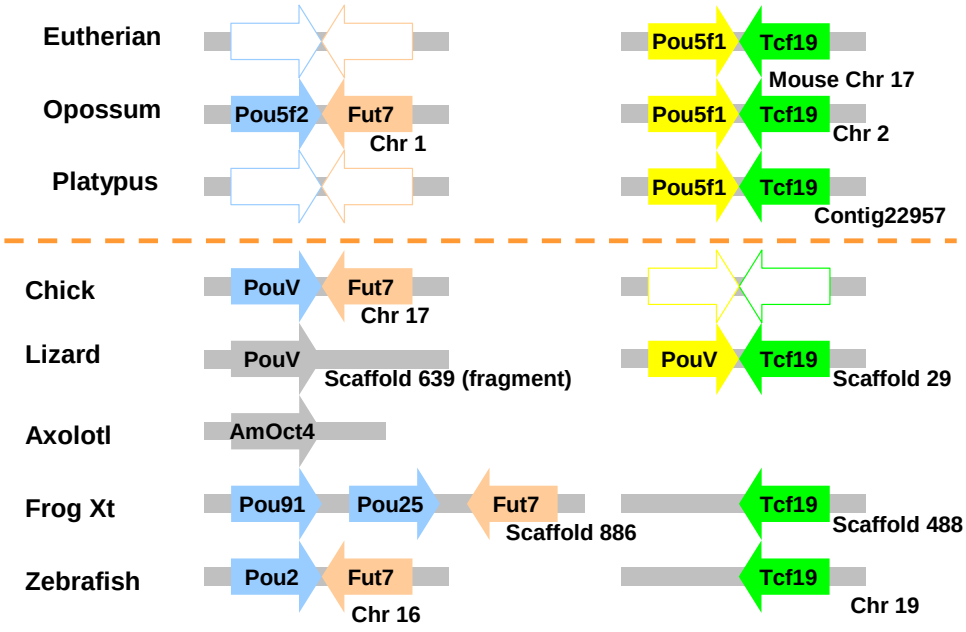


Figure 58. Emergence of Pou5f1 in a mammalian genomic context predates the evolution of mammals
Yellow = orthologs of Pou5f1
Blue = orthologs of Pou2
Grey = uncertain due to insufficient surrounding sequence data
Open block arrows = gene not found

Moreover, when we revisit the *Pou5f1* gene synteny map, it can be seen that lizard *PouV* gene is already syntenic with the *Tcf19* gene, just like in the mammalian genomic context, predating the later divergence between reptiles and mammals (Fig. 58). This suggests that the earliest opportunity for functional novelty in Oct4 via differences in cis-regulatory regions already exist before the platypus.

The gene synteny map is then used to create a reconstructed history of the evolution of Oct4. A recent study has found an additional *Pou5f2* ortholog in the platypus (Niwa *et al.* 2008) and latest research suggests that the axolotl *AmOct4* is more similar to mammalian *Pou5f1* than to *Pou2* related homologs (Frankenberg *et al.* 2010). These findings have been incorporated in my reconstruction of Oct4 history (Fig. 60). As you can see, *Pou5f2* is likely the ancestral gene, found in the same genomic context in the fish and the frog.

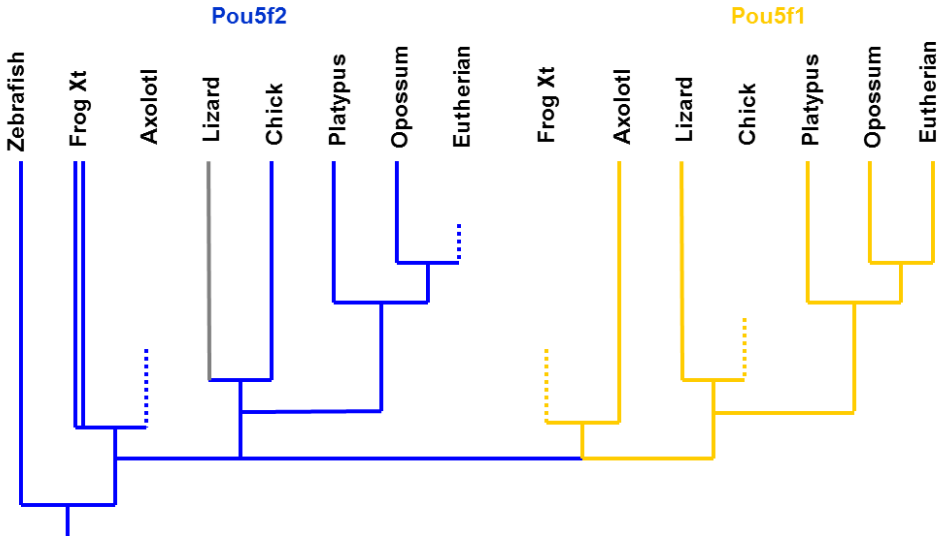


Figure 59. Reconstructed evolutionary history of Oct4
Yellow = orthologs of Pou5f1
Blue = orthologs of Pou5f2
Grey = uncertain due to insufficient sequence data
Dashed line = possible gene death

The gene duplication event that led to the emergence of paralog *Pou5f1* most likely occurred before the divergence between amphibians and other tetrapods. *Pou5f1* is already in the mammalian genomic context in the lizard. Notably, both *Pou5f1* and *Pou5f2* exist in the non-eutherian mammals shown here, suggesting that the loss of *Pou5f2* occurred at the base of the eutherians.

2. Based on the multi-species alignment, eutherian-specific amino acid changes in the Oct4 DBD were identified as having the potential to affect oct-sox binding, and thus affect the pluripotent functions of the protein.

These highly-conserved coding sequence changes are most likely to be important to eutherian mammals. However, Oct4 is not only involved in pluripotency – it also plays a crucial role in the maintenance of the germline (Kehler *et al.* 2004) and the differentiation of cardiomyocytes (Zeineddine *et al.* 2006). Since the focus of the current project is on pluripotency, any importance of the sequence changes to these other functions were not evaluated. For example, the unexpected propensity for the platypus Oct4 chimera to induce cardiomyocytes even under ES media conditions with high LIF, hints to the tantalizing possibility that platypus Oct4 DBD may have an enhanced ability to direct the differentiation of heart muscle compared to its eutherian counterpart.

Moreover, it is not clear whether the absence of eutherian-specific changes actually inhibit oct-sox binding at the molecular level. Gel shift experiments done by a collaborator Ralf Jausch indicated that the platypus Oct4 DBD binds even more strongly

to the canonical oct-sox element with mouse Sox2 compared to the binding between mouse Oct4 DBD and mouse Sox2,. This result, in conjunction with my results from the iPS experiments strongly suggests that the direction of the change is opposite to my initial postulation.

5.2 Cis-evolution of Critical Genes

Some developmental biologists believe that mutations to the cis-regulatory regions of genes are relatively more important to the evolution of morphological features than coding sequence mutations (Carroll 2008). In the early stages of this study, it was believed that coding sequence changes to transcription factors are essentially as important to cell-type evolution as cis-regulatory changes, since transcription factors bind to specific DNA binding sites and thus the cis-acting and trans-acting changes operate in a continuum.

However, all three core members of the Oct4-Sox2-Nanog regulatory network are critical genes for pre-implantation development and also have important functions in other aspects of development, such as neural and germ cell development. Thus, the potential for lethal mutations is high, and the whole network as a system needs to maintain robustness against small changes at the molecular level just to survive the developmental process. So although these genes have many downstream targets, the cell-level effects of small coding sequence changes may be masked instead of amplified – via compensatory mechanisms between the members of the network to prevent the whole network from disintegrating due to a small number of replication errors.

As such the coding sequence of highly-interconnected, critical genes is not really the most optimal target if one wishes to look for small mutations that can result in large phenotypic effects. Another approach is to target the cis-regulatory regions. For example, in the *Nanog* promoter assays discussed earlier in Section 4.3, point mutations of only 3 base pairs could significantly reduce the promoter activity of the Sox2-Oct4 site. Moreover, there is no evidence of a Sox2-Oct4 binding site in non-eutherian mammal or chick *Nanog*. In a similar vein, in the Niwa paper the researchers examined the CR4 region of the platypus *Pou5f1* promoter and found that the auto-regulatory element involved in the reciprocal inhibition between Pou5f1 and Cdx2 is missing (Niwa *et al.* 2008). This is important because high Cdx2 expression is essential for placental formation. When promoter assays were performed on the CR4 region they found that it had no enhancer function. They postulate that this difference may result in the simpler placenta of the platypus in contrast with the sophisticated eutherian placenta.

5.3 Future Work

Since the platypus Oct4 can maintain pluripotency, while the zebrafish paralog cannot, one way forward is to study the *PouV* gene in the *Anolis carolinensis* to find out if the pluripotent capability of the Oct4 protein itself already exists in the lizard. The gene vicinity can also be compared with the mouse and the platypus to see if there are any differences in the cis-regulatory elements.

Apart from the study of pluripotency, from a purely technical perspective it would be interesting to perform more quantitative experiments to find out how much more efficient the platypus Oct4 chimera is in the induction of iPS relative to mouse Oct4, and to characterize the specific amino acid changes that lead to improved efficiency. The use of platypus Oct4 chimera as a supplement for the directed differentiation of cardiomyocytes can also be further investigated.

Bibliography

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 Dev* **17**: 126-140

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., Gifford D.K., Melton D.A., Jaenisch R. and Young R.A. (2005) Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122**: 1-10

Bradley A., Evans M.J., Kaufman M.H. and Robertson E.J. (1984) Formation of germ line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**: 255-256

Bradley A., Zheng B. and Liu P. (1998) EGF, epithelium and Modifying the mouse genome, Thirteen years of manipulating the mouse genome: a personal history. *Int. J. Dev. Biol.* **42**: 943-950

Brinster R.L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *J. Exp. Med*. **140**: 1049-1056

Carroll S.B. (2005) Evolution at two levels: On genes and form. *PLoS Biol* **3**: 1159-1166

Carroll S.B. (2008) Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134**: 25-36

Carsona C.T., Pagratisa M. and Parr B.A. (2004) Tbx12 regulates eye development in Xenopus embryos. *Biochem Biophys Res Comms* **318**: 485-489

Chambers I., Colby D., Robertson M., Nichols J., Lee S., Tweedie S. and Smith A. (2003) Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* **113**: 643-655

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., Ng H.H. (2005) Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol.* **25** :6031-6046.

Cretekos C.J.,Wang Y., Green E.D., NISC Comparative Sequencing Program, Martin J.F., Rasweiler IV J.J. and Behringer R.R. (2008) Regulatory divergence modifies limb length between mammals. *Genes & Dev* **22**: 141-151

Darwin, C. R. (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 1st edition,1st issue. John Murray, London.

Davidson E.H. (2001) Genomic Regulatory Systems: Development and Evolution. Academic Press, San Diego USA.

Davidson E.H., Rast J.P., Oliveri P., Ransick A., Calaestani C., Yuh C., Minokawa T., Amore G., Hinman V., Arenas-Mena C., Otim O., Brown T., Livi C.B., Lee P.Y., Revilla R., Rust A.G., Pan Z.J., Schilstra M.J., Clarke P.J.C., Arnone M.I., Rowen L., Cameron R.A., McClay D.R., Hood L. and Bolouri H. (2002) A Genomic Regulatory Network for Development. *Science* **295**: 1669-1678

Ellis P., Fagan B.M., Magness S. T., Hutton S., Taranova O., Hayashi S., McMahon A., Rao M. and Pevny L. (2004) SOX2, a Persistent Marker for Multipotential Neural Stem Cells Derived from Embryonic Stem Cells, the Embryo or the Adult. *Dev Neurosci* **26**:148-165

Enders A.C. (2002). Implantation in the nine-banded armadillo: how does a single blastocyst form four embryos? *Placenta* **23**: 71-85.

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

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. *Dev Biol* **337**: 162-170

Gompel N., Prud'homme B., Wittkopp P.J., Kassner V.A. and Carroll S.B. (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433**: 481-487

Gould S.J. (1989) Wonderful Life. Norton, New York USA.

Hamburger V. and Hamilton H.L. (1951) A series of normal stages in the development of the chick embryo. *J Morphol* **38**:49–92

Hughes R.L. and Hall L.S. (1998) Early development and embryology of the platypus. *Philos Trans R Soc Lond B Biol Sci.* **353**:1101-14.

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

Kimura, M. (1983) The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge UK.

King M.C., Wilson A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* **188:** 107–116.

Kumano A., Sasaki M., Budipitojo T., Kitamura N., Krause W.J. and Yamada J. (2005) Immunohistochemical Localization of Gastrin-releasing Peptide, Neuronal Nitric Oxide Synthase and Neurone-specific Enolase in the Uterus of the North American Opossum, Didelphis virginiana. *Anat. Histol. Embryol.* **34**: 225–231

Kuroda T., Tada M., Kubota H., Kimura H., Hatano S.Y., Suemori H., Nakatsuji N. and Tada T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol.* **25**: 2475-2485.

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

Loh Y.H., Wu Q., Chew J.L., Vega V.B., Zhang W., Chen X., Bourque G., George J., Leong B., Liu J., Wong K.W., Sung K.W., Lee C.W.H., Zhao X.D., Chiu K.P., Lipovich L., Kuznetsov V.A., Robson P., Stanton L.W., Wei C.L., Ruan Y., Lim B. and Ng H.H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genetics* **38**: 431-440

Luo Z., Crompton A.W. and Sun A. (2001) A New Mammaliaform from the Early Jurassic and Evolution of Mammalian Characteristics. *Science* **292**: 1535-1540

Martin G.R. and Evans M.J. (1974) The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**:163-172

McGinnis W., Levine M.S., Hafen E., Kuroiwa A. and Gehring W.J. (1984). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* **308**: 428-433

Mesnard D., Filipe M., Belo J.A. and Zernicka-Goetz M. (2004) The Anterior-Posterior Axis emerges respecting the morphology of the mouse embryo that changes and aligns with the uterus before gastrulation. *Curr Biol* **14**:184–196

Miller C.T.,Beleza S.,Pollen A.A.,Schluter D., Kittles R.A., Shriver M.D. and Kingsley D.M. (2007) cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. *Cell* **131**, 1179–1189

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., Nishimoto M.1, Saito T., Ninomiya M., Sawamoto K., Okano H., Muramatsu M., Oguro H., Iwama A. and Okuda A. (2006) The *Sox2* regulatory region 2 functions as a neural stem cell specific enhancer in the telencephalon. *J Biol Chem* **281**:13374-13381

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

Nakagawa M., Koyanagi M., Tanabe K., Takahashi K., Ichisaka T., Aoi T., Okita K., Mochiduki Y., Takizawa N. and Yamanaka S. (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotech* **26**: 101-106

Nichols J., Zevnik B., Anastassiadis K., Niwa H., Klewe-Nebenius I.D., Chambers I., Scholer 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.

Niwa H., Sekita Y., Tsend-Ayush E. and Grutzner F. (2008) Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. *Evol Dev* **10**: 671-682

Orr H.A. (2005) The Genetic Theory of Adaptation: A Brief History. *Nat Rev Gen* **6**: 119-127

Pan G. and Pei D. (2005) The Stem Cell Pluripotency Factor NANOG Activates Transcription with Two Unusually Potent Subdomains at its C Terminus. *J Biol Chem* **280**: 1401-1407

Pennisi E. (2008) Deciphering the Genetics of Evolution. *Science* **321**: 760-763

Peppler R.D. and Stone S.C. (1980) Plasma progesterone level during delayed implantation, gestation and postpartum period in the armadillo. *Lab Anim Sci*. **30**:188-191

Purugganan M.D. (1998) The molecular evolution of development. *BioEssays* **20**:700–711

Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R. and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048-2059.

Renfree M.B. and Shaw G. (2000) Diapause. *Annu Rev Physiol* **62**:353-375

Rinkenberger J.L., Cross J.C. and Werb Z. (1997) Molecular genetics of implantation in the mouse. *Dev Genetics* **21**:6-20

Rodda D.J., Chew J.L., Lim L.H., Loh Y.H., Wang B., Ng H.H. and Robson P. (2005) Transcriptional Regulation of Nanog by OCT4 and SOX2. *J Biol Chem* **280**: 24731-24737

Sagai T., Hosoya M., Mizushina Y., Tamura M. and Shiroishi T. (2005) Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**: 797-803

Silva J., Nichols J., Theunissen T.W., Guo G., van Oosten A.L., Barrandon O., Wray J., Yamanaka S., Chambers I. and Smith A. (2009) Nanog Is the Gateway to the Pluripotent Ground State. *Cell* **138**: 722-737

Soodeen-Karamath S. and Gibbins A.M.V. (2001) Apparent Absence of oct 3/4 From the Chicken Genome. *Mol Reprod Dev* **58**:137-148

Springer M.S., Murphy W.J., Eizirik E. and O'Brien S.J. (2003) Placental Mammal Diversification and the Cretaceous-Tertiary Boundary. *Proc Natl Acad Sci* **100**:1056-1061

Stadtfeld M., Maherali N., Breault D.T. and Hochedlinger K. (2008) Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse. *Cell Stem Cell* **2**:230-240

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

Tam P.P.L. and Rossant J. (2003) Mouse embryonic chimeras: tools for studying mammalian development. *Development* **130**: 6155-6163

Tomioka M., Nishimoto M., Miyagi S., Katayanagi T., Fukui N., Niwa H., Muramatsu M. and Okuda A. (2002) Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res.* **30**: 3202-3213.

Wakimoto B.T., Turner F.R., Kaufman T.C. (1984) Defects in embryogenesis in mutants associated with the antennapedia gene complex of Drosophila melanogaster. *Dev Biol.* **102**:147-172.

Wang J., Rao S., Chu J., Shen X., Levasseur D.N., Theunissen T.W. and Orkin S.H. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**: 364-368

Weaver C. and Kimelman D. (2004) Move it or lose it: axis specification in Xenopus. *Development* **131**: 3491-3499

Wilkins A.S. (2002) The Evolution of Development Pathways. Sinauer Associates, Inc. Sunderland, Massachussetts USA.

Williams, Jr. D.C., Cai M. and Clore G.M. (2004) Molecular Basis for Synergistic Transcriptional Activation by Oct1 and Sox2 Revealed from the Solution Structure of the 42-kDa Oct1_Sox2_*Hoxb1*-DNA Ternary Transcription Factor Complex. *J Biol Chem* **279**: 1449-1457

Yousef A. and Selwood L. (1993) Embryonic Development in Culture of the Marsupials *Antechinus stuartii* (Macleay) and *Sminthopsis macroura* (Spencer) during Preimplantation Stages. *Reprod. Fertil. Dev.* **5**: 445-458

Yuan H., Corbi N., Basilico C. And Dailey L. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**: 2635-2645

Zeineddine D., Papadimou E., Chebli K.,Gineste M., Liu J., Grey C., Thurig S., Behfar A., Wallace V.A., Skerjanc I.S. and Puce´ M. (2006) Oct-3/4 Dose Dependently Regulates Specification of Embryonic Stem Cells toward a Cardiac Lineage and Early Heart Development. *Dev Cell* **11**: 535-546

Appendix A

BAC Library Screening Database

Oct4									
S/No	Species	Filter Set	Filter No.	Panel/Field	Grid Location	Vector/Position	Plate range	Plate ID	BAC Clone ID
1	Elephant	VMRC-15	7E	5	C6	E	313-318	317	VM15-317C6
2				3	G7	C	301-306	303	VM15-303G7
3				3	E7	B	295-300	297	VM15-297E7
4			8E	2	J19	A	337-342	338	VM15-338J19
5				2	J11	G	373-378	374	VM15-274J11
6				2	J7	C	349-354	350	VM15-350J7
7				2	A9	A	337-342	338	VM15-338A9
8				6	P8	G	373-378	378	VM15-378P8
9				6	P8	C	349-354	354	VM15-354P8
10				6	J23	D	355-360	360	VM15-360J23
11				6	C3	D	355-360	360	VM15-360C3
12				6	A11	E	361-366	366	VM15-366A11
13				3	E2	A	337-342	339	VM15-339E2
14			9E	2	H8	B	391-396	392	VM15-392H8
15				5	C23	B	391-396	395	VM15-395C23
16				3	F2	F	415-420	417	VM15-417F2
17			10E	4	J13	A	433-438	436	VM15-436J13
18				2	C6	D	451-456	452	VM15-452C6
19				1	E1	G	469-474	469	VM15-469E1
20			11E	2	M19	B	487-492	488	VM15-488M19
21			12E	4	L16	C	541-546	544	VM15-544L16
22				4	I15	H	571-576	574	VM15-574I15
23				4	F1	D	547-552	550	VM15-550F1
24				3	I17	B	535-540	537	VM15-537I17
25			13E	6	F9	A	577-582	582	VM15-582F9
26				3	M18	A	577-582	579	VM15-579M18
27				3	N9	A	577-582	579	VM15-579N9
28				3	J4	C	589-594	591	VM15-591J4
29	Opossum	VMRC-6	8K	1	F3	B	343-348	343	VM6-343F3
30			10K	6	D5	F	463-468	468	VM6-468D5
31			11K	4	C11	C	493-498	496	VM6-496C11
32			12K	4	A18	E	553-558	556	VM6-556A18
33			13K	6	E12	B	583-588	588	VM6-588E12
34			14K	6	P22	F	655-660	660	VM6-660P22
35	Kangaroo	ME_KBa	A	6	B11	7	Plate 42	Plate 42	42B11
36			A	3	G5	1	Plate 3	Plate 3	3G5
37			A	1	G8	6	Plate 31	Plate 31	31G8
38			A	1	J11	1	Plate 1	Plate 1	1J11
39			A	5	K2	8	Plate 47	Plate 47	47K2
40			B	2	N10	3	Plate 14	Plate 62	62N10
41			B	4	B10	1	Plate 4	Plate 52	52B10
42			B	1	F19	1	Plate 1	Plate 49	49F19
43			B	1	O24	4	Plate 19	Plate 67	67O24
44			C	2	A16	1	Plate 2	Plate 98	98A16
45			C	6	B3	8	Plate 48	Plate 144	144B3
46			C	6	P6	4	Plate 24	Plate 120	120P6
47			C	3	E15	3	Plate 15	Plate 111	111E15
48			C	3	N20	5	Plate 27	Plate 123	123N20
49			C	4	D5	6	Plate 34	Plate 130	130D5
50			C	4	K5	1	Plate 4	Plate 100	100K5
51			C	5	K8	5	Plate 29	Plate 125	125K8
52			D	3	A20	3	Plate 15	Plate 159	159A20
53			D	1	A8	3	Plate 13	Plate 157	157A8
54			E	3	K20	8	Plate 45	Plate 237	237K20
55			F	3	F3	8	Plate 45	Plate 285	285F3
56			F	3	N22	3	Plate 15	Plate 255	255N22
57			F	3	F24	3	Plate 15	Plate 255	255F24
58			F	4	G20	8	Plate 46	Plate 286	286G20
59			G	3	C12	5	Plate 27	Plate 315	315C12
60			H	6	N18	2	Plate 12	Plate 348	348N18
61			H	3	K23	3	Plate 15	Plate 351	351K23
62			I	6	A9	8	Plate 48	Plate 432	432A9
63			I	4	O2	4	Plate 22	Plate 406	406O2
64			I	1	F13	1	Plate 1	Plate 385	385F13
65			I	1	G15	5	Plate 25	Plate 409	409G15
66			J	1	O18	7	Plate 37	Plate 469	469O18
67			K	3	J2	5	Plate 27	Plate 507	507J2
68			K	1	O11	1	Plate 1	Plate 481	481O11
69			K	1	L20	7	Plate 37	Plate 517	517L20
70			L	4	D1	8	Plate 46	Plate 574	574D1
71			L	1	L18	4	Plate 19	Plate 547	547L18
72			L	5	P9	2	Plate 11	Plate 539	539P9
73			M	2	G5	5	Plate 26	Plate 602	602G5
74			M	1	B9	6	Plate 31	Plate 607	607B9
75			M	1	G22	1	Plate 1	Plate 577	577G22
76	Platypus	OA_Bb	B	5	H4	2	Plate 11	Plate 59	59H4
77			C	3	H8	2	Plate 9	Plate 105	105H8
78			C	1	E20	1	Plate 1	Plate 97	97E20
79			C	5	I19	5	Plate 29	Plate 125	125I19
80			D	2	J10	1	Plate 2	Plate 146	146J10
81			E	3	M7	3	Plate 15	Plate 207	207M7
82			J	4	M12	5	Plate 28	Plate 460	460M12
83			J	4	A15	6	Plate 34	Plate 466	466A15
84	Armadillo	VMRC-5	1	2	F16	E	25-30	Plate 26	VM5-26F16
85			2	2	K17	F	79-84	Plate 80	VM5-80K17
86			4	6	L8	E	169-174	Plate 174	VM5-174L8
87			4	6	C10	F	175-180	Plate 180	VM5-180C10

Nanog									
S/No	Species	Filter Set	Filter No.	Panel/Field	Grid Location	Vector/Position	Plate range	Plate ID	BAC Clone ID
1	Elephant	VMRC-15	7E	1	E22	E	313-318	313	VM15-313E22
2				1	G23	B	295-300	295	VM15-295G23
3				2	L6	F	319-324	320	VM15-320L6
4				3	A3	H	331-336	333	VM15-333A3
5			8E	2	C5	D	355-360	356	VM15-356C5
6				2	G11	G	373-378	374	VM15-374G11
7			9E	5	E18	H	427-432	431	VM15-431E18
8				5	F19	H	427-432	431	VM15-431F19
9				2	B2	E	409-414	410	VM15-410B2
10				6	A21	F	415-420	420	VM15-420A21
11				6	L8	B	391-396	396	VM15-396L8
12			10E	1	M21	B	439-444	439	VM15-439M21
13				2	A10	H	475-480	476	VM15-476A10
14				6	C4	G	469-474	474	VM15-474C4
15			11E	4	C24	B	487-492	490	VM15-490C24
16			12E	6	G18	C	541-546	546	VM15-546G18
17	Armadillo	VMRC-5	1C	5	G7	D	19-24	23	VM5-23G7
18				5	J21	D	19-24	23	VM5-23J21
19				2	L24	H	43-48	44	VM5-44L24
20			2C	1	B18	H	91-96	91	VM5-91B18
21				2	A4	F	79-84	80	VM5-80A4
22				2	E5	H	91-96	92	VM5-92E5
23			3C	5	B14	C	109-114	113	VM5-113B14
24				5	E19	F	127-132	131	VM5-131E19
25				3	L3	B	103-108	105	VM5-105L3
26			4C	4	I17	H	187-192	190	VM5-190I17
27				1	P19	E	169-174	169	VM5-169P19
28				2	O13	G	181-186	182	VM5-182O13
29				6	B13	B	151-156	156	VM5-156B13
30				3	G23	B	151-156	153	VM5-153G23
31			5C	2	F11	A	193-198	194	VM5-194F11
32				5	F20	B	199-204	203	VM5-203F20
33				5	G8	D	211-216	215	VM5-215G8
34				2	C4	G	229-234	230	VM5-230C4
35			6C	2	J13	A	241-246	242	VM5-242J13
36				1	B15	E	265-270	265	VM5-265B15
37				3	J5	H	283-288	285	VM5-285J5
38	Opossum	VMRC-6	8K	4	N14	A	337-342	340	VM6-340N14
39				2	E1	C	349-354	350	VM6-350E1
40				6	L20	F	367-372	372	VM6-372L20
41				5	D13	E	361-366	365	VM6-365D13
42				3	K7	E	361-366	363	VM6-363K7
43			9K	2	D9	C	397-402	398	VM6-398D9
44			12K	1	D2	C	541-546	541	VM6-541D2
45			13K	2	P20	B	583-588	584	VM6-584P20
46			14K	4	G23	F	655-660	658	VM6-658G23
47				2	O9	F	655-660	656	VM6-656O9
48	Kangaroo	ME_KBa	C	2	J2	5	Plate 26	Plate 122	122J2
49			C	3	N7	2	Plate 9	Plate 105	105N7
50	Platypus	OA_Bb	K	6	L4	7	Plate 42	Plate 522	522L4

Appendix B

BAC Screening Protocol

This protocol has three parts:

I. Radioactive Work Protocol

II. BAC High-density Filter Screening

III. Protocol for reading BAC IDs

Part I. Radioactive Work Protocol

This information is to be used as an introduction to basic procedures and safety in research work involving radioactive materials. It is specific to Robson Lab, Genome Institute of Singapore, and reflects the recommended procedures in Jan 2007.

For further details and clarifications please consult the current radiation work committee.

General Procedures

1. Please be suitably attired prior to entry into radiation room.

Attire:

- Long-sleeved lab coat
- Plastic goggles (if not wearing spectacles)
- **TLD Badge** (the black dosimeter tag, to be worn on the collar/lab coat chest pocket)
- Covered shoes.

2. Enter using access pass.

3. Please put on double-gloves after entering the room.

The outer layer is to be discarded into radioactive waste box immediately when contaminated (or suspect to be contaminated). The inner layer, if not contaminated, can be worn until the end of the experiments.

4. Turn on the Geiger counter and begin checking the work area.

Turn the Geiger counter's dial to"X1". Turn audible to "On". It should start clicking randomly and sparsely (on average about once a second). Free the pancake scanner arm and scan yourself and the work area by hovering the pancake over the area of interest. The wire mesh detector must face the direction you want to scan. **Do not let the detector touch anything - it may become contaminated.** To verify that the counter is working, simply open the radioactive waste box and hover the pancake over it – the counter should click vigorously (several clicks per second). Close the waste box when done, and check your gloves with the counter.

It is recommended to check other areas such as: Door handle, light switch, fridge/freezer handle, outside of radioactive waste boxes and heating blocks. This is to verify that the previous user did not leave any contamination behind.

5. Set up the Geiger counter to monitor your work area.

Rotate the pancake to face the work area. Hover your gloves or other items over the detector to check for contaminations.

6. Proceed with radiation work.

Guidelines:

- **Be very sure of your protocol.** If possible, rehearse the procedure without the radioactive material first.
- **Minimize exposure time.** The radio-material should be in its lead container or in the acrylic housing box almost all of the time. You should be behind the acrylic shield whenever the radio-material is not covered adequately. If you must raise the tube into the air, work quickly (but do not rush!) and return the tube into container. Remember that your fingers are still absorbing the radiation even though your body is behind the shield. The clattering counter will encourage you to work quickly.
- **Handle the material confidently and cautiously**. Avoid accidental splashes and spills. If these occur, clean up promptly. Check often with counter.
- **Check your gloves often.** Change when in doubt. Err on the side of wastefulness.

7. Clean up any contaminations promptly.

Use the Radiac detergent solution to wipe up the spill. Throw the contaminated paper towels into the large radioactive waste box. Repeat until the counter reads only background levels.

8. When the experiment is complete, scan the entire work area again.

This practice protects the next user, who could be you.

9. Scan yourself.

Check gloves, coat sleeves, your lab coat, pants and don't forget your shoes as well. This practice protects you and lab members outside the radiation room. Also scan any items or reagents that you are taking out of the room, such as film cassettes or buffer bottles.

10. Throw away your gloves.

Into appropriate bins.

11. You can now leave the radiation room. Turn off the lights and lock the room if there are no further users.

Handling 32P

1. The radio-material will come in a yellow lead container. It is cylindrical and very heavy.

2. When first delivered, bring it into the radiation room. Open the box and remove the packing material and notes. Be aware that the lead container is heavy, do not let it drop! Store the lead container at 4 deg C.

3. Once you are ready to use the 32P, take the lead container out of the fridge and place behind the acrylic shield. Set up the Geiger counter as usual.

4. Break the paper seal on the side holding the cap of the lead container. Turn anti-clockwise to loosen the cap until it stops turning. Now lift the cap away. Note: The cap is heavy!

5. Loosen the internal plastic cap by turning anticlockwise. Lift the cap and place it inside facing up. You will see a splash-guard with a depression in the middle. The Geiger counter will start to clatter.

6. Now use a short 10ul filter-tip (only use the short one! The long one may splinter) and jab it firmly (not too hard) into the depression to loosen the splash guard. Lift vertically, and shoot the tip+splash-guard into radioactive waste bin. Note: The Geiger counter will scream like mad. Be prepared for the sound.

7. Prepare your sample tubes on acrylic box and open their caps. Draw the required amount of 32P (red color liquid) and transfer into your sample tubes. Work quickly to minimize exposure. Close caps and cover the acrylic box.

8. Re-cap the plastic cap and lead cap on the yellow container. Check the lead container and work area for any splash contamination with the Geiger counter.

9. If completely used up, put the lead container into the radiation waste area. Otherwise, return the container to 4 deg C.

Part II. BAC High-density Filter Screening

Overview

Here are the recommended reagent amounts:

Number of filters per library = 8 or more

Amount of initial [Gamma32P]-ATP needed per filter = 1.0ul (10uCi)

Exposure time = 1-3 hours using storage phosphor screen and Typhoon phosphoimager.

Approx. 1 hour exposure per 10 000 cpm (measured 1 cm above the filter).

1. **5' End Labeling reaction**

   Dilute the oligo probes (~30bp desalt quality) to 5 uM (usually 1:20 dilution).

   Reaction Mix

	Per 10ul
Nuclease-free water (Ambion)	2

Probe (5 uM)	1
[Gamma-32P]-ATP	5
10X Kinase Buffer (NEB)	1
T4 Polynucleotide Kinase (Ambion)	1

(inside the radioactive room)

- gently mix
- incubate at 37C for 1 h

2. Preparation of NucAway spin columns

- tap column to settle dry gel
- hydrate using 650ul nuclease-free water
- cap, vortex, tap out air bubbles and leave at room temperature 5 – 15 min
- can be stored up to 3 days at 4C if needed
- spin column at 750g for 2 min (put into elute tube)
- check orientation
- discard elute tube
- apply sample directly to centre of gel bed (don’t touch sides or gel surface)
- place column in collection tube, using the same orientation as the first spin
- spin column at 750g for 2 min
- discard spin column into radioactive waste container
- store sample at -20C (radioactive room freezer)

3. Scintillation counter (optional step)

- 1ul of labeled probe + 1ul of scintillation fluid in an eppendorf tube

- go to level 4 scintillation counter, select 32P option

- obtain 1 minute average

- typical readings about 1 million counts per minute (cpm)

4. Hybridization using UltraHyb-Oligo (Ambion)

- preheat a 125ml pack to 55C for 5 min to dissolve precipitated materials

- take the BAC high-density filters and separate with a piece of nylon mesh between 2 filters

- up to 4 filters (+3 mesh) can fit into one long hybridization bottle

- add a minimum of 50ml of UltraHyb into bottle (for 4 filters, add 80ml)

- set hybridization oven to 42C, prehyb the blot for 30 min with low rotisserie speed

- decant the hyb solution into a 50ml falcon tube

- add ~1 million cpm/ml (final conc) of labeled probes from Step 2 into falcon tube

Usually 10ul for 8 filters, 15ul for 13 filters (at least 1ul per filter)

- cap the falcon tube and invert a few time to mix

- add the hot mixture back into the hyb bottle.

- hyb overnight in the oven at 42C for 14-24h

5. Washing and mounting the hot filters

- pour away the hot hyb buffer into liquid radiowaste bottle

- immediately add 50ml 1xSSC (+0.5%SDS) and wash in hyb oven 42C for 1 hour

- (optional) pour out and repeat the wash step

- pour out the 1xSSC and place on paper towels to absorb excess hot solution

- when sufficiently dry, prepare transparent plastic sheets

- mount the filter between two transparent plastic sheets

- store the mounted filters in an acrylic container at 4C.

6. Visualizing labeled filters in phosphoimager

-Collect large storage phosphor screens from Level 4

- Place hot filter on the velvet side of the X-ray cassette, with pencil marks/printed numbers facing up

- Position the phosphor screen between the enhancer (white surface) and the hot filter, with the white part of the phosphor facing the filter.

- Close the cassette and leave it for a duration dependent on this formula:

1 hour exp per 10 000 cpm at 1 cm distance.

Eg. A 3 000 cpm hot filter must be exposed for at least 3 hours 20 minutes

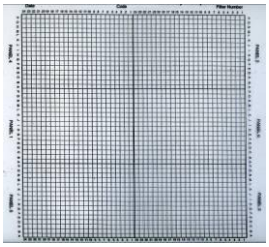
-open cassette, wipe clean and return the hot filter back into 4C.

- bring the phosphor screen down to Level 4 inside its cardboard box to use Typhoon

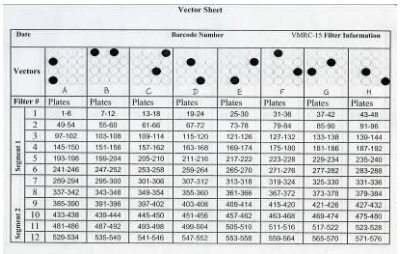
Part III. Protocol for Reading BAC IDs

The protocols for reading BAC clone identity are different for BACPAC or AGI/CUGI resources. Please refer to the company documentation for specific details.

BACPAC Protocol For Reading BAC IDs



Overlay -
Read Panel No., Coordinates



Vector Sheet

Date | Barcode Number | VMRC-15 Filter Information

Vectors	A	B	C	D	E	F	G	H
Filter #	Plates	Plates	Plates	Plates	Plates	Plates	Plates	Plates
Segment 1 — 1	1-6	7-12	13-18	19-24	25-30	31-36	37-42	43-48
2	49-54	55-60	61-66	67-72	73-78	79-84	85-90	91-96
3	97-102	103-108	109-114	115-120	121-126	127-132	133-138	139-144
4	145-150	151-156	157-162	163-168	169-174	175-180	181-186	187-192
5	193-198	199-204	205-210	211-216	217-222	223-228	229-234	235-240
6	241-246	247-252	253-258	259-264	265-270	271-276	277-282	283-288
Segment 2 — 7	289-294	295-300	301-306	307-312	313-318	319-324	325-330	331-336
8	337-342	343-348	349-354	355-360	361-366	367-372	373-378	379-384
9	385-390	391-396	397-402	403-408	409-414	415-420	421-426	427-432
10	433-438	439-444	445-450	451-456	457-462	463-468	469-474	475-480
11	481-486	487-492	493-498	499-504	505-510	511-516	517-522	523-528
12	529-534	535-540	541-546	547-552	553-558	559-564	565-570	571-576

Vector Sheet -
Filter No., Vector, Plate Range

Plate Locator -
Determine exact plate

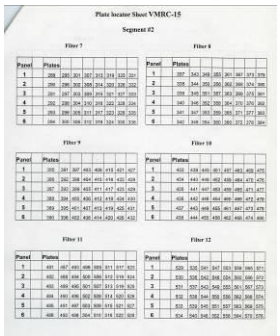
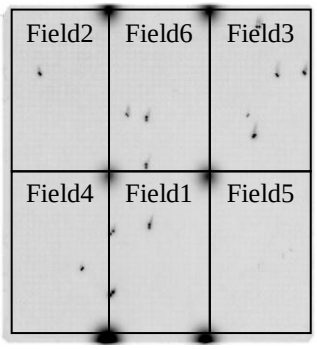


Plate locator Sheet VMRC-15

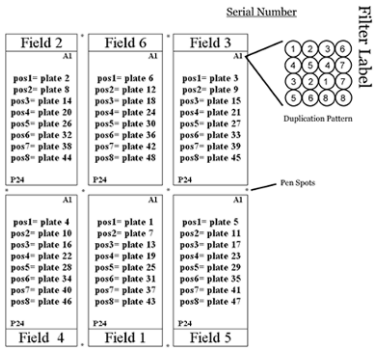
Segment #2

Filter 7 | Filter 8 | Filter 9 | Filter 10 | Filter 11 | Filter 12

AGI/CUGI Protocol For Reading BAC IDs

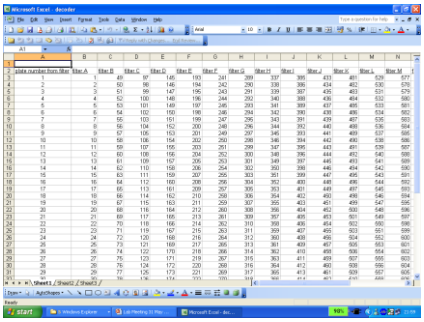


Grid Pattern – Field No., Coordinates



Field Chart –
Duplication Pattern, Plate Range

Decoder –
Filter No., Exact plate No.



Appendix C

Real-time PCR Protocol

This protocol has three parts:

IV. Preparing RNA

V. Preparing cDNA

VI. BioMark operation

Part I. RNA preparation

Harvesting cells

1. Wash the cells in 1 x PBS.
2. Add 1ml Trizol to dish/well. Pipette up and down to disperse cells.
3. Transfer into a 1.7ml microfuge tube.
4. Proceed to RNA extraction or store at -80C immediately.

RNA Extraction (adapted from Kevin's protocol)

1. Incubate at room temp for 5 min.
2. Add 200ul chloroform. Shake vigorously for 15 sec.
3. Incubate at room temp for 3 min.
4. Spin at 13000 rpm, 4C for 15 min.
5. The liquid will separate into two phases. Carefully pipette ~450ul of the top phase into a new tube, avoiding the protein interface.
6. Add 450ul 70% ethanol. Mix by inverting tube.
7. Apply 700ul to RNA kit column (from Qiagen RNeasy Mini Kit).
8. Spin at 13000 rpm, 15 sec (this and subsequent spin steps at room temp). Discard flow-through.

9. Add remainder from step 7 to column.
10. Spin at 13000 rpm, 15 sec. Discard flow-through.
11. Add 700ul buffer RW1 to column.
12. Spin at 13000 rpm, 15 sec. Discard collection tube.
13. Transfer column to new collection tube (provided by kit).
14. Add 500ul buffer RPE.
15. Spin at 13000 rpm, 15 sec. Discard flow-through.
16. Add 500ul buffer RPE.
17. Spin at 13000 rpm, 15 sec. Discard collection tube.
18. Transfer column to clean centrifuge tube (not provided by kit)
19. Dry the column by spinning 13000 rpm for 1 min. Discard tube.
20. Place column in a new microfuge tube for elution (provided by kit)
21. Add 30ul RNase-free water directly to the membrane. Let it stand for 1 min.
22. Spin at 13000 rpm, 1 min.
23. Repeat steps 21-22
24. Quantitate by Nanodrop
25. Store at -80C.

Check RNA yield and quality

1. Use 1.5ul per sample in the Nanodrop machine. A good yield of RNA should be ~ 100 – 1000 ng/ul depending on cell number.
2. For RNA, optimal 260/280 ratio is around **1.9**
3. For RNA, optimal 260/230 ratio is around **1.6**

4. Check the UV spectrum, if it there is a smooth curve peaking at 260 nm, the RNA quality is good.

Part II. cDNA preparation

Reverse Transcription
(Using ABI High-capacity cDNA Reverse Transcription Kit (4368813)

1. Prepare reactions on ice and remember to use filter tips.

2. Two sets of reactions to be made: RT mix and RNA mix.

3. For the RT mix, prepare a master mix using the below ratios:

10× RT Buffer	2.0
25× dNTP Mix (100 mM)	0.8
10× RT Random Primers	2.0
MultiScribe™ Reverse Transcriptase	1.0
Nuclease-free H2O	4.2
Total per Reaction	10.0 ul

4. For the RNA mix, prepare individually the below:

RNA sample	**(1 ug** equivalent volume)
Nuclease-free H2O	Top up to 10 ul
Total per Reaction	10.0 ul

5. Pipette 10 µL of RT master mix into each well of 8-tube PCR strip or 96-well reaction plate.

6. Pipette 10 µL of RNA sample into each well, pipetting up

and down two times to mix.

7. Seal the plates or tubes.

8. Briefly centrifuge the plate or tubes to spin down the

contents and to eliminate any air bubbles.

9. Place the plate or tubes on ice until you are ready to load the

thermal cycler (set rxn volume to 20ul):

Temperature (°C)	25	37	85	4
Time	10 min	120 min	5 sec	∞

10. Store the cDNA at -20C.

Pooling TaqMan assays

1. Combine equal volumes of each 20X Taqman Gene Expression Assay, up to 100 assays (max. 48 for BioMark).For BioMark, add 10ul of each assay in a microfuge tube.
2. Dilute the pooled TaqMan assays using TE buffer such that each assay is at final concentration of 0.2X. For BioMark, add 520ul TE for final volume of 1ml.

Pre-Amplification

1. Prepare the preamplification mix using the below ratios:

TaqMan PreAmp Master Mix (2X)	5.0
Pooled assay mix (0.2X)	2.5
cDNA sample	2.5
Total per Reaction	10.0 ul

2. Place the 8-strip or plate into PCR machine with these cycle conditions:

Temperature (°C)	95	95	60	4
Time	10 min (hold)	15sec	4 min	∞
		└← 10 cycles ┘		

3. Upon completion, immediately place on ice.
4. Dilute 1:5 using TE buffer (Add 40ul TE buffer to each reaction)
5. Store at -20C.

Part III. BioMark operation

Assay and Sample Preparation

1. Remember to use filter tips.

2. Two sets of reactions to be made: Assays and Sample mix

3. For the Assays, prepare individually using the below ratios:

20X TaqMan Gene Expression Assay	2.5
DA Assay Loading reagent	2.5
Total volume per Reaction	5.0

4. For the Sample, prepare a master mix using these ratios:

TaqMan Universal PCR Master Mix	2.5
DA Sample Loading reagent	0.25
cDNA sample(from preAmp step)	2.25
Total volume per Reaction	5.0

5. Vortex briefly, spin down and place these reactions on ice while priming the chip.

Priming the Chip

1. Open a new pack.

2. **Caution!** Use the chip within 24 hours of opening.

3. Using the syringe provided, inject control line fluid into each of the two accumulators on the chip.

4. Remember to insert the needle all the way into the gasket. Fill up to the lowest mark on the accumulator well.

5. **Caution!** Do not spill control line fluid on any other part of the chip.

6. Place the chip into NanoFlex IFC controller. Note the A1 position.

7. Press “Admin”, the password is admin.

8. Select 113x Chip Prime script to run (approx 10 min to complete)

9. **Caution!** Load the chip within 30 min after priming.

Loading the Chip

1. Remove primed chip and peel the protective blue film from bottom of chip.

2. Place the chip on the black-coloured work station.

3. Load the Assays using a short 10ul filter tip. Assay inlets are on the left side of the chip. Do not go pass the first stop on the pipette, avoid introducing bubbles. It's OK to load slightly less than 5ul.
4. Load the Samples in the same manner. Sample inlets are on the right side of the chip.
5. Place the chip into NanoFlex IFC controller.
6. This time select the sample loading script to run (approx 1 hour to complete)
7. **Caution!** Start the run within 4 hours after loading.

Power Up the BioMark Instrument

1. Press the big round button (jacketed by octagonal plastic) on the left side of the instrument.
2. Press the square button on the right side.
3. Press the green switch on the left side.
4. Click the BioMark Data Collection Software icon to launch the software.
5. Check the status bar to make sure that the cooling process has started (approx 1 hour to complete)
6. Instrument will be ready when it is cooled to -5C. Begin the run within an hour of that.

Starting the run

1. Remove loaded chip and remove any dust over the top of chip centre using a small piece of tape.
2. Place the chip into the BioMark Instrument loading tray. Note the A1 position.
3. Click Start

4. Click Load Chip
5. Type the barcode number
6. Browse to file location for saving data.
7. Application type: Gene Expression
8. Assay: Single probe
9. Browse to find thermal protocol file (only one available)
10. Select Auto Exposure, Passive Reference ROX, Probe Type FAM-MGB.
11. Verify chip run info.
12. Start Chip Run (approx 3 hours to complete)

Data analysis

1. Click the BioMark Real-Time PCR Analysis software icon to launch software.
2. Click Open Chip Run
3. Analysis setting is Auto (Detectors). Click Analyze.
4. Select Sample Setup to label your Samples.
5. Select Detector Setup to label your Assays.
6. Export data as a .csv file. Transfer file out using a thumbdrive.

Copy and paste wholesale into Cell A1 of the "BioMark_16x48_version2" Excel template made by Andrew and Lee Thean.

Appendix D

Oct4 DBD VP16 / EnR Microarray Results

List of genes altered by 1.5 fold either up or down				
Mouse VP16 up				
Gene Name	Common	Genbank	Product	RefSeq
scl41083.21_226-S	Mtmr4	NM_133215	myotubularin related protein 4	NM_133215
scl083456.10_6-S	Mov10l1	NM_031260	Moloney leukemia virus 10-like 1	NM_031260
scl067106.7_0-S	Arch	NM_025970	zinc finger and BTB domain containing 8 opposite strand	NM_025970
scl0004065.1_58-S	Abhd1	NM_021304	abhydrolase domain containing 1	NM_021304
scl27081.7_488-S	2610019P18Rik	NM_178612	hypothetical protein LOC66455	NM_178612
scl013627.1_210-S	Eef1a1	XM_203909	eukaryotic translation elongation factor 1 alpha 1	
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl0003019.1_66-S	9530090G24Rik	NM_145537	putative alpha-mannosidase	NM_145537
scl37811.6.1_5-S	Gstt1	NM_008185	glutathione S-transferase, theta 1	NM_008185
scl0077771.2_16-S	A330102K23Rik	NM_153409	TGF-beta induced apotosis protein 2	NM_153409
scl0001589.1_43-S	Pnpo	NM_134021	pyridoxine 5'-phosphate oxidase	NM_134021
scl51379.4_394-S	2010002N04Rik	NM_134133	putative small membrane protein NID67	NM_134133
GI_27370425-S				
scl0003642.1_1-S	Ercc8	NM_028042	excision repaiross-complementing rodent repair deficiency, complei	NM_028042
scl54338.5_728-S	9430034D17Rik	NM_029891	hypothetical protein LOC77286	NM_029891
scl24282.10_504-S	Ltb4dh	NM_025968	leukotriene B4 12-hydroxydehydrogenase	NM_025968
scl39235.7_89-S	Arhgdia	NM_133796	Rho GDP dissociation inhibitor (GDI) alpha	NM_133796
scl53716.6_16-S	DXBwg1396e	NM_029836	nucleolar TGF-beta1 target protein isoform a	NM_029836
scl000025.1_30-S	4930584N22Rik	NM_026654	TOE1 homolog	NM_026654
GI_52421325-S				
Mouse VP16 down				
Gene Name	Common	Genbank	Product	RefSeq
scl0017289.2_279-S	Mertk	NM_008587	c-mer proto-oncogene tyrosine kinase	NM_008587
scl31801.4.1_81-S	Il11	NM_008350	interleukin 11	NM_008350
scl41500.10.1_4-S	Cias1	NM_145827	cold autoinflammatory syndrome 1 homolog	NM_145827
scl016582.1_68-S	Kifc3	NM_010631	kinesin family member C3	NM_010631
scl47450.4_93-S	Hoxc6	NM_010465	homeobox C6	NM_010465
scl0020563.2_90-S	Slit2	NM_178804	slit homolog 2	NM_178804
scl0021826.1_238-S	Thbs2	NM_011581	thrombospondin 2	NM_011581
GI_75677459-S				
scl0001310.1_38-S	Ccng1	NM_009831	cyclin G1	NM_009831
scl012696.5_98-S	Cirbp	NM_007705	cold inducible RNA binding protein	NM_007705
GI_60678283-S				
scl023886.1_155-S	Gdf15	NM_011819	growth differentiation factor 15	NM_011819
scl47445.7_14-S	Copz1	NM_019817	coatomer protein complex, subunit zeta 1	NM_019817
scl21892.2.1_81-S	1110055J05Rik	NM_026394	hypothetical protein LOC67828	NM_026394
scl0056068.2_195-S	Ammecr1	NM_019496	AMMECR1 protein	NM_019496
scl000155.1_43-S	Lin7b	NM_011698	lin 7 homolog b	NM_011698
scl53162.3.1_182-S	Gpr120	NM_181748	G protein-coupled receptor 120	NM_181748
scl0230868.3_5-S	BC055811	NM_198610	immunoglobin superfamily, member 21	NM_198610
GI_85702174-A				
scl0002693.1_3-S	Mutyh	NM_133250	mutY homolog	NM_133250
scl0003546.1_11-S	Scotin	NM_026381	scotin isoform 1	NM_025858
scl23469.10.1_15-S	Hkr3	NM_133879	GLI-Kruppel family member HKR3	NM_133879
scl42956.7_401-S	Jundm2	NM_030887	Jun dimerization protein 2	NM_030887
scl54930.9.1_1-S	Hprt1	NM_013556	hypoxanthine guanine phosphoribosyl transferase 1	NM_013556
scl021809.2_58-S	Tgfb3	NM_009368	transforming growth factor, beta 3	NM_009368
scl24121.4.1_71-S	Cdkn2a	NM_009877	cyclin-dependent kinase inhibitor 2A	NM_009877
scl26371.6_22-S	Cxcl10	NM_021274	chemokine (C-X-C motif) ligand 10	NM_021274
gi_7305154_ref_NM_(	Hprt1	NM_013556	hypoxanthine guanine phosphoribosyl transferase 1	NM_013556
scl0001034.1_561-S	Crbn	NM_021449	cereblon isoform 1	NM_021449
scl0321000.2_14-S	4933421E11Rik	NM_177309	receptor-interacting factor 1	NM_177309
scl33485.7_10-S	Herpud1	NM_022331	homocysteine-inducible, endoplasmic reticulum stress-inducible, ul	NM_022331
scl0002582.1_20-S		NM_178718		
scl0103889.1_67-S	Hoxb2	NM_134032	homeo box B2	NM_134032
scl19424.2.1_75-S		XM_149113		
scl074761.10_0-S	1200013A08Rik	NM_024263	limitrin	NM_024263
scl019073.1_109-S	Prg1	NM_011157	proteoglycan 1, secretory granule	NM_011157
scl24606.5_24-S	1200013A08Rik	NM_024263	limitrin	NM_024263
scl37309.10_320-S	Mmp13	NM_008607	matrix metalloproteinase 13	NM_008607
scl0016998.2_125-S	Ltbp3	NM_008520	latent transforming growth factor beta binding protein 3	NM_008520
GI_22129490-S				
scl0003434.1_18-S	Mmp13	NM_008607	matrix metalloproteinase 13	NM_008607
scl39273.6_263-S	Lgals3bp	NM_011150	lectin, galactoside-binding, soluble, 3 binding protein	NM_011150
scl0003597.1_94-S	Mmp3	NM_010809	matrix metalloproteinase 3	NM_010809
scl37305.10.1_6-S	Mmp10	NM_019471	matrix metalloproteinase 10	NM_019471
scl030839.9_291-S	Fbxw5	NM_013908	F-box and WD-40 domain protein 5	NM_013908
scl18860.2_108-S	Grem1	NM_011824	gremlin 1	NM_011824
scl42842.7.1_0-S	D12Ertd647e	NM_026790	hypothetical protein LOC52668 isoform 1	NM_026790
scl053606.2_17-S	G1p2	NM_015783	interferon, alpha-inducible protein	NM_015783
scl020296.2_11-S	Ccl2	NM_011333	chemokine (C-C motif) ligand 2	NM_011333
scl37307.8.1_29-S	Mmp3	NM_010809	matrix metalloproteinase 3	NM_010809
scl016145.5_111-S	Igtp	NM_018738	interferon gamma induced GTPase	NM_018738
scl0001526.1_22-S	Irf1	NM_008390	interferon regulatory factor 1	NM_008390
scl29554.6.1_30-S	Usp18	NM_011909	ubiquitin specific protease 18	NM_011909

Mouse EnR up				
Gene Name	Common	Genbank	Product	RefSeq
scl26197.25_374-S	Rutbc2	NM_172718	RUN and TBC1 domain containing 2	NM_172718
scl083456.10_6-S	Mov10l1	NM_031260	Moloney leukemia virus 10-like 1	NM_031260
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl40901.16_69-S	Stat5a	NM_011488	signal transducer and activator of transcription 5A	NM_011488
scl013176.2_9-S		XM_147280		
scl41083.21_226-S	Mtmr4	NM_133215	myotubularin related protein 4	NM_133215
scl0077771.2_16-S	A330102K23Rik	NM_153409	TGF-beta induced apotosis protein 2	NM_153409
scl28889.10.1_70-S	MGC59076	NM_178413	hypothetical protein LOC232078	NM_0010339
scl018642.4_28-S	Pfkm	NM_021514	phosphofructokinase, muscle	NM_021514
scl0004065.1_58-S	Abhd1	NM_021304	abhydrolase domain containing 1	NM_021304
scl013532.4_62-S	Dub2	NM_010089	deubiquitinating enzyme 2	NM_010089
scl0003023.1_84-S	Fnbp4	NM_018828	formin binding protein 4	NM_018828
scl51379.4_394-S	2010002N04Rik	NM_134133	putative small membrane protein NID67	NM_134133
scl54338.5_728-S	9430034D17Rik	NM_029891	hypothetical protein LOC77286	NM_029891
scl0099375.1_136-S	Cul4a	NM_146207	cullin 4A	NM_146207
scl0003019.1_66-S	9530090G24Rik	NM_145537	putative alpha-mannosidase	NM_145537
scl0093742.1_83-S	Pard3	NM_033620	partitioning-defective protein 3 homolog isoform 2	NM_0010135
scl0003972.1_558-S	Ankrd17	NM_030886	ankyrin repeat domain protein 17 isoform a	NM_030886
scl0003712.1_12-S	Tm4sf17	NM_028841	tetraspanin 17	NM_028841
scl017354.18_45-S	Mllt10	NM_010804	myeloid/lymphoid or mixed lineage-leukemia translocation to 10 hoı	NM_010804
scl50141.7.1_283-S	Bak1	NM_007523	BCL2-antagonist/killer 1	NM_007523
scl0004106.1_50-S	Usp46	NM_177561	ubiquitin specific protease 46	NM_177561
scl0002386.1_172-S	Hs1bp3	NM_021429	HS1-binding protein 3	NM_021429
scl40193.3.1_118-S	Hand1	NM_008213	heart and neural crest derivatives expressed transcript 1	NM_008213
scl00234736.1_63-S	Rfwd3	NM_146218	ring finger and WD repeat domain 3	NM_146218
scl0003331.1_104-S	Bcl2l1	NM_009743	Bcl2-like 1	NM_009743
scl000914.1_84-S	Cyp20a1	XM_129747	cytochrome P450, family 20, subfamily A, polypeptide 1	
scl0001298.1_25-S	1300013J15Rik	NM_026183	hypothetical protein LOC67473	NM_026183
scl014697.10_11-S	Gnb5	NM_010313	guanine nucleotide-binding protein, beta-5 subunit isoform 1	NM_010313
scl54853.6_573-S	Zfp275	NM_031494	Zinc finger protein 275	NM_031494
scl00102423.2_7-S	AA589481	NM_172162	MBD2 (methyl-CpG-binding protein)-interacting zinc finger protein	NM_172162
scl0001399.1_22-S	Prpsap2	NM_144806	phosphoribosyl pyrophosphate synthetase-associated protein 2	NM_144806
scl00209318.1_9-S	Gps1	NM_145370	G protein pathway suppressor 1	NM_145370
scl50934.9.2_2-S	Pacsin1	NM_011861	protein kinase C and casein kinase substrate in neurons 1	NM_011861
scl38890.17_310-S	P4ha1	NM_011030	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydrox	NM_011030
scl065246.3_62-S	Xpo7	NM_023045	exportin 7	NM_023045
Mouse EnR down				
Gene Name	Common	Genbank	Product	RefSeq
scl15754.9.1_26-S	Rcor3	NM_144814	REST corepressor 3	NM_144814
scl0003434.1_18-S	Mmp13	NM_008607	matrix metalloproteinase 13	NM_008607
scl25019.5.1_161-S	Edn2	NM_007902	endothelin 2	NM_007902
scl00013.1_82-S	Arfip2	NM_029802	ADP-ribosylation factor interacting protein 2	NM_029802
scl21680.4_681-S	Kcnc4	NM_145922	potassium voltage gated channel, Shaw-related subfamily, member	NM_145922
scl39973.20.1_51-S	4933427D14Rik	NM_028963	hypothetical protein LOC74477	NM_028963
scl0014664.2_299-S	Slc6a9	NM_008135	solute carrier family 6 member 9	NM_008135
scl17644.2.1_29-S	Twist2	NM_007855	twist homolog 2	NM_007855
scl42842.7.1_0-S	D12Ertd647e	NM_026790	hypothetical protein LOC52668 isoform 1	NM_026790
scl0001660.1_2-S	Jmjd2b	NM_172132	jumonji domain containing 2B	NM_172132
scl16598.23.1_93-S	Ptprn	NM_008985	protein tyrosine phosphatase, receptor type, N	NM_008985
scl29554.6.1_30-S	Usp18	NM_011909	ubiquitin specific protease 18	NM_011909
scl0054635.2_40-S	Pdgfc	NM_019971	platelet-derived growth factor, C polypeptide	NM_019971
scl32353.2.4_40-S	Ndufc2	NM_024220	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 2	NM_024220
scl072185.1_8-S	2810427I04Rik	NM_028146	hypothetical protein LOC72185	NM_028146
GI_71274129-S				
GI_22129490-S				
scl0053382.2_104-S	Txnl1	NM_016792	thioredoxin-like 1	NM_016792
scl0003597.1_94-S	Mmp3	NM_010809	matrix metalloproteinase 3	NM_010809
scl37305.10.1_6-S	Mmp10	NM_019471	matrix metalloproteinase 10	NM_019471
scl19424.2.1_75-S		XM_149113		
scl18860.2_108-S	Grem1	NM_011824	gremlin 1	NM_011824
scl37309.10_320-S	Mmp13	NM_008607	matrix metalloproteinase 13	NM_008607
scl37307.8.1_29-S	Mmp3	NM_010809	matrix metalloproteinase 3	NM_010809
scl030839.9_291-S	Fbxw5	NM_013908	F-box and WD-40 domain protein 5	NM_013908

Human VP16 up				
Gene Name	Common	Genbank	Product	RefSeq
scl067106.7_0-S	Arch	NM_025970	zinc finger and BTB domain containing 8 opposite strand	NM_025970
scl013627.1_210-S	Eef1a1	XM_203909	eukaryotic translation elongation factor 1 alpha 1	
scl017354.18_45-S	Mllt10	NM_010804	myeloid/lymphoid or mixed lineage-leukemia translocation to 10 hoı	NM_010804
scl018715.2_20-S	Pim2	NM_138606	serine-threonine protein kinase pim-2 isoform 1	NM_138606
scl083456.10_6-S	Mov10l1	NM_031260	Moloney leukemia virus 10-like 1	NM_031260
scl39640.3.1_127-S	1700001P01Rik	XM_126645	hypothetical protein LOC72215	
scl20206.6_190-S	Dstn	NM_019771	destrin	NM_019771
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl0001287.1_52-S	Ascc2	NM_029291	ASC-1 complex subunit P100	NM_029291
scl0001399.1_22-S	Prpsap2	NM_144806	phosphoribosyl pyrophosphate synthetase-associated protein 2	NM_144806
scl24987.27.1_4-S	Inpp5b	NM_008385	inositol polyphosphate-5-phosphatase B	NM_008385
scl015574.1_13-S	Hus1	NM_008316	Hus1 homolog	NM_008316
scl51147.8.1_46-S	T	NM_009309	brachyury	NM_009309
scl0214987.1_149-S	5830457O10Rik	NM_145412	hypothetical protein LOC214987	NM_145412
scl43616.1.2_258-S	1700024P04Rik	XM_127485	hypothetical protein LOC69382	XM_127485
scl0003019.1_66-S	9530090G24Rik	NM_145537	putative alpha-mannosidase	NM_145537
scl013176.2_9-S		XM_147280		
scl27267.9.1_12-S	Myl2	NM_010861	myosin, light polypeptide 2, regulatory, cardiac, slow	NM_010861
scl0001707.1_32-S	Axin1	XM_128515	axin 1	
scl074143.9_44-S	Opa1	NM_133752	optic atrophy 1 homolog	NM_133752
scl0001589.1_43-S	Pnpo	NM_134021	pyridoxine 5'-phosphate oxidase	NM_134021
scl52796.6.1_4-S	Acy3	NM_027857	aspartoacylase-3	NM_027857
scl060365.6_119-S	Rbm8a	NM_025875	RNA binding motif protein 8a	NM_025875
scl54338.5_728-S	9430034D17Rik	NM_029891	hypothetical protein LOC77286	NM_029891
scl020969.4_58-S	Sdc1	NM_011519	syndecan 1	NM_011519
scl0399675.1_18-S	Tdpoz4	NM_207272	TD and POZ domain containing 4	NM_207272
scl0026442.1_315-S	Psma5	NM_011967	proteasome (prosome, macropain) subunit, alpha type 5	NM_011967
Human VP16 down				
Gene Name	Common	Genbank	Product	RefSeq
scl0002109.1_724-S	Dnajb4	NM_025926	DnaJ (Hsp40) homolog, subfamily B, member 4	NM_025926
scl067412.7_37-S	6330407J23Rik	NM_026138	hypothetical protein LOC67412	NM_026138
scl42956.7_401-S	Jundm2	NM_030887	Jun dimerization protein 2	NM_030887
scl0001595.1_53-S	9530058B02Rik	NM_026633	hypothetical protein LOC68241	NM_026633
scl47963.9.1_68-S	Fzd6	NM_008056	frizzled 6	NM_008056
scl0017082.1_132-S	Il1rl1	NM_010743	interleukin 1 receptor-like 1 isoform a	NM_001025E
scl0014165.1_114-S	Fgf10	NM_008002	fibroblast growth factor 10	NM_008002
scl000096.1_251-S	1600012H06Rik	NM_026451	hypothetical protein LOC67912	NM_026451
scl00108086.1_95-S	Ubce7ip1	NM_080561	ubiquitin conjugating enzyme 7 interacting protein 1	NM_080561
scl053606.2_17-S	G1p2	NM_015783	interferon, alpha-inducible protein	NM_015783
scl16260.3.61_18-S	Elf3	NM_007921	E74-like factor 3	NM_007921
scl0230868.3_5-S	BC055811	NM_198610	immunoglobin superfamily, member 21	NM_198610
scl16598.23.1_93-S	Ptprn	NM_008985	protein tyrosine phosphatase, receptor type, N	NM_008985
scl42430.2_236-S	Foxa1	NM_008259	forkhead box A1	NM_008259
scl39634.14_208-S	Plxdc1	NM_028199	plexin domain containing 1	NM_028199
scl25795.24.1_9-S	2210010N04Rik	XM_149712	hypothetical protein LOC70381 isoform 1	XM_149712
scl17475.2.1_264-S	4930429O20Rik	NM_029025	hypothetical protein LOC74626	NM_029025
scl30446.19.1_20-S	9130012B15Rik	NM_030221	NAD synthetase 1	NM_030221
scl0020698.1_299-S	Sphk1	NM_011451	sphingosine kinase 1 isoform 1	NM_011451
scl016145.5_111-S	Igtp	NM_018738	interferon gamma induced GTPase	NM_018738
gi_7305154_ref_NM_0	Hprt1	NM_013556	hypoxanthine guanine phosphoribosyl transferase 1	NM_013556
scl0002041.1_2-S	5830417I10Rik	XM_194592	hypothetical protein LOC76022 isoform 1	XM_194592
scl45278.7.1_11-S	Dnajd1	NM_025384	DnaJ (Hsp40) homolog, subfamily D, member 1	NM_025384
scl0003523.1_1-S	BC021438	NM_145416	hypothetical protein LOC215194	NM_145416
scl00268482.1_212-S	Krt1-12	NM_010661	keratin complex 1, acidic, gene 12	NM_010661
scl33485.7_10-S	Herpud1	NM_022331	homocysteine-inducible, endoplasmic reticulum stress-inducible, ut	NM_022331
scl39273.6_263-S	Lgals3bp	NM_011150	lectin, galactoside-binding, soluble, 3 binding protein	NM_011150
scl37305.10.1_6-S	Mmp10	NM_019471	matrix metalloproteinase 10	NM_019471
scl39973.20.1_51-S	4933427D14Rik	NM_028963	hypothetical protein LOC74477	NM_028963
scl39344.12.1_117-S	Fdxr	NM_007997	ferredoxin reductase	NM_007997
scl00224014.1_30-S	Fgd4	NM_139232	FYVE, RhoGEF and PH domain containing 4 isoform alpha	NM_139232
scl37309.10_320-S	Mmp13	NM_008607	matrix metalloproteinase 13	NM_008607
scl0002451.1_30-S	5730502D15Rik	NM_026485	hypothetical protein LOC67976	NM_026485
scl47450.4_93-S	Hoxc6	NM_010465	homeobox C6	NM_010465
GI_46909570-S	Gata6	NM_010258	GATA binding protein 6	NM_010258
scl37307.8.1_29-S	Mmp3	NM_010809	matrix metalloproteinase 3	NM_010809
scl0003514.1_179-S	Pml	NM_008884	promyelocytic leukemia isoform 1	NM_008884
scl0001526.1_22-S	Irf1	NM_008390	interferon regulatory factor 1	NM_008390
scl42842.7.1_0-S	D12Ertd647e	NM_026790	hypothetical protein LOC52668 isoform 1	NM_026790
scl17780.6_86-S	Stk16	NM_011494	serine/threonine kinase 16	NM_011494
scl030839.9_291-S	Fbxw5	NM_013908	F-box and WD-40 domain protein 5	NM_013908
scl18860.2_108-S	Grem1	NM_011824	gremlin 1	NM_011824
scl29554.6.1_30-S	Usp18	NM_011909	ubiquitin specific protease 18	NM_011909

Human EnR up				
Gene Name	Common	Genbank	Product	RefSeq
scl0004206.1_0-S	BC034507	XM_131888	claudin 12	
scl51379.4_394-S	2010002N04Rik	NM_134133	putative small membrane protein NID67	NM_134133
scl53716.6_16-S	DXBwg1396e	NM_029836	nucleolar TGF-beta1 target protein isoform a	NM_029836
scl013627.1_210-S	Eef1a1	XM_203909	eukaryotic translation elongation factor 1 alpha 1	
scl41083.21_226-S	Mtmr4	NM_133215	myotubularin related protein 4	NM_133215
scl0003520.1_0-S	Dibd1	NM_133981	disrupted in bipolar disorder 1 homolog	NM_133981
scl072139.1_117-S	2610044O15Rik	NM_153780	hypothetical protein LOC72139	NM_153780
scl27081.7_488-S	2610019P18Rik	NM_178612	hypothetical protein LOC66455	NM_178612
scl54853.6_573-S	Zfp275	NM_031494	Zinc finger protein 275	NM_031494
scl42351.23.1_41-S	4930447C04Rik	NM_029444	Six6 opposite strand transcript 1	NM_029444
scl0399675.1_18-S	Tdpoz4	NM_207272	TD and POZ domain containing 4	NM_207272
scl0002074.1_21-S	Bnipl	NM_134253	BNIP-2 similar	NM_134253
scl17780.6_86-S	Stk16	NM_011494	serine/threonine kinase 16	NM_011494
scl0003316.1_65-S	1810020C19Rik	XM_130317	hypothetical protein LOC69113 isoform 1	XM_130317
scl41295.13.702_28-S	P2rx5	NM_033321	purinergic receptor P2X5	NM_033321
scl17438.3_636-S	Lmod1	NM_053106	leiomodin 1 (smooth muscle)	NM_053106
scl50962.10_636-S	Tmem8	NM_021793	transmembrane protein 8 (five membrane-spanning domains)	NM_021793
scl0000117.1_15-S	Taf6	NM_009315	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associat	NM_009315
scl0067845.2_211-S	Zfp364	NM_026406	Rabring 7	NM_026406
scl47241.11_30-S	4921532K09Rik	NM_026149	chronic myelogenous leukemia tumor antigen 66	NM_026149
scl17475.2.1_264-S	4930429O20Rik	NM_029025	hypothetical protein LOC74626	NM_029025
scl071275.2_75-S	4933437F05Rik	XM_127023	hypothetical protein LOC71275	XM_127023
scl00278672.2_86-S	1110051B16Rik	NM_183389	hypothetical protein LOC278672	NM_183389
scl083456.10_6-S	Mov10l1	NM_031260	Moloney leukemia virus 10-like 1	NM_031260
scl0022070.1_31-S	Tpt1	NM_009429	tumor protein, translationally-controlled 1	NM_009429
scl36160.67.1_2-S	Col5a3	NM_016919	procollagen, type V, alpha 3	NM_016919
scl0014972.1_210-S	H2-K1	NM_001001E	histocompatibility 2, K1, K region	NM_001001E
scl0003155.1_68-S				
scl015371.3_18-S	Hmx1	NM_010445	H6 homeo box 1	NM_010445
scl067106.7_0-S	Arch	NM_025970	zinc finger and BTB domain containing 8 opposite strand	NM_025970
scl0021974.1_259-S	Top2b	NM_009409	topoisomerase (DNA) II beta	NM_009409
scl45932.17_232-S	Tm9sf2	NM_080556	transmembrane 9 superfamily member 2	NM_080556
scl00268482.1_212-S	Krt1-12	NM_010661	keratin complex 1, acidic, gene 12	NM_010661
scl0070373.1_243-S	1700020O03Rik	NM_027405	hypothetical protein LOC70373	NM_027405
scl0001660.1_2-S	Jmjd2b	NM_172132	jumonji domain containing 2B	NM_172132
scl00230259.2_287-S	E130308A19Rik	NM_153158	hypothetical protein LOC230259 isoform 2	NM_001015E
scl33556.12_71-S	Gpt2	NM_173866	glutamic pyruvate transaminase (alanine aminotransferase) 2	NM_173866
scl40982.2.1_13-S	Hoxb1	NM_008266	homeobox B1	NM_008266
scl15995.6_195-S	C130085G02Rik	XM_136364	dual specificity phosphatase 27 (putative)	
scl0018378.2_220-S	Omp	NM_011010	olfactory marker protein	NM_011010
scl45217.22.1_37-S	2810028N01Rik	NM_028315	RIKEN cDNA 2810028N01	NM_028315
scl0067475.1_65-S	1300013B24Rik	NM_026184	endoplasmic oxidoreductase 1 beta	NM_026184
scl020510.12_56-S	Slc1a1	NM_009199	solute carrier family 1 (neuronal/epithelial high affinity glutamate tra	NM_009199
scl00103537.2_139-S	Mbtd1	NM_134012	mbt domain containing 1	NM_134012
scl50188.17.10_115-S	1200003M09Rik	NM_027880	hypothetical protein LOC71718	NM_027880
scl0227634.1_319-S	Camsap1	XM_129375	calmodulin regulated spectrin-associated protein 1 isoform 1	XM_129375
scl25657.15_226-S	Cbfa2t1h	NM_009822	CBFA2T1 identified gene homolog	NM_009822
scl39991.5_43-S	Spag7	NM_172561	sperm associated antigen 7	NM_172561
scl0001234.1_2-S	Slc37a3	NM_028123	solute carrier family 37 (glycerol-3-phosphate transporter), member	NM_028123
scl0003019.1_66-S	9530090G24Rik	NM_145537	putative alpha-mannosidase	NM_145537
scl00108686.2_43-S	A430106J12Rik	NM_176841	Hook-related protein 1	NM_176841
scl16771.16.1_41-S	Pms1	NM_153556	postmeiotic segregation increased 1	NM_153556
scl0020843.2_126-S	Stag2	NM_021465	stromal antigen 2	NM_021465
scl26197.25_374-S	Rutbc2	NM_172718	RUN and TBC1 domain containing 2	NM_172718
scl0026442.1_315-S	Psma5	NM_011967	proteasome (prosome, macropain) subunit, alpha type 5	NM_011967
scl000319.1_7-S	Acin1	NM_019567	apoptotic chromatin condensation inducer 1 isoform 1	NM_019567
scl0077771.2_16-S	A330102K23Rik	NM_153409	TGF-beta induced apotosis protein 2	NM_153409
scl0098733.2_280-S	AW822216	NM_178884	hypothetical protein LOC98733	NM_178884
scl0068149.1_319-S	Otub2	NM_026580	OTU domain, ubiquitin aldehyde binding 2	NM_026580
scl00244891.1_136-S		NM_175536		
scl050505.11_70-S	Ercc4	NM_015769	excision repair cross-complementing rodent repair deficiency, comp	NM_015769
scl0015519.1_14-S	Hspca	NM_010480	heat shock protein 1, alpha	NM_010480
scl00280487.1_157-S		NM_183299		
scl0109154.1_152-S	2410014A08Rik	NM_175403	hypothetical protein LOC109154	NM_175403
scl00209318.1_9-S	Gps1	NM_145370	G protein pathway suppressor 1	NM_145370
scl22236.8_4-S	Ccna2	NM_009828	cyclin A2	NM_009828
scl0001399.1_22-S	Prpsap2	NM_144806	phosphoribosyl pyrophosphate synthetase-associated protein 2	NM_144806
scl53489.23_479-S	Pacs1	XM_283545	phosphofurin acidic cluster sorting protein 1	
scl47278.57_323-S	4432411E13Rik	XM_196130	extraembryonic development protein isoform 1	XM_196130
scl29111.14_3-S	Slc37a3	NM_028123	solute carrier family 37 (glycerol-3-phosphate transporter), member	NM_028123
scl0012928.2_60-S	Crk	NM_133656	v-crk sarcoma virus CT10 oncogene homolog	NM_133656
scl0001298.1_25-S	1300013J15Rik	NM_026183	hypothetical protein LOC67473	NM_026183
scl39506.28.1461_10	Slc4a1	NM_011403	solute carrier family 4 (anion exchanger), member 1	NM_011403
scl19043.4.1_203-S	Pramel7	NM_178250	preferentially expressed antigen in melanoma like 7	NM_178250
scl53821.10.1_71-S	1700008I05Rik	XM_136071	t-complex 11 protein	
scl24103.5_0-S	5830433M19Rik	NM_026368	hypothetical protein LOC67770	NM_026368
scl30446.19.1_20-S	9130012B15Rik	NM_030221	NAD synthetase 1	NM_030221
scl0232821.5_303-S	BC018462	NM_146178	hypothetical protein LOC232821	NM_146178
scl011532.9_123-S	Adh5	NM_007410	alcohol dehydrogenase 5 (class III), chi polypeptide	NM_007410
scl18712.18.1_54-S	Brrn1	NM_144818	barren homolog	NM_144818
scl0093762.1_285-S	Smarca5	NM_053124	SWI/SNF related, matrix associated, actin dependent regulator of c	NM_053124

scl35222.4_425-S	Myd88	NM_010851	myeloid differentiation primary response gene 88	NM_010851
scl0001405.1_534-S	Spag9	NM_027569	sperm associated antigen 9 isoform 2	NM_0010254
scl50156.18_539-S		NM_153140		
scl000781.1_68-S	Kif21b	NM_019962	kinesin family member 21B	NM_019962
scl0017156.2_32-S	Man1b	NM_010763	mannosidase, alpha, class 1A, member 2	NM_010763
scl26728.9.1_113-S	Rbks	NM_153196	ribokinase	NM_153196
scl068280.6_0-S		XM_131189		
scl000041.1_12_REV	Rad9	NM_011237	RAD9 homolog	NM_011237
scl00267019.1_256-S	Rps15a	NM_170669	ribosomal protein S15a	NM_170669
scl0277333.1_280-S	MGC68323	NM_199472	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)-like	NM_199472
scl50308.28_413-S	Map3k4	NM_011948	mitogen activated protein kinase kinase kinase 4	NM_011948
scl44299.4.1_276-S	Chrm3	NM_033269	cholinergic receptor, muscarinic 3, cardiac	NM_033269
scl38890.17_310-S	P4ha1	NM_011030	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydrox	NM_011030
scl0022420.2_231-S	Wnt6	NM_009526	wingless-related MMTV integration site 6	NM_009526
scl077574.1_0-S	3321401G04Rik	XM_133096	hypothetical protein LOC77574	
scl053886.1_142-S	Cdkl2	NM_177270	cyclin-dependent kinase-like 2 (CDC2-related kinase) isoform 1	NM_016912
scl49449.1.769_0-S	Abat	NM_172961	4-aminobutyrate aminotransferase	NM_172961
scl00170460.2_242-S	Stard5	NM_023377	StAR-related lipid transfer protein 5	NM_023377
scl20206.6_190-S	Dstn	NM_019771	destrin	NM_019771
scl20194.20.1_79-S	Sec23b	NM_019787	SEC23B	NM_019787
scl069612.1_312-S	2310037I24Rik	NM_133714	hypothetical protein LOC69612	NM_133714
scl29880.12.1_23-S	0610039N19Rik	NM_026159	all-trans-13,14-dihydroretinol saturase	NM_026159
scl059014.1_56-S	Rrs1	NM_021511	RRS1 ribosome biogenesis regulator homolog	NM_021511
scl0016619.1_79-S	Klk27	NM_020268	kallikrein 27	NM_020268
scl37811.6.1_5-S	Gstt1	NM_008185	glutathione S-transferase, theta 1	NM_008185
scl36871.20.1_240-S	Parp6	XM_134863	poly (ADP-ribose) polymerase family, member 6 isoform 1	XM_134863
scl014050.16_119-S	Eya3	NM_010166	eyes absent 3 homolog isoform 2	NM_010166
scl0075302.2_295-S	Asxl2	NM_172421	polycomb group protein ASXH2 homolog	NM_172421
scl0018008.1_3-S	Nes	NM_016701	nestin	NM_016701
scl000935.1_50-S	LOC217536	XM_122407	similar to transcription factor B2, mitochondrial	XM_122407
scl00213464.2_151-S	Rbbp5	NM_172517	retinoblastoma binding protein 5	NM_172517
scl31678.22.1_32-S	Sfrs16	NM_016680	splicing factor, arginine/serine-rich 16 isoform L	NM_016680
scl36985.26_298-S	Usp28	NM_175482	ubiquitin specific protease 28	NM_175482
scl020755.2_79-S	Sprr2a	NM_011468	small proline-rich protein 2A	NM_011468
scl53989.25_581-S	Zfp261	NM_019831	zinc finger protein 261	NM_019831
scl41204.5.67_14-S	Unc119	NM_011676	UNC-119 homolog	NM_011676
scl0230848.1_316-S	BC059177	NM_198248	zinc finger and BTB domain containing 40	NM_198248
scl34926.9_67-S	3110010F15Rik	NM_026067	histone mRNA 3' end-specific exonuclease	NM_026067
scl011787.1_19-S	Apbb2	NM_009686	amyloid beta (A4) precursor protein-binding, family B, member 2	NM_009686
scl0194237.1_178-S	BC057371	NM_177572	hypothetical protein LOC194237	NM_177572
scl50796.3_29-S	Ly6g6c	NM_023463	lymphocyte antigen 6 complex G6C	NM_023463
scl29631.10_10-S	Creld1	NM_133930	cysteine-rich with EGF-like domains 1	NM_133930
scl0024061.1_292-S	Smc1l1	NM_019710	SMC1 structural maintenance of chromosomes 1-like 1	NM_019710
scl47063.13.1_72-S	Top1mt	XM_128145	DNA topoisomerase 1, mitochondrial	
scl0002244.1_36-S		XM_128857		
scl0072981.1_191-S	Prkrir	NM_028410	interferon-inducible, double stranded RNA dependent inhibitor, prote	NM_028410
scl47783.4_206-S	Zfp7	NM_145916	zinc finger protein 7	NM_145916
scl0073197.1_297-S	D19Ertd703e	NM_029456	SAPS domain family, member 3	NM_029456
scl39652.23.1463_24	Kpnb1	NM_008379	karyopherin (importin) beta 1	NM_008379
scl069178.1_5-S	Snx5	NM_024225	sorting nexin 5	NM_024225
scl24618.9_190-S	A530082C11Rik	NM_177186	solute carrier family 35, member E2	NM_177186
scl18851.3_427-S	6430601A21Rik	NM_175466	hypothetical protein LOC228491	NM_175466
scl0004065.1_58-S	Abhd1	NM_021304	abhydrolase domain containing 1	NM_021304
scl54338.5_728-S	9430034D17Rik	NM_029891	hypothetical protein LOC77286	NM_029891
scl30789.23.1_30-S	Copb1	NM_033370	coatomer protein complex, subunit beta 1	NM_033370
scl29845.15.1_28-S	Rtkn	NM_133641	rhotekin isoform 1	NM_009106
scl34187.26.1_95-S	Nup133	NM_172288	nucleoporin 133	NM_172288
scl48775.5_265-S	Emp2	NM_007929	epithelial membrane protein 2	NM_007929
scl37831.9_401-S	Tfam	NM_009360	transcription factor A, mitochondrial	NM_009360
scl017079.3_90-S	Ly78	NM_008533	CD180 antigen	NM_008533
scl015382.13_13-S	Hnrpa1	NM_010447	heterogeneous nuclear ribonucleoprotein A1	NM_010447
scl30495.11.1_64-S		XM_146151		
scl072313.2_0-S	2510002A14Rik	NM_028194	hypothetical protein LOC72313	NM_028194
scl019654.2_30-S	Rbm6	NM_011251	RNA binding motif protein 6 isoform a	NM_011251
scl16664.9_588-S	Lancl1	NM_021295	LanC (bacterial lantibiotic synthetase component C)-like 1	NM_021295
scl0003028.1_68-S		NM_026568		
scl38274.14.1_69-S	Si	NM_021882	silver	NM_021882
scl011702.3_204-S	Amd1	NM_009665	S-adenosylmethionine decarboxylase 1	NM_009665
scl28557.1_516-S	Srgap3	NM_153070	brain stress early protein Gbi isoform 1	NM_080448
scl0004152.1_41-S	Chfr	NM_172717	checkpoint with forkhead and ring finger domains	NM_172717
scl0224619.1_310-S	Traf7	NM_153792	Tnf receptor-associated factor 7	NM_153792
scl31153.10.1_27-S	Rlbp1	NM_020599	retinaldehyde binding protein 1	NM_020599
scl25667.3_320-S	C430048L16Rik	XM_355471	hypothetical protein LOC77604	
scl17620.15_302-S	Apg4b	NM_174874	autophagin 1	NM_174874
scl30692.5.1_111-S	Lat	NM_010689	linker for activation of T cells	NM_010689
scl00213391.1_239-S	Rassf4	NM_178045	Ras association (RalGDS/AF-6) domain family 4	NM_178045
scl027418.18_28-S	Mkln1	NM_013791	muskelin 1, intracellular mediator containing kelch motifs	NM_013791
scl074143.9_44-S	Opa1	NM_133752	optic atrophy 1 homolog	NM_133752
scl33570.4.1_25-S	Klf1	NM_010635	Kruppel-like factor 1 (erythroid)	NM_010635
scl000571.1_203-S	Cklfsf4	NM_153582	chemokine-like factor super family 4	NM_153582
scl42191.8.1_132-S	4930534B04Rik	XM_127104	hypothetical protein LOC75216	
scl47211.15_135-S	Rad21	NM_009009	RAD21 homolog	NM_009009
scl000738.1_3831-S	Nfat5	NM_133957	nuclear factor of activated T-cells 5 isoform b	NM_018823
scl00140917.1_97-S	Dclre1b	NM_133865	DNA cross-link repair 1B, PSO2 homolog isoform b	NM_0010253
scl38761.18.1_172-S	Ggt1	NM_008116	gamma-glutamyltransferase 1	NM_008116
scl33425.15_229-S	D230025D16Rik	NM_145604	lin-10	NM_145604
scl0020842.1_84-S	Stag1	NM_009282	stromal antigen 1	NM_009282
scl33321.20.266_29-S	Cog4	NM_133973	component of oligomeric golgi complex 4	NM_133973
scl0064164.1_275-S	Ifrg15	NM_022329	interferon alpha responsive	NM_022329
scl00231380.2_31-S	5730469D23Rik	NM_172712	hypothetical protein LOC231380	NM_172712
scl39235.7_89-S	Arhgdia	NM_133796	Rho GDP dissociation inhibitor (GDI) alpha	NM_133796
scl52780.10_0-S	Slc3a2	NM_008577	solute carrier family 3 (activators of dibasic and neutral amino acid t	NM_008577
scl0213498.14_294-S	Arhgef11	XM_159702	Rho guanine nucleotide exchange factor (GEF) 11	
scl0017925.1_325-S	Myo9b	NM_015742	myosin IXb	NM_015742
scl52961.19_173-S	Add3	NM_013758	adducin 3 (gamma)	NM_013758
scl075541.1_190-S	1700019G17Rik	NM_029331	hypothetical protein LOC75541	NM_029331
scl40379.30_407-S	Ranbp17	NM_023146	RAN binding protein 17	NM_023146
scl49778.7_54-S	Sh3gl1	NM_013664	SH3-domain GRB2-like 1	NM_013664
scl16633.10.1_180-S	Pecr	NM_023523	peroxisomal trans 2-enoyl CoA reductase	NM_023523
scl015574.1_13-S	Hus1	NM_008316	Hus1 homolog	NM_008316

scl00207213.2_277-S	Tdpoz1	NM_148949	TD and POZ domain containing 1	NM_148949
scl0057247.1_172-S	Zfp276	NM_020497	zinc finger protein 276	NM_020497
scl0074143.1_234-S	Opa1	NM_133752	optic atrophy 1 homolog	NM_133752
scl0381944.2_84-S	Dub1a	NM_201409	deubiquitinating enzyme 1a	NM_201409
scl38628.18.8_108-S	Txnrd1	NM_015762	thioredoxin reductase 1	NM_015762
scl53939.16_446-S	Abcb7	XM_356348	ATP-binding cassette, sub-family B (MDR/TAP), member 7 isoform	XM_356348
scl40252.14.1_42-S	Phf15	NM_199299	PHD finger protein 15	NM_199299
scl0020466.2_206-S	Sin3a	NM_011378	transcriptional regulator, SIN3A	NM_011378
scl30501.18.1_150-S	Tp53i5	NM_178381	Trp53 inducible protein 5	NM_178381
scl0013171.2_124-S	Dbt	NM_010022	dihydrolipoamide branched chain transacylase E2	NM_010022
scl29327.6.1_41-S		NM_010156		
scl070737.1_295-S		XM_131052		
scl00241520.2_184-S	D430039N05Rik	NM_175514	hypothetical protein LOC241520	NM_175514
scl38886.15_179-S	Cbara1	NM_144822	calcium binding atopy-related autoantigen 1	NM_144822
scl38333.7.174_81-S	Cdk4	NM_009870	cyclin-dependent kinase 4	NM_009870
scl53368.5_483-S	5730409F24Rik	NM_181403	vacuolar protein sorting 37C	NM_181403
scl0170734.1_214-S	Zfp371	NM_133204	zinc finger protein 371	NM_133204
scl18683.9.1_140-S	Zc3hdc8	NM_020594	zinc finger CCCH type containing 8	NM_020594
scl54409.6_0-S	Tcte1l	NM_025975	t-complex-associated-testis-expressed 1-like	NM_025975
scl083921.9_127-S	Tmem2	NM_031997	transmembrane protein 2	NM_0010337
scl000404.1_21-S	Cpb2	NM_019775	carboxypeptidase B2 (plasma)	NM_019775
scl0001563.1_89-S	elf1; elf3; SPTB2	NM_175836	spectrin beta 2 isoform 2	NM_009260
scl17355.9_0-S	Glul	NM_008131	glutamine synthetase	NM_008131
scl34177.5_600-S	2310022B05Rik	NM_175149	hypothetical protein LOC69551	NM_175149
GI_6753645-S	Dlx2	NM_010054	distal-less homeobox 2	NM_010054
scl060365.6_119-S	Rbm8a	NM_025875	RNA binding motif protein 8a	NM_025875
scl00242585.2_286-S	Slc35d1	NM_177732	solute carrier family 35 (UDP-glucuronic acid/UDP-N-acetylgalactos	NM_177732
scl46893.13.1_7-S	AV014541; 6330	NM_178869	tubulin tyrosine ligase-like 1	NM_178869
scl33129.9_499-S	Suv420h2	NM_146177	suppressor of variegation 4-20 homolog 2	NM_146177
scl44239.8_168-S	Zfp307	NM_023685	zinc finger protein 306	NM_023685
scl0016195.2_10-S	Il6st	NM_010560	interleukin 6 signal transducer	NM_010560
scl54440.26.1_53-S	Hdac6	NM_010413	histone deacetylase 6	NM_010413
scl36601.18.1_52-S		XM_356175		
scl31063.18_269-S	2810439K08Rik	NM_028343	hypothetical protein LOC72759	NM_028343
scl40689.20_176-S	Sep-09	NM_017380	septin 9	NM_017380
scl0002862.1_78-S	Dnaic1	NM_175138	dynein, axonemal, intermediate chain 1	NM_175138
scl00329002.1_170-S		NM_177832		
scl39311.17_24-S	Srp68	NM_146032	signal recognition particle 68	NM_146032
scl068634.3_3-S	1110025I09Rik	NM_026795	BBP-like protein 2 isoform 1	NM_026795
scl32084.18.1_39-S	Slc5a11	NM_146198	sodium/glucose cotransporter KST1	NM_146198
scl072124.8_39-S	Seh1l	NM_028112	sec13-like protein	NM_028112
scl0003760.1_927-S	Elmo1	NM_198093	engulfment and cell motility 1 isoform 1	NM_080288
scl072026.10_10-S	Trmt1	NM_028063	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase 1	NM_028063
scl0001321.1_142-S	Gps2	XM_126221	G protein pathway suppressor 2	
scl00218214.2_131-S	Aof1	NM_172262	amine oxidase, flavin containing 1	NM_172262
scl24426.8_231-S	Wdr40a	NM_026893	WD repeat domain 40A	NM_026893
scl41405.25_9-S	Gas7	NM_008088	growth arrest specific 7	NM_008088
scl0240892.1_60-S	C130085G02Rik	XM_136364	dual specificity phosphatase 27 (putative)	
scl018519.18_103-S	Pcaf	NM_020005	p300/CBP-associated factor	NM_020005
scl0021969.2_46-S	Top1	NM_009408	topoisomerase (DNA) I	NM_009408
scl0102866.1_231-S	Pls3	NM_145629	plastin 3 precursor	NM_145629
scl41791.4.1_29-S	Tmem17	NM_153596	transmembrane protein 17	NM_153596
scl054201.1_27-S	Zfp316	NM_017467	zinc finger protein 316	NM_017467
scl022631.1_10-S	Ywhaz	NM_011740	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activatior	NM_011740
scl000974.1_4-S	1200016D23Rik	NM_028776	ezrin-binding partner PACE-1	NM_028776
scl0223649.1_255-S	BC011468	NM_144847	nuclear receptor binding protein 2	NM_144847
scl0068796.2_313-S	1110039B18Rik	NM_144525	hypothetical protein LOC68796	NM_144525
scl33661.14.3_4-S	Tom1	NM_011622	target of myb1 homolog	NM_011622
scl41427.25.1_159-S	Elac2	NM_023479	elaC homolog 2	NM_023479
scl39660.7_333-S	Pnpo	NM_134021	pyridoxine 5'-phosphate oxidase	NM_134021
scl0075292.1_77-S	Prkcn	NM_029239	protein kinase D3	NM_029239
scl42676.6_93-S	Rab10	NM_016676	RAB10, member RAS oncogene family	NM_016676
scl34392.3.1_200-S	E130303B06Rik	NM_198299	hypothetical protein LOC102124	NM_198299
scl0225898.22_280-S	BC022146	NM_144872	echinoderm microtubule associated protein like 3	NM_144872
scl0217779.3_307-S	2610022K04Rik	NM_153121	LysM, putative peptidoglycan-binding, domain containing 1	NM_028134
scl00231290.1_128-S	E130304D01	NM_173403	solute carrier family 10 (sodium/bile acid cotransporter family), men	NM_173403
scl0319757.3_203-S	Smo	NM_176996	smoothened homolog	NM_176996
scl066797.3_51-S	Cntnap2	XM_358363	contactin associated protein-like 2 isoform b	
scl050917.1_181-S	Galns	NM_016722	galactosamine (N-acetyl)-6-sulfate sulfatase	NM_016722
scl21098.22.1_27-S	Pkn3	NM_153805	protein kinase N3	NM_153805
scl50176.16.1_35-S	Msln	NM_018857	mesothelin	NM_018857
scl51365.21_674-S	A630042L21Rik	NM_134134	hypothetical protein LOC106894	NM_134134
scl35703.13_436-S	Map2k1	NM_008927	mitogen activated protein kinase kinase 1	NM_008927
scl0003023.1_84-S	Fnbp4	NM_018828	formin binding protein 4	NM_018828
scl32646.8.1_89-S	Ldh3	NM_013580	lactate dehydrogenase 3, C chain, sperm specific	NM_013580
scl014897.1_29-S	Trip12	NM_133975	thyroid hormone receptor interactor 12	NM_133975
scl0022145.1_239-S	Tuba4	NM_009447	tubulin, alpha 4	NM_009447
scl00230257.2_191-S	Rod1	NM_144904	ROD1 regulator of differentiation 1 isoform 1	NM_144904
scl47614.23.1_38-S	Shank3	NM_021423	SH3/ankyrin domain gene 3	NM_021423
scl0056347.2_235-S	Eif3s8	NM_019646	eukaryotic translation initiation factor 3, subunit 8	
scl0381644.24_249-S	BC062951	NM_199032	centrosomal protein 4	NM_199032
scl054382.1_21-S	Tcstv1	NM_018756	2-cell-stage, variable group, member 1	NM_018756
scl47326.13.2_27-S	Cct5	NM_007637	chaperonin subunit 5 (epsilon)	NM_007637
scl066624.1_138-S	5730406I15Rik	NM_025668	signal peptidase complex subunit 2 homolog	NM_025668
scl25133.32.1_60-S	Nrd1	NM_146150	nardilysin, N-arginine dibasic convertase, NRD convertase 1	NM_146150
scl41115.27.1_34-S	Myohd1	NM_025414	myosin head domain containing 1	NM_025414
scl0002461.1_0-S		NM_178718		
scl53187.33.1_13-S	Mphosph1	XM_193936	M-phase phosphoprotein 1 isoform 1	XM_193936
scl054194.1_276-S	Akap8l	NM_017476	A kinase (PRKA) anchor protein 8-like	NM_017476
scl43611.1.1_320-S		NM_026606		
scl26132.13_30-S	B830028P19Rik	NM_172998	hypothetical protein LOC269695	NM_172998
scl30468.3.45_2-S	Ins2	NM_008387	insulin II	NM_008387
scl023948.13_25-S	Mmp17	NM_011846	matrix metalloproteinase 17	NM_011846
scl0054635.2_40-S	Pdgfc	NM_019971	platelet-derived growth factor, C polypeptide	NM_019971
scl020430.29_22-S	Cyfip1	NM_011370	cytoplasmic FMR1 interacting protein 1	NM_011370
scl16846.27.1_59-S	Eif5b	NM_019570	REV1-like	NM_019570
scl25019.5.1_161-S	Edn2	NM_007902	endothelin 2	NM_007902
scl000549.1_38-S	Cbfb	NM_022309	core binding factor beta	NM_022309
scl45910.20.1_29-S	4930452B06Rik	NM_028934	hypothetical protein LOC74430	NM_028934
scl41013.18_227-S	5730593F17Rik	NM_172543	RIKEN cDNA 5730593F17	NM_172543

scl053610.12_290-S	Nono	NM_023144	non-POU-domain-containing, octamer binding protein	NM_023144
scl28573.11_3-S	Crbn	NM_175357	cereblon isoform 1	NM_021449
scl0017257.1_80-S	Mecp2	NM_010788	methyl CpG binding protein 2	NM_010788
scl45197.6_141-S	Fbxl3	NM_015822	F-box and leucine-rich repeat protein 3	NM_015822
scl42484.15.73_1-S	Strn3	NM_052973	striatin, calmodulin binding protein 3	NM_052973
scl0004144.1_174-S	Zfp644	XM_358356	zinc finger motif enhancer binding protein 2	
scl33938.5.1_209-S	2310043K02Rik	NM_025869	dual specificity phosphatase 26	NM_025869
scl53577.7_510-S	Prps2	NM_026662	phosphoribosyl pyrophosphate synthetase 2	NM_026662
scl23188.11.1_55-S	4933406N12Rik	NM_028937	hypothetical protein LOC74434	NM_028937
scl48790.9.1_54-S	Nagpa	NM_013796	N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminid	NM_013796
scl0003444.1_11-S	Mre11a	NM_018736	meiotic recombination 11 homolog A	NM_018736
scl0077371.2_183-S	Sec24a	NM_175255	SEC24 related gene family, member A	NM_175255
scl53350.20_138-S	AV312086	NM_172635	hypothetical protein LOC225929	NM_172635
scl000768.1_18-S	5230400G24Rik	NM_029409	hypothetical protein LOC75734	NM_029409
scl37239.6.39_1-S	A230050P20Rik	NM_175687	hypothetical protein LOC319278	NM_175687
scl066680.1_207-S	3230401D17Rik	NM_025699	oxidative stress responsive 1	NM_025699
scl00110052.1_94-S	Dek	NM_025900	DEK oncogene (DNA binding)	NM_025900
scl31237.29_375-S	Tjp1	NM_009386	tight junction protein 1	NM_009386
scl40586.13.1_225-S	Hormad2	XM_126016	HORMA domain containing 2 isoform 1	XM_126016
scl0054632.2_248-S	Ftsj1	NM_133991	Ftsj homolog	NM_133991
scl0056491.1_59-S	Vapb	NM_019806	vesicle-associated membrane protein, associated protein B and C	NM_019806
scl0105835.22_280-S	Rutbc3	NM_134091	RUN and TBC1 domain containing 3	NM_134091
scl017470.1_35-S	Cd200	NM_010818	Cd200 antigen	NM_010818
scl54745.40.1_21-S	Tnrc11	NM_021521	mediator of RNA polymerase II transcription, subunit 12 homolog	NM_021521
scl24999.4_423-S	Lmyc1	NM_008506	lung carcinoma myc related oncogene 1	NM_008506
scl0001589.1_43-S	Pnpo	NM_134021	pyridoxine 5'-phosphate oxidase	NM_134021
scl076044.3_63-S	5830426I05Rik	NM_133762	more than blood	NM_133762
scl32836.5_532-S	Zfp30	NM_013705	zinc finger protein 30	NM_013705
scl020778.1_265-S	Scarb1	NM_016741	scavenger receptor class B, member 1	NM_016741
scl16333.16_85-S	Mcm6	NM_008567	minichromosome maintenance protein 6	NM_008567
scl0014390.1_51-S	Gabpa	NM_008065	GA repeat binding protein, alpha	NM_008065
scl0067072.1_237-S	Cdc37l1	NM_025950	cell division cycle 37 homolog (S. cerevisiae)-like 1	NM_025950
scl067455.1_261-S	Klhl13	NM_026167	kelch-like 13	NM_026167
scl00234736.1_63-S	Rfwd3	NM_146218	ring finger and WD repeat domain 3	NM_146218
scl49765.36_30-S	Ptprs	NM_011218	protein tyrosine phosphatase, receptor type, S	NM_011218
scl18731.9_1-S	AI851464	XM_130551	hypothetical protein LOC99100 isoform 1	XM_130551
scl25228.16_571-S	A430091O22Rik	NM_183024	raver2	NM_183024
scl17290.14_518-S	1200016D23Rik	NM_028776	ezrin-binding partner PACE-1	NM_028776
scl0099889.1_5-S	Arfip1	XM_130985	ADP-ribosylation factor interacting protein 1 isoform 1	XM_130985
scl0059069.1_170-S	Tpm3	NM_022314	tropomyosin 3, gamma	NM_022314
scl0003069.1_215-S	Rbm12	NM_029397	RNA binding motif protein 12	NM_029397
scl37323.16_486-S	Cwf19l2	NM_027545	CWF19-like 2, cell cycle control	NM_027545
scl28393.5.1_16-S	Kcna6	NM_013568	potassium voltage-gated channel, shaker-related, subfamily, memb	NM_013568
scl026428.1_91-S	Orc4l	NM_011958	origin recognition complex subunit 4	NM_011958
scl0028185.2_148-S	Tomm70a	NM_138599	translocase of outer mitochondrial membrane 70 homolog A	NM_138599
scl18001.16.1_91-S	Ercc5	NM_011729	excision repair cross-complementing rodent repair deficiency, comp	NM_011729
scl27787.7_486-S	0610040J01Rik	NM_029554	hypothetical protein LOC76261	NM_029554
scl084652.1_47-S	Drctnnb1a	NM_053090	down-regulated by Ctnnb1, a	NM_053090
scl000011.1_73-S	Strn3	NM_052973	striatin, calmodulin binding protein 3	NM_052973
scl021667.1_237-S	Tdgf1	NM_011562	teratocarcinoma-derived growth factor	NM_011562
scl00226432.1_235-S	Ipo9	XM_129442	importin 9 isoform 1	XM_129442
scl52796.6.1_4-S	Acy3	NM_027857	aspartoacylase-3	NM_027857
scl52406.11.1_60-S	Actr1a	NM_016860	ARP1 actin-related protein 1 homolog A	NM_016860
scl20431.14.1_58-S		XM_130428		
scl0067419.2_138-S	3632451O06Rik	NM_026142	hypothetical protein LOC67419	NM_026142
scl18631.16_116-S	Slc23a2	NM_018824	solute carrier family 23 (nucleobase transporters), member 2	NM_018824
scl0014211.2_328-S	Smc2l1	NM_008017	structural maintenance of chromosomes 2-like 1	NM_008017
scl43536.32_392-S	Ipo11	NM_029665	importin 11	NM_029665
scl42244.15_187-S	Aldh6a1	NM_134042	aldehyde dehydrogenase family 6, subfamily A1	NM_134042
scl0235072.13_139-S	Sep-07	NM_009859	cell division cycle 10 homolog	NM_009859
scl058810.1_30-S	Akr1a4	NM_021473	aldo-keto reductase family 1, member A4 (aldehyde reductase)	NM_021473
scl067345.15_210-S	Herc4	NM_026101	hect domain and RLD 4	NM_026101
scl48625.8_280-S	Bcl6	NM_009744	B-cell leukemia/lymphoma 6	NM_009744
scl36411.19_355-S	Lrrfip2	XM_284541	leucine rich repeat (in FLII) interacting protein 2 isoform 1	XM_284541
scl00193670.1_279-S	1700022N24Rik	NM_145355	ring finger protein 185	NM_145355
scl29818.18.1_0-S	Alms1	XM_355793	Alstrom syndrome 1	
scl0074053.2_257-S	Grip1	NM_130891	glutamate receptor interacting protein 1 isoform 1	NM_028736
scl52903.34_0-S	Plcb3	NM_008874	phospholipase C, beta 3	NM_008874
scl0002017.1_169-S	Col25a1	NM_198711	collagen, type XXV, alpha 1 isoform 1	NM_029838
scl15870.13_49-S	Akt3	NM_011785	thymoma viral proto-oncogene 3	NM_011785
scl27718.20_150-S	AI836376	NM_178896	DCN1, defective in cullin neddylation 1, domain containing 4	NM_178896
scl31391.7.1_2-S	Rcn3	NM_026555	reticulocalbin 3	NM_026555
scl0109135.16_234-S	Plekha5	NM_144920	phosphoinositol 3-phosphate-binding protein-2	NM_144920
scl0072278.1_319-S	Ccpg1	NM_028181	cell cycle progression 8 protein	NM_028181
scl026377.1_5-S	Dapp1	NM_011932	dual adaptor for phosphotyrosine and 3-phosphoinositides 1	NM_011932
scl47390.18.104_38-S	Tars	NM_033074	threonyl-tRNA synthetase	NM_033074
scl0066645.2_243-S	Pspc1	NM_025682	paraspeckle protein 1	NM_025682
scl41729.5.1_4-S	3300001G02Rik	NM_030093	RIKEN cDNA 3300001G02	NM_030093
scl0019205.2_0-S	Ptbp1	NM_008956	polypyrimidine tract binding protein 1	NM_008956
scl12614.1.1_13-S	Ube2q	NM_027315	ubiquitin-conjugating enzyme E2Q	NM_027315
scl21344.10_377-S	9630050M13Rik	XM_194000	hypothetical protein LOC269233	XM_194000
scl42037.39_16-S	Cdc42bpb	NM_183016	Cdc42 binding protein kinase beta	NM_183016
scl0003164.1_24-S	Rbm18	NM_026434	RNA binding motif protein 18	NM_026434
scl0378466.1_307-S		NM_194351		
scl0069354.1_68-S	Slc38a4	NM_027052	solute carrier family 38, member 4	NM_027052
scl0015387.1_0-S	Hnrpk	NM_025279	heterogeneous nuclear ribonucleoprotein K	NM_025279
scl24617.16.29_1-S	Cdc2l1	NM_007661	cell division cycle 2-like 1	NM_007661
scl0213473.2_27-S		XM_135033		
scl41416.15.1_75-S	Myh3	XM_354614	myosin, heavy polypeptide 3, skeletal muscle, embryonic	XM_354614
scl0067459.2_215-S	Nvl	NM_026171	nuclear VCP-like	NM_026171
scl000094.1_33-S	Ppp2r5d	NM_009358	delta isoform of regulatory subunit B56, protein phosphatase 2A	NM_009358
scl46280.17_461-S	Wdr23	NM_133734	WD repeat domain 23	NM_133734
scl0003602.1_637-S	Gnb5	NM_138719	guanine nucleotide-binding protein, beta-5 subunit isoform 1	NM_010313
scl0002119.1_21-S	Prcc	NM_033573	papillary renal cell carcinoma translocation-associated gene produc	NM_033573
scl0003213.1_223-S	Trib3	NM_144554	tribbles homolog 3 isoform 1	NM_144554
scl24729.4_232-S	4732496O08Rik	NM_172877	hypothetical protein LOC242736	NM_172877
scl26863.16_246-S	Wdr1	NM_011715	WD repeat domain 1	NM_011715
scl0245688.12_1-S	Rbbp7	NM_009031	retinoblastoma binding protein 7	NM_009031
scl25451.20.1_292-S	Invs	NM_010569	inversin	NM_010569
scl42934.12.1_88-S	Adck1	NM_028105	aarF domain containing kinase 1	NM_028105

scl50605.12.1_93-S	Fsd1	NM_183178	fibronectin type 3 and SPRY domain-containing protein	NM_183178
scl013807.1_30-S	Eno2	NM_013509	enolase 2, gamma neuronal	NM_013509
scl0071782.1_119-S	D5Ertd585e	NM_027922	hypothetical protein LOC71782	NM_027922
scl012366.11_305-S	Casp2	NM_007610	caspase 2	NM_007610
scl21036.31.1_8-S	Mapkap1	NM_177345	mitogen-activated protein kinase associated protein 1	NM_177345
scl34515.8.1_201-S		NM_177902		
scl0065970.2_260-S	D15Ertd366e	NM_023063	epithelial protein lost in neoplasm	NM_023063
scl0066691.1_167-S	4432404J10Rik	NM_025709	GTPase activating protein and VPS9 domains 1	NM_025709
scl36347.8_506-S	Endogl1	NM_172456	endonuclease G-like 1	NM_172456
scl41889.9.1_252-S	2410008K03Rik	XM_125970	hypothetical protein LOC71962	
scl00230249.2_159-S	AI314180	NM_172381	expressed sequence AI314180	NM_172381
scl000354.1_6-S	Slc22a17	NM_021551	solute carrier family 22 (organic cation transporter), member 17	NM_021551
scl0019684.1_297-S	Rdx	NM_009041	radixin	NM_009041
gi_30794511_ref_NM_	Hmbs	NM_013551	hydroxymethylbilane synthase	NM_013551
scl37245.7_152-S	2210010B09Rik	NM_172919	hypothetical protein LOC244721	NM_172919
scl54226.16.1_58-S	4933424A10Rik	NM_177293	hypothetical protein LOC320923	NM_177293
scl41716.5.1_57-S	9630054F20Rik	NM_173784	dendritic cell-derived ubiquitin-like protein	NM_173784
scl47892.24_470-S	D15Ertd621e	NM_145959	hypothetical protein LOC210998	NM_145959
scl056275.3_18-S	Rbm14	NM_019869	RNA binding motif protein 14	NM_019869
scl3664.1.1_276-S	Dio3	NM_172119	deiodinase, iodothyronine type III	NM_172119
scl46200.4_648-S	A030013D21	NM_177628	hypothetical protein LOC219148	NM_177628
scl072020.1_37-S	1600021C16Rik	NM_028059	zinc finger protein 654	NM_028059
scl35387.5.1_9-S	Tex264	NM_011573	testis expressed gene 264	NM_011573
scl059033.25_59-S	Slc4a8	NM_021530	solute carrier family 4 (anion exchanger), member 8	NM_021530
scl0114565.1_16-S	Zfp295	NM_175428	zinc finger protein 295	NM_175428
scl24866.28.1_30-S	Map3k6	NM_016693	mitogen-activated protein kinase kinase kinase 6	NM_016693
scl001000710.2_193-S	Aprin	NM_175310	androgen-induced prostate proliferative shutoff associated protein A	NM_175310
scl53815.3_1-S	Rex3	NM_009052	brain expressed, X-linked 1	NM_009052
scl18423.21_359-S	Rbl1	NM_011249	retinoblastoma-like protein 1	NM_011249
scl38642.20.1_2-S	Ankrd24	NM_027480	ankyrin repeat domain 24	NM_027480
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl00217995.1_300-S	B130016L12Rik	NM_144835	BAP28 protein	NM_144835
scl0381724.1_1-S	BC061212	NM_198667	hypothetical protein LOC381724	NM_198667
scl0272027.1_270-S	BC057893	NM_173033	hypothetical protein LOC272027	NM_173033
scl50999.7.56_77-S	Tce1	NM_027141	splA/ryanodine receptor domain and SOCS box containing 3	NM_027141
scl0014701.2_54-S	Gng12	NM_025278	guanine nucleotide binding protein (G protein), gamma 12	NM_025278
scl38024.23.1_253-S	A530089I17Rik	NM_133999	Sac domain-containing inositol phosphatase 3	NM_133999
scl43030.13_217-S	4930539P14Rik	NM_133798	hypothetical protein LOC97827	NM_133798
scl078672.2_29-S		XM_196563		
scl37833.23.1_1-S	Bicc1	NM_031397	bicaudal C homolog 1	NM_031397
scl0110651.20_30-S	Rps6ka3	NM_148945	ribosomal protein S6 kinase polypeptide 3	NM_148945
scl34186.14_327-S	Abcb10	NM_019552	ATP-binding cassette, sub-family B, member 10	NM_019552
scl0233824.2_3-S	Cog7	XM_133861	component of oligomeric golgi complex 7	
scl0012808.2_292-S	Cobl	NM_172496	cordon-bleu protein	NM_172496
scl30660.11.1_1-S	Kif22	NM_145588	kinesin family member 22	NM_145588
scl0066404.2_152-S	2410001C21Rik	NM_025542	hypothetical protein LOC66404	NM_025542
scl31496.6.1_91-S	Scn1b	NM_011322	sodium channel, voltage-gated, type I, beta	NM_011322
scl25953.10_428-S	Wbscr16	NM_033572	Williams-Beuren syndrome chromosome region 16 homolog	NM_033572
scl35262.19_0-S	1300006C19Rik	XM_358385	source of immunodominant MHC-associated peptides	
scl19346.22_0-S	Golga1	NM_029793	golgi autoantigen, golgin subfamily a, 1	NM_029793
scl0319710.13_85-S	4930488L10Rik	NM_028127	FERM domain containing 6	NM_028127
scl23536.7_579-S	2610305D13Rik	NM_145078	hypothetical protein LOC112422	NM_145078
scl00102607.2_257-S	Snx19	NM_028874	sorting nexin 19	NM_028874
scl41758.26.1_24-S	Pnpt1	NM_027869	polyribonucleotide nucleotidyltransferase 1	NM_027869
scl36067.18_4-S	Aplp2	NM_009691	amyloid beta (A4) precursor-like protein 2	NM_009691
scl000342.1_1-S	1700009P03Rik	NM_134077	cutaneous T-cell lymphoma tumor antigen se70-2	NM_134077
scl36817.19_229-S	Dpp8	NM_028906	dipeptidylpeptidase 8	NM_028906
scl48818.18.1_43-S	Trap1	NM_026508	TNF receptor-associated protein 1	NM_026508
scl00216551.2_0-S	1110067D22Rik	NM_173752	hypothetical protein LOC216551	NM_173752
scl32350.23.1_118-S	2610034N24Rik	NM_027256	hypothetical protein LOC101861	NM_027256
scl31585.9.1_1-S	Dll3	NM_007866	delta-like 3	NM_007866
scl51248.6.1_66-S	8030462N17Rik	NM_178670	hypothetical protein LOC212163	NM_178670
scl0002269.1_18-S	Matr3	NM_010771	matrin 3	NM_010771
scl0070396.1_30-S	2210409M21Rik	NM_133728	hypothetical protein LOC70396	NM_133728
scl0003648.1_67-S	Zfp346	NM_012017	zinc finger protein 346	NM_012017
scl32765.8.1_43-S	2900093B09Rik	NM_021387	hypothetical protein LOC58188	NM_021387
scl32180.18_394-S	Parva	NM_020606	parvin, alpha	NM_020606
scl0022589.1_142-S	Atrx	NM_009530	alpha thalassemia/mental retardation syndrome X-linked homolog	NM_009530
scl056398.7_324-S	1500003O03Rik	NM_019769	calcium binding protein P22	NM_019769
scl00013.1_82-S	Arfip2	NM_029802	ADP-ribosylation factor interacting protein 2	NM_029802
scl25160.11.1_73-S	C030002N13Rik	NM_145550	Yip1 domain family, member 1	NM_145550
scl069757.1_0-S	Leng1	NM_027203	leukocyte receptor cluster (LRC) member 1	NM_027203
scl0093699.1_40-S	Pcdhga12	NM_033574	protocadherin gamma subfamily B, 1	NM_033574
scl30988.16_625-S	2700017M01Rik	NM_028292	protein phosphatase methylesterase 1	NM_028292
scl074522.7_319-S	Zcwcc1	NM_198162	microrchidia 2A	NM_198162
scl067291.6_198-S	3110023B02Rik	NM_152807	hypothetical protein LOC67291	NM_152807
scl28889.10.1_70-S	MGC59076	NM_178413	hypothetical protein LOC232078	NM_0010339
scl0055946.2_3-S	Ap3m1	NM_018829	adaptor-related protein complex 3, mu 1 subunit	NM_018829
scl44563.12_95-S	2810446P07Rik	NM_175187	hypothetical protein LOC72745	NM_175187
scl00219189.1_160-S		XM_127785		
scl0002669.1_204-S	Epb4.1l4b	NM_019427	erythrocyte protein band 4.1-like 4b	NM_019427
scl35747.13.1_171-S	B230114P05Rik	NM_172444	hypothetical protein LOC207596	NM_172444
scl50396.10.5_23-S	Msh6	NM_010830	mutS homolog 6	NM_010830
scl36550.12_47-S	Amotl2	NM_019764	angiomotin like 2	NM_019764
scl00218294.2_195-S	Cdc14b	NM_172587	CDC14 cell division cycle 14 homolog B	NM_172587
scl45408.20.1_29-S	Ephx2	NM_007940	epoxide hydrolase 2, cytoplasmic	NM_007940
scl16146.28_80-S	Lamc1	NM_010683	laminin, gamma 1	NM_010683
scl37898.15_404-S	Ddx21	NM_019553	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	NM_019553
scl26570.35_194-S	Centd1	XM_132099	centaurin, delta 1 isoform 1	XM_132099
scl0002471.1_21-S	Fbln1	NM_010180	fibulin 1	NM_010180
scl015032.1_236-S		NM_010396		
scl00229782.2_107-S	Slc35a3	NM_144902	solute carrier family 35 (UDP-N-acetylglucosamine (UDP-GlcNAc) t	NM_144902
scl0170719.14_114-S	Oxr1	NM_130885	oxidation resistance 1	NM_130885
scl19476.9.1_2-S	D2Wsu81e	NM_172660	hypothetical protein LOC227695	NM_172660
scl0066105.2_68-S	Ube2d3	NM_025356	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	NM_025356
scl35223.18_513-S	Oxsr1	XM_135264	oxidative-stress responsive 1 isoform 1	XM_135264
scl33269.35.1_230-S	Plcg2	NM_172285	phospholipase C, gamma 2	NM_172285
scl27730.23_48-S	Atp10d	NM_153389	ATPase, Class V, type 10D	NM_153389
scl19495.14.1_36-S	Gtf3c5	NM_148928	general transcription factor III polypeptide 5	NM_148928
scl0003636.1_5-S	Tgfbi	NM_009369	transforming growth factor, beta induced	NM_009369

scl35367.18_438-S	Sema3f	NM_011349	sema domain, immunoglobulin domain (Ig), short basic domain, se	NM_011349
scl066989.6_1-S	AI451943; AW54	NM_025888	hypothetical protein LOC66989	NM_025888
scl00214137.2_222-S	B130017I01Rik	NM_172525	PTPL1-associated RhoGAP 1	NM_172525
scl00319263.1_175-S	A030012M09Rik	NM_183028	hypothetical protein LOC319263	NM_183028
scl011429.19_155-S	Aco2	NM_080633	aconitase 2, mitochondrial	NM_080633
scl17563.43_11-S				
scl22078.18_184-S	Kpna4	NM_008467	karyopherin alpha 4	NM_008467
scl00319876.2_224-S	Cobll1	NM_177025	Cobl-like 1	NM_177025
scl53163.10.1_0-S	1200008O12Rik	NM_028760	hypothetical protein LOC74107	NM_028760
scl00233056.2_80-S	6330581L23Rik	NM_146185	hypothetical protein LOC233056	NM_146185
scl071919.4_35-S	D15Ertd682e	NM_028003	hypothetical protein LOC71919	NM_028003
scl0218333.1_228-S	BC018507	XM_358313	hypothetical protein LOC218333 isoform 1	XM_358313
scl38323.13.1_7-S	Arhgap9	NM_146011	Rho GTPase activating protein 9	NM_146011
scl46272.20.1_37-S	Rec8L1	NM_020002	REC8-like 1	NM_020002
scl18732.66_476-S	Fbn1	NM_007993	fibrillin 1	NM_007993
scl00214498.1_240-S	Hrpt2	NM_145991	hyperparathyroidism 2 homolog	NM_145991
scl18270.11_281-S	Aurka	NM_011497	serine/threonine protein kinase 6	NM_011497
scl21209.20_282-S	Yme1l1	NM_013771	YME1-like 1	NM_013771
scl00224640.2_275-S	Lemd2	NM_146075	LEM domain containing 2	NM_146075
scl46872.13_7-S	Cerk	NM_145475	ceramide kinase	NM_145475
scl00213541.2_318-S	Ythdf2	NM_145393	high glucose-regulated protein 8	NM_145393
scl0001631.1_196-S	Rpo1-1	NM_009085	RNA polymerase 1-1	NM_009085
scl066840.1_106-S	Wdr45l	NM_025793	Wdr45 like	NM_025793
scl078825.5_201-S	5830417C01Rik	NM_024282	hypothetical protein LOC78825	NM_024282
scl29109.11_491-S	Mkrn1	NM_018810	makorin, ring finger protein, 1	NM_018810
scl33381.16_572-S	Cdh1	NM_009864	cadherin 1	NM_009864
scl014745.1_6-S	Edg2	NM_010336	endothelial differentiation, lysophosphatidic acid G-protein-coupled r	NM_010336
scl53938.12_633-S	Zdhhc15	NM_175358	zinc finger, DHHC domain containing 15	NM_175358
scl26258.25_611-S	Evi5	NM_007964	ecotropic viral integration site 5	NM_007964
scl068743.2_26-S	Anln	NM_028390	anillin	NM_028390
scl51494.33_412-S	E030006K04Rik	NM_139206	ARF-GAP, RHO-GAP, ankyrin repeat and pleckstrin homology dor	NM_139206
scl00227094.2_263-S	5330401P04Rik	NM_172654	hypothetical protein LOC227094	NM_172654
scl0027364.1_305-S	Srr	NM_013761	serine racemase	NM_013761
scl000086.1_135-S	AA959742	NM_133807	hypothetical protein LOC98238	NM_133807
scl0068198.2_271-S	Ndufb2	NM_026612	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2	NM_026612
scl19759.21.1_156-S	2610031L17Rik	NM_133701	U5 snRNP-associated 102 kDa protein	NM_133701
scl0014696.1_196-S	Gnb4	NM_013531	guanine nucleotide-binding protein, beta-4 subunit	NM_013531
scl0083921.1_123-S	Tmem2	NM_031997	transmembrane protein 2	NM_0010337
scl00227399.2_17-S	AW555814	NM_173760	hypothetical protein LOC227399	NM_173760
scl33851.12.1_1-S	1810047C23Rik	NM_138668	hypothetical protein LOC192169	NM_138668
scl16258.13.1_64-S	Rnpep	NM_145417	arginyl aminopeptidase (aminopeptidase B)	NM_145417
scl0068916.1_226-S	Cdkal1	NM_144536	CDK5 regulatory subunit associated protein 1-like 1	NM_144536
scl53173.13_456-S	E430027O22Rik	XM_129248	BTAF1 RNA polymerase II, B-TFIID transcription factor-associated,	XM_129248
scl021341.1_249-S	Taf1c	NM_021441	TATA box binding protein (Tbp)-associated factor, RNA polymerase	NM_021441
scl00217449.2_221-S	Ttc15	NM_178811	tetratricopeptide repeat domain 15	NM_178811
scl43172.15_129-S	Prpf39	NM_177806	PRP39 pre-mRNA processing factor 39 homolog	NM_177806
scl52369.13.4_3-S	Xpnpep1	NM_133216	X-prolyl aminopeptidase (aminopeptidase P) 1, soluble	NM_133216
scl024135.1_208-S	Zfp68	NM_013844	Zinc finger protein 68	NM_013844
scl0068263.1_199-S	Pdhb	NM_024221	pyruvate dehydrogenase (lipoamide) beta	NM_024221
scl070356.1_235-S	St13	NM_133726	suppression of tumorigenicity 13	NM_133726
scl52956.24.4_13-S	Nfkb2	NM_019408	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2	NM_019408
scl26009.10.1_18-S	Slc15a4	NM_133895	solute carrier family 15, member 4	NM_133895
scl056392.8_7-S	Shoc2	NM_019658	soc-2 (suppressor of clear) homolog	NM_019658
scl47033.26.1_53-S	2810439M11Rik	NM_183091	I-kappa-B-related protein	NM_183091
scl0001452.1_0-S	Flcn	NM_146018	folliculin	NM_146018
scl070428.20_289-S	Polr3b	NM_027423	RNA polymerase III subunit RPC2	NM_027423
scl0067456.1_141-S	1200009B18Rik	NM_026168	PTX1 protein isoform 1	NM_026168
scl21103.16.1_274-S	Gle1l	XM_130106	GLE1 RNA export mediator-like (yeast	
scl011764.20_169-S	Ap1b1	NM_007454	adaptor protein complex AP-1, beta 1 subunit	NM_007454
scl0065973.1_101-S	Asph	NM_023066	aspartyl beta-hydroxylase isoform 1	NM_023066
scl0072193.1_269-S	Sfrs2ip	XM_128178	splicing factor, arginine/serine-rich 2, interacting protein	
scl24976.4.1_46-S	2310005N01Rik	NM_027310	hypothetical protein LOC70088	NM_027310
scl027419.6_267-S	Naglu	NM_013792	alpha-N-acetylglucosaminidase	NM_013792
scl32944.30.1_12-S	Arhgef1	NM_008488	Rho guanine nucleotide exchange factor (GEF) 1	NM_008488
scl0053621.2_255-S	Cnot4	NM_016877	CCR4-NOT transcription complex, subunit 4	NM_016877
scl073122.1_292-S	Tgfbrap1	XM_129857	TGF beta receptor associated protein -1	
scl068564.2_264-S	1110001M19Rik	XM_110931	82-kD FMRP Interacting Protein	
scl37796.6_295-S	Col6a1	NM_009933	procollagen, type VI, alpha 1	NM_009933
scl00226182.1_36-S	Taf5	NM_177342	TAF5 RNA polymerase II, TATA box binding protein (TBP)-associat	NM_177342
scl54291.11_376-S	Zdhhc9	NM_172465	zinc finger, DHHC domain containing 9	NM_172465
scl24962.21.1_11-S	Clspn	NM_175554	claspin	NM_175554
scl0011491.1_129-S	Adam17	NM_009615	a disintegrin and metalloprotease domain 17	NM_009615
scl052857.1_311-S	D7Bwg0611e	NM_027898	hypothetical protein LOC52857	NM_027898
scl49330.22.4_270-S	Chrd	NM_009893	chordin	NM_009893
scl099712.1_51-S	Cept1	NM_133869	choline/ethanolaminephosphotransferase 1	NM_133869
scl054644.12_28-S	DXImx46e	NM_138604	hypothetical protein LOC54644	NM_138604
scl0074043.2_32-S	Pex26	NM_028730	peroxisome biogenesis factor 26	NM_028730
scl33522.11_60-S	Chd9	XM_284439	chromodomain helicase DNA binding protein 9	
scl48725.21.1_175-S	2310008H04Rik	NM_146068	hypothetical protein LOC224008	NM_146068
scl0013844.2_257-S	Ephb2	XM_204072	Eph receptor B2	
scl000087.1_18-S	Cog1	NM_013581	component of oligomeric golgi complex 1	NM_013581
scl0270163.13_79-S		NM_173018		
scl23133.14.29_13-S	Gmps	XM_130877	guanine monphosphate synthetase	
scl0078929.1_75-S	Polr3h	NM_030229	polymerase (RNA) III (DNA directed) polypeptide H	NM_030229
scl24307.19_191-S	Catnal1	NM_018761	catenin (cadherin associated protein), alpha-like 1	NM_018761
scl38900.29.1_38-S	Ranbp2	NM_011240	RAN binding protein 2	NM_011240
scl46160.15.1_21-S	Clu	NM_013492	clusterin	NM_013492
scl0224109.1_93-S	E430025L02Rik	NM_146069	leucine rich repeat containing 33	NM_146069
scl000513.1_2582-S		NM_172638		
scl27710.19_132-S	Fip1l1	NM_024183	FIP1 like 1	NM_024183
scl52545.7.1_87-S		NM_181587		
scl0067452.2_1-S	1200006O19Rik	NM_026164	intracellular membrane-associated calcium-independent phospholip	NM_026164
scl36669.38.1_11-S	Myo6	NM_008662	myosin VI	NM_008662
scl23942.14.1_32-S	Plk3	NM_013807	polo-like kinase 3	NM_013807
scl18778.5_150-S	2810002D13Rik	NM_025657	hypothetical protein LOC66606	NM_025657
scl00209737.1_267-S	Knsl7	NM_010620	kinesin-like 7	NM_010620
scl0066660.1_237-S	5730555F13Rik	NM_025690	modulator of estrogen induced transcription isoform a	NM_025690
scl0072199.2_276-S	Mms19l	NM_028152	MMS19 (MET18 S. cerevisiae)-like	NM_028152
scl0003232.1_17-S	Stam2	NM_019667	signal transducing adaptor molecule (SH3 domain and ITAM motif)	NM_019667
scl39328.7_317-S	Grb2	NM_008163	growth factor receptor bound protein 2	NM_008163

scl070396.3_74-S	2210409M21Rik	NM_133728	hypothetical protein LOC70396	NM_133728
scl33390.7_348-S	Lypla3	NM_133792	lysophospholipase 3	NM_133792
scl46398.42.1_4-S	Ktn1	NM_008477	kinectin 1	NM_008477
scl00217734.2_317-S	Pomt2	NM_153415	protein-O-mannosyltransferase 2	NM_153415
scl0229473.1_11-S	D930015E06Rik	NM_172681	RIKEN cDNA D930015E06	NM_172681
scl0022427.2_165-S	Wrn	NM_011721	Werner syndrome protein	NM_011721
scl20444.23.1_130-S	Bub1b	NM_009773	budding uninhibited by benzimidazoles 1 homolog, beta	NM_009773
scl0002780.1_2-S	Exosc10	NM_016699	exosome component 10	NM_016699
scl37772.15_266-S	Pwp2h	NM_029546	PWP2 periodic tryptophan protein homolog	NM_029546
scl18618.8_0-S	3300001M20Rik	NM_175113	hypothetical protein LOC66926	NM_175113
scl47694.7_354-S	2310042L06Rik	NM_172428	hypothetical protein LOC76457	NM_172428
scl28043.12_286-S	Klhl7	NM_026448	SBBI26 homolog	NM_026448
scl0063913.2_46-S	Niban	NM_022018	niban protein	NM_022018
scl013132.15_5-S	Dab2	NM_023118	disabled homolog 2 isoform b	NM_0010087
scl36035.14_69-S	Chek1	NM_007691	checkpoint kinase 1 homolog	NM_007691
scl54150.7.1_69-S	Renbp	NM_023132	renin binding protein	NM_023132
scl52206.9_527-S	Rnf138	NM_019706	ring finger protein 138 isoform 2	NM_019706
scl36938.4.5_104-S	Crabp1	NM_013496	cellular retinoic acid binding protein I	NM_013496
scl0027368.2_99-S	Tbl2	NM_013763	transducin (beta)-like 2	NM_013763
scl0236546.4_11-S	AF067061	NM_199060	hypothetical protein LOC236546	NM_199060
scl38536.23.1_129-S	Fgd6	NM_053072	FYVE, RhoGEF and PH domain containing 6	NM_053072
scl0093840.1_217-S	Ltap	NM_033509	loop tail associated protein	NM_033509
scl35708.10_49-S	Smad3	NM_016769	MAD homolog 3	NM_016769
scl0002117.1_17-S	BC028528	NM_153513	hypothetical protein LOC229600	NM_153513
scl54548.7.1_64-S	Hadh2	NM_016763	hydroxyacyl-Coenzyme A dehydrogenase type II	NM_016763
scl0019240.2_329-S	Tmsb10	NM_025284	thymosin, beta 10	NM_025284
scl0066354.2_231-S	Skiip	NM_025507	SKI interacting protein	NM_025507
scl36822.13_234-S		XM_134902		
scl22717.15.1_30-S	Wdr47	NM_181400	WD repeat domain 47	NM_181400
scl31838.2_538-S	Fgf15	NM_008003	fibroblast growth factor 15	NM_008003
scl0075452.2_305-S	Ascc2	NM_029291	ASC-1 complex subunit P100	NM_029291
scl19490.7_664-S	Gtf3c4	NM_172977	general transcription factor IIIC, polypeptide 4	NM_172977
scl45319.27_70-S	Rb1	NM_009029	retinoblastoma 1	NM_009029
scl0240174.1_158-S		NM_183021		
scl0068477.1_272-S	1110007A06Rik	NM_024288	hypothetical protein LOC68477	NM_024288
scl39416.15_560-S	1500002M01Rik	NM_133702	nucleolar protein 11	NM_133702
scl54036.9.4_13-S	Maged1	NM_019791	melanoma antigen family D, 1	NM_019791
scl0268933.8_8-S	Wdr24	NM_173741	WD repeat domain 24	NM_173741
scl30646.3_608-S	8030466O12Rik	NM_146203	hypothetical protein LOC233893	NM_146203
scl019656.1_169-S	Rbmxrt	NM_009033	RNA binding motif protein, X chromosome retrogene	NM_009033
scl015331.1_246-S	Hmgn2	NM_016957	high mobility group nucleosomal binding domain 2	NM_016957
scl00329877.1_15-S	1700065A05Rik	XM_283972	hypothetical protein LOC329877 isoform 1	XM_283972
scl43917.11.1_43-S	BC021381	NM_145382	hypothetical protein LOC212483	NM_145382
scl45036.29.1_11-S	Vps41	NM_172120	vacuolar protein sorting 41	NM_172120
scl34532.9_187-S	Dnaja2	NM_019794	DnaJ (Hsp40) homolog, subfamily A, member 2	NM_019794
scl0227738.1_321-S	Lrsam1	NM_199302	leucine rich repeat and sterile alpha motif containing 1	NM_199302
scl54546.18_0-S	Smc1l1	NM_019710	SMC1 structural maintenance of chromosomes 1-like 1	NM_019710
scl0054219.2_27-S	425O18-1	NM_019421	putative VLDL lipoprotein receptor precursor	NM_019421
scl0001603.1_1171-S	Arl6ip2	NM_178050	ADP-ribosylation factor-like 6 interacting protein 2 isoform 1	NM_019717
scl0001498.1_3411-S	Ankfy1	NM_009671	ankyrin repeat and FYVE domain containing 1	NM_009671
scl067732.2_80-S	4833421E05Rik	NM_026347	hypothetical protein LOC67732	NM_026347
scl00271564.2_122-S	D330038K10Rik	NM_173028	vacuolar protein sorting 13A	NM_173028
scl43387.27.1_23-S	Smc6l1	NM_025695	SMC6 protein	NM_025695

Human EnR down				
Gene Name	Common	Genbank	Product	RefSeq
scl0003527.1_242-S	Ddx6	NM_007841	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6	NM_007841
scl0100986.1_13-S	Akap9	NM_194462	A kinase (PRKA) anchor protein (yotiao) 9	NM_194462
GI_21703919-A				
GI_71725393-S				
scl0232989.3_1-S	Hnrpul1	NM_178089	EIB-55kDa associated protein 5 isoform 1	NM_144922
scl27096.8.1_1-S	0610009M14Rik	NM_023910	TSC22-related inducible leucine zipper 2	NM_023910
scl0264895.2_92-S	BC018371	NM_153807	hypothetical protein LOC264895	NM_153807
scl46501.15.1_0-S	Nek4	NM_011849	NIMA (never in mitosis gene a)-related expressed kinase 4	NM_011849
scl066993.2_1-S	Smarcd3	NM_025891	SWI/SNF related, matrix associated, actin dependent regulator of c	NM_025891
GI_58372137-S				
scl37031.1_4-S		NM_030256		
scl0021356.1_24-S	Tapbp	NM_009318	TAP binding protein isoform 1	NM_0010253
scl0233490.1_48-S	Zf	NM_145151	HCF-binding transcription factor Zhangfei	NM_145151
scl0243529.4_9-S	H1fx	NM_198622	H1 histone family, member X	NM_198622
scl0004119.1_46-S	Rbpsuh	NM_009035	recombining binding protein suppressor of hairless	NM_009035
scl012048.3_57-S	Bcl2l1	NM_009743	Bcl2-like 1	NM_009743
scl014630.7_323-S	Gclm	NM_008129	glutamate-cysteine ligase , modifier subunit	NM_008129
scl0223642.1_184-S	Zc3hdc3	NM_172121	zinc finger CCCH type containing 3	NM_172121
scl0002721.1_25-S	Ece1	NM_199307	endothelin converting enzyme 1	NM_199307
scl22634.5_179-S	Pitx2	NM_011098	paired-like homeodomain transcription factor 2	NM_011098
scl0003304.1_37-S	Xrn2	NM_011917	5'-3' exoribonuclease 2	NM_011917
scl020416.3_0-S	Shc1	NM_011368	src homology 2 domain-containing transforming protein C	NM_011368
GI_85702146-A				
GI_83776566-I				
scl29554.6.1_30-S	Usp18	NM_011909	ubiquitin specific protease 18	NM_011909
GI_85677490-S				
scl22954.14.1_9-S	Crtc2	XM_110709	transducer of regulated cAMP response element-binding protein (CREB) 2	
GI_71274129-S				
scl0016688.1_300-S	Krt2-6b	NM_010669	keratin complex 2, basic, gene 6b	NM_010669
scl0214987.1_149-S	5830457O10Rik	NM_145412	hypothetical protein LOC214987	NM_145412
scl35671.6_128-S	Lactb	NM_030717	lactamase, beta	NM_030717
scl0019294.1_121-S	Pvrl2	NM_008990	poliovirus receptor-related 2	NM_008990
scl48815.9.1_11-S	Crebbp	XM_148699	CREB binding protein	
scl0002781.1_26-S	Pafah2	NM_133880	platelet-activating factor acetylhydrolase 2	NM_133880
scl22447.5.1_50-S	Zzz3	NM_198416	zinc finger, ZZ domain containing 3	NM_198416
GI_85986646-S				
scl39032.1_64-S	Tspyl1	NM_009433	testis-specific protein, Y-encoded-like 1	NM_009433
scl000674.1_0-S	Itgb1	NM_010578	integrin beta 1 (fibronectin receptor beta)	NM_010578
scl00353282.2_264-S		NM_177386		
GI_51921350-S				
scl0002677.1_6-S	Nol6	NM_139236	nucleolar RNA-associated protein long isoform	NM_139236
GI_58743328-S				
scl22891.6.29_2-S	Tcfl1	NM_009336	transcription factor-like 1	NM_009336
scl0098366.1_13-S	Smap1	NM_028534	stromal membrane-associated protein 1	NM_028534
GI_60678283-S				
scl31521.10_99-S	1810054G18Rik	NM_029377	hypothetical protein LOC75660	NM_029377
GI_85986662-S				
scl50606.7.1_142-S	Shd	NM_009168	src homology 2 domain-containing transforming protein D	NM_009168
scl0017289.2_279-S	Mertk	NM_008587	c-mer proto-oncogene tyrosine kinase	NM_008587
scl32292.3.1_56-S	Phox2a	NM_008887	paired-like homeobox 2a	NM_008887
scl023886.1_155-S	Gdf15	NM_011819	growth differentiation factor 15	NM_011819
GI_31981052-S				
scl0001402.1_96-S	Tnip1	NM_021327	TNFAIP3 interacting protein 1	NM_021327
scl38444.3.1_70-S	Phlda1	NM_009344	pleckstrin homology-like domain, family A, member 1	NM_009344
scl00231070.1_184-S	Insig1	NM_153526	insulin induced gene 1	NM_153526
scl0018616.2_167-S	Peg3	NM_008817	paternally expressed 3 isoform 2	NM_0010109
GI_31340603-S				
scl0001319.1_14-S	Hmmr	NM_013552	hyaluronan mediated motility receptor (RHAMM)	NM_013552
scl000775.1_119-S	Pappa2	XM_355248	pappalysin 2	XM_355248
scl019247.3_3-S	Ptpn11	NM_011202	protein tyrosine phosphatase, non-receptor type 11	NM_011202
scl22226.9.1_81-S	Nudt6	NM_153561	nudix (nucleoside diphosphate linked moiety X)-type motif 6	NM_153561
scl0004188.1_86-S	Zfp326	NM_018759	zinc finger protein 326	NM_018759
scl00116848.1_137-S	Baz2a	NM_054078	bromodomain adjacent to zinc finger domain, 2A	NM_054078
GI_55769575-I				
scl31141.20.4_13-S	Anpep	NM_008486	alanyl (membrane) aminopeptidase	NM_008486
GI_84993769-S				
scl016477.1_58-S	Junb	NM_008416	Jun-B oncogene	NM_008416
scl19750.8.1_24-S	Thedc1	NM_145921	thioesterase domain containing 1	NM_145921
scl00105148.1_7-S	Iars	NM_172015	isoleucine-tRNA synthetase	NM_172015
scl00074.1_47-S	Kirrel2	NM_172898	kin of IRRE-like 2	NM_172898
scl33714.1_27-S	Jund1	NM_010592	jun D proto-oncogene	NM_010592
GI_9055307-A				
scl0013688.1_69-S	Eif4ebp2	NM_010124	eukaryotic translation initiation factor 4E binding protein 2	NM_010124
scl0270802.1_64-S	BC048403	NM_173022	hypothetical protein LOC270802	NM_173022
GI_49227375-S				
GI_51921290-A				
scl20561.14_317-S	E430002G05Rik	NM_173749	regeneration associated muscle protease	NM_173749
scl0002764.1_60-S	Nol6	NM_139237	nucleolar RNA-associated protein long isoform	NM_139236
scl38403.1.6_135-S	Lrrc10	NM_146242	leucine rich repeat containing 10	NM_146242
scl25096.6.1_229-S	Tal1	NM_011527	T-cell acute lymphocytic leukemia 1	NM_011527
GI_27734061-S				
scl0016969.1_242-S	Zbtb7	NM_010731	zinc finger and BTB domain containing 7	NM_010731
scl25876.6.1_3-S	Mospd3	NM_030037	motile sperm domain containing 3	NM_030037
scl38644.7_75-S	Aes	NM_010347	amino-terminal enhancer of split	NM_010347
scl0066119.2_58-S	1110002E23Rik	NM_025365	hypothetical protein LOC66119	NM_025365
scl37305.10.1_6-S	Mmp10	NM_019471	matrix metalloproteinase 10	NM_019471
scl40317.1.1637_8-S		XM_147531		
scl33632.3.219_84-S	D8Ertd69e	XM_194424	HIV-1 induced protein HIN-1 isoform 1	XM_194424
GI_58801305-S				
scl39273.6_263-S	Lgals3bp	NM_011150	lectin, galactoside-binding, soluble, 3 binding protein	NM_011150
scl014313.3_29-S	Fst	NM_008046	follistatin	NM_008046
scl000180.1_17-S	Numbl	NM_010950	numb-like	NM_010950
scl28780.8_93-S	Cyp26b1	NM_175475	cytochrome P450, family 26, subfamily b, polypeptide 1	NM_175475
GI_31342014-S				
scl24507.4_95-S	Prdm13	XM_131309	similar to PR domain containing 13 isoform 1	XM_131309
GI_85702174-A				
scl013680.2_48-S	Ddx19a	NM_007916	Ddx19-like protein	NM_007916
scl0001799.1_81-S	Pcqap	NM_033609	positive cofactor 2, glutamine/Q-rich-associated protein	NM_033609
scl018223.10_5-S	Numbl	NM_010950	numb-like	NM_010950

scI018173.9_2-S	Slc11a1	NM_013612	solute carrier family 11 (proton-coupled divalent metal ion transporte	NM_013612
scI24183.14_264-S	Nfib	NM_008687	nuclear factor I/B	NM_008687
GI_31343133-S				
scI49315.8.1_6-S	Ahsg	NM_013465	alpha-2-HS-glycoprotein	NM_013465
GI_58743352-S				
scI067291.1_2-S	3110023B02Rik	NM_152807	hypothetical protein LOC67291	NM_152807
scI020588.27_42-S	Smarcc1	NM_009211	SWI/SNF related, matrix associated, actin dependent regulator of c	NM_009211
scI51773.17.1_8-S	Mbd1	NM_013594	methyl-CpG binding domain protein 1	NM_013594
scI26701.7.1_2-S	Tnip2	NM_139064	TNFAIP3 interacting protein 2	NM_139064
scI000925.1_8-S	Pctk3	NM_008795	PCTAIRE-motif protein kinase 3	NM_008795
scI40193.3.1_118-S	Hand1	NM_008213	heart and neural crest derivatives expressed transcript 1	NM_008213
scI29506.6.9_30-S	Chd4	NM_145979	chromodomain helicase DNA binding protein 4	NM_145979
scI0003150.1_14-S		NM_133852		
GI_58801287-S				
scI0002693.1_3-S	Mutyh	NM_133250	mutY homolog	NM_133250
scI34628.21.1_19-S	Arhgap10	NM_030113	Rho GTPase activating protein 10	NM_030113
scI077781.2_11-S	Epm2aip1	NM_175266	EPM2A (laforin) interacting protein 1	NM_175266
scI25795.24.1_9-S	2210010N04Rik	XM_149712	hypothetical protein LOC70381 isoform 1	XM_149712
GI_58801355-S				
scI48157.24.1_63-S	Wdr9	NM_145125	bromodomain and WD repeat domain containing 1	NM_145125
scI0019727.2_202-S	Rfxank	NM_011266	regulatory factor X-associated ankyrin-containing protein isoform 2	NM_0010255
scI018146.8_16-S	Npdc1	NM_008721	neural proliferation, differentiation and control gene 1	NM_008721
scI012013.5_256-S	Bach1	NM_007520	BTB and CNC homology 1	NM_007520
scI075786.2_11-S	Ckap5	XM_130287	cytoskeleton associated protein 5	
scI0003679.1_4-S	Col4a3bp	NM_023420	procollagen, type IV, alpha 3 (Goodpasture antigen) binding protein	NM_023420
scI22801.5.1_50-S	Ngfb	NM_013609	nerve growth factor, beta	NM_013609
scI26994.6.1_84-S	Nptx2	NM_016789	neuronal pentraxin 2	NM_016789
scI0014221.2_140-S	Fjx1	NM_010218	four jointed box 1	NM_010218
scI0017755.1_176-S	Mtap1b	NM_008634	microtubule-associated protein 1 B	NM_008634
scI31730.8.4_24-S	Sepw1	NM_009156	selenoprotein W, muscle 1	NM_009156
scI50509.4_211-S	Lbh	NM_029999	limb-bud and heart	NM_029999
scI00229542.2_217-S	C430014D17Rik	NM_139304	transcription repressor p66 component of the MeCP1 complex	NM_139304
GI_59797055-S				
scI00114716.1_17-S	Spred2	NM_033523	sprouty-related protein with EVH-1 domain 2	NM_033523
scI52748.5.1_9-S	2900055D14Rik	NM_028411	hypothetical protein LOC72982	NM_028411
scI0078428.2_263-S	A030010B05Rik	NM_030100	within bgcn homolog	NM_030100
scI50739.5_359-S	Zfp57	NM_009559	zinc finger protein 57	NM_0010137
scI0001956.1_0-S	Elf2	NM_023502	ets family transcription factor ELF2A2	NM_023502
scI46098.2.1_25-S	Htr2a	NM_172812	5-hydroxytryptamine (serotonin) receptor 2 A	NM_172812
scI0003840.1_14-S	Ppp1r14c	NM_133485	PKC-potentiated PP1 inhibitory protein	NM_133485
scI015284.1_28-S	Hlx1	NM_008250	H2.0-like homeo box gene	NM_008250
scI018550.2_67-S	Furin	NM_011046	furin (paired basic amino acid cleaving enzyme)	NM_011046
scI0004115.1_39-S	Add1	NM_013457	adducin 1 (alpha) isoform 1	NM_0010244
scI27354.15.1_3-S		XM_355680		
GI_56606022-S				
scI021869.1_239-S	Titf1	NM_009385	thyroid transcription factor 1	NM_009385
scI53162.3.1_182-S	Gpr120	NM_181748	G protein-coupled receptor 120	NM_181748
scI0003780.1_7-S	Akap12	AK053844	A kinase (PRKA) anchor protein (gravin) 12	NM_031185
scI00231798.1_133-S	Lrch4	NM_146164	leucine rich repeat protein 4, neuronal	NM_146164
GI_84370287-S				
scI55067.6_230-S	Timm17b	NM_011591	translocator of inner mitochondrial membrane 17b	NM_011591
GI_83816896-I				
scI00076.1_24-S	Mlstd2	NM_027379	male sterility domain containing 2	NM_026143
scI49931.1.1_57-S	Olfr99	NM_146515	olfactory receptor 99	NM_146515
scI0381085.13_66-S	BC045600	NM_198647	TBC1 domain family, member 22B	NM_198647
GI_72384360-S				
scI076800.3_8-S	Usp42	XM_132483	ubiquitin specific protease 42	
scI0328801.4_107-S	0610030H11Rik	NM_026712	zinc finger protein 414	NM_026712
scI054204.2_14-S	Sep-01	NM_017461	septin 1	NM_017461
GI_13385611-S				
scI00118449.1_23-S		NM_080451		
scI0068501.1_28-S	1110014D18Rik	NM_026746	hypothetical protein LOC68501	NM_026746
scI46547.15.1_55-S	Anxa11	NM_013469	annexin A11	NM_013469
scI28856.13.1_8-S	Tcf3	NM_009332	transcription factor 3	NM_009332
GI_58801413-S				
scI35386.1.1_202-S		NM_027488		
scI000931.1_1146-S	Bmpr2	NM_007561	bone morphogenic protein receptor, type II (serine/threonine kinase)	NM_007561
IGLC1_J00587_Ig_lambda_constant_1		XM_148393		
scI48798.3.1_6-S	1500031H01Rik	XM_358753	hypothetical protein LOC207740	XM_358753
scI18858.8.1_25-S	Sgne1	NM_009162	secretory granule neuroendocrine protein 1, 7B2 protein	NM_009162
GI_71480159-S				
GI_77861890-S				
scI030839.9_291-S	Fbxw5	NM_013908	F-box and WD-40 domain protein 5	NM_013908
scI53823.22.1_4-S	Nxf7	NM_130888	nuclear RNA export factor 7 isoform 2	NM_130888
scI19424.2.1_75-S		XM_149113		
scI020354.1_53-S	Sema4d	NM_013660	semaphorin 4D	NM_013660
GI_22129490-S				
scI18860.2_108-S	Grem1	NM_011824	gremlin 1	NM_011824
scI53147.9.1_60-S	Cyp2c39	NM_010003	cytochrome P450, family 2, subfamily c, polypeptide 39	NM_010003
scI41500.10.1_4-S	Cias1	NM_145827	cold autoinflammatory syndrome 1 homolog	NM_145827
scI47828.2.1_30-S	4930572J05Rik	NM_198607	mesenchymal stem cell protein DSCD75 homolog	NM_198607
scI37695.8.1_48-S	Nfic	NM_008688	nuclear factor I/C isoform a	NM_008688

Eleph VP16 up				
Gene Name	Common	Genbank	Product	RefSeq
scl0004206.1_0-S	BC034507	XM_131888	claudin 12	
scl030839.9_291-S	Fbxw5	NM_013908	F-box and WD-40 domain protein 5	NM_013908
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl51379.4_394-S	2010002N04Rik	NM_134133	putative small membrane protein NID67	NM_134133
scl0001399.1_22-S	Prpsap2	NM_144806	phosphoribosyl pyrophosphate synthetase-associated protein 2	NM_144806
scl021667.1_237-S	Tdgf1	NM_011562	teratocarcinoma-derived growth factor	NM_011562
scl067106.7_0-S	Arch	NM_025970	zinc finger and BTB domain containing 8 opposite strand	NM_025970
scl0022217.2_66-S	Usp12	NM_011669	ubiquitin specific protease 12	NM_011669
scl41083.21_226-S	Mtmr4	NM_133215	myotubularin related protein 4	NM_133215
scl24618.9_190-S	A530082C11Rik	NM_177186	solute carrier family 35, member E2	NM_177186
scl011532.9_123-S	Adh5	NM_007410	alcohol dehydrogenase 5 (class III), chi polypeptide	NM_007410
scl0067845.2_211-S	Zfp364	NM_026406	Rabring 7	NM_026406
scl36686.18.1_100-S	Gclc	NM_010295	glutamate-cysteine ligase, catalytic subunit	NM_010295
scl46893.13.1_7-S	AV014541; 6330	NM_178869	tubulin tyrosine ligase-like 1	NM_178869
scl00278672.2_86-S	1110051B16Rik	NM_183389	hypothetical protein LOC278672	NM_183389
scl0001603.1_1171-S	Arl6ip2	NM_178050	ADP-ribosylation factor-like 6 interacting protein 2 isoform 1	NM_019717
scl29111.14_3-S	Slc37a3	NM_028123	solute carrier family 37 (glycerol-3-phosphate transporter), member	NM_028123
scl0067475.1_65-S	1300013B24Rik	NM_026184	endoplasmic oxidoreductase 1 beta	NM_026184
scl36521.18.1_56-S		XM_356182		
scl0019684.1_297-S	Rdx	NM_009041	radixin	NM_009041
scl019656.1_169-S	Rbmxrt	NM_009033	RNA binding motif protein, X chromosome retrogene	NM_009033
scl45217.22.1_37-S	2810028N01Rik	NM_028315	RIKEN cDNA 2810028N01	NM_028315
scl071275.2_75-S	4933437F05Rik	XM_127023	hypothetical protein LOC71275	XM_127023
scl36160.67.1_2-S	Col5a3	NM_016919	procollagen, type V, alpha 3	NM_016919
scl33661.14.3_4-S	Tom1	NM_011622	target of myb1 homolog	NM_011622
scl50308.28_413-S	Map3k4	NM_011948	mitogen activated protein kinase kinase kinase 4	NM_011948
scl00239217.2_149-S	Kctd12	NM_177715	potassium channel tetramerisation domain containing 12	NM_177715
scl36985.26_298-S	Usp28	NM_175482	ubiquitin specific protease 28	NM_175482
scl37245.7_152-S	2210010B09Rik	NM_172919	hypothetical protein LOC244721	NM_172919
scl022631.1_10-S	Ywhaz	NM_011740	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation	NM_011740
scl0070373.1_243-S	1700020O03Rik	NM_027405	hypothetical protein LOC70373	NM_027405
scl0070675.2_209-S	5730538E15Rik	NM_173443	valosin-containing protein (p97)/p47 complex-interacting protein p13	NM_173443
scl013627.1_210-S	Eef1a1	XM_203909	eukaryotic translation elongation factor 1 alpha 1	
scl47241.11_30-S	4921532K09Rik	NM_026149	chronic myelogenous leukemia tumor antigen 66	NM_026149
scl013532.4_62-S	Dub2	NM_010089	deubiquitinating enzyme 2	NM_010089
scl0015519.1_14-S	Hspca	NM_010480	heat shock protein 1, alpha	NM_010480
scl072139.1_117-S	2610044O15Rik	NM_153780	hypothetical protein LOC72139	NM_153780
scl17554.3.1_315-S	En1	NM_010133	engrailed 1	NM_010133
scl0022221.2_82-S	Ubp1	NM_013699	upstream binding protein 1	NM_013699
scl0026442.1_315-S	Psma5	NM_011967	proteasome (prosome, macropain) subunit, alpha type 5	NM_011967
Eleph VP16 down				
Gene Name	Common	Genbank	Product	RefSeq
scl41500.10.1_4-S	Cias1	NM_145827	cold autoinflammatory syndrome 1 homolog	NM_145827
scl38650.11.1_216-S		XM_125716		
scl17644.2.1_29-S	Twist2	NM_007855	twist homolog 2	NM_007855
scl39344.12.1_117-S	Fdxr	NM_007997	ferredoxin reductase	NM_007997
scl0013222.1_296-S	Defcr-rs2	NM_007847	defensin related cryptdin, related sequence 2	NM_007847
scl49357.1.362_28-S		XM_147222		
scl33485.7_10-S	Herpud1	NM_022331	homocysteine-inducible, endoplasmic reticulum stress-inducible, ut	NM_022331
scl45267.5.1_68-S	1190002H23Rik	NM_025427	response gene to complement 32	NM_025427
scl0001660.1_2-S	Jmjd2b	NM_172132	jumonji domain containing 2B	NM_172132
scl30286.11_185-S	Irf5	NM_012057	interferon regulatory factor 5	NM_012057
scl39611.13_705-S	9930017A07Rik	NM_172564	C-terminal tensin-like	NM_172564
scl00108086.1_95-S	Ubce7ip1	NM_080561	ubiquitin conjugating enzyme 7 interacting protein 1	NM_080561
scl38761.18.1_172-S	Ggt1	NM_008116	gamma-glutamyltransferase 1	NM_008116
scl20391.15_58-S	B830009D23Rik	NM_175285	transmembrane protein 62	NM_175285
scl42430.2_236-S	Foxa1	NM_008259	forkhead box A1	NM_008259
scl33021.4.288_87-S	Pglyrp1	NM_009402	peptidoglycan recognition protein 1	NM_009402
scl40626.1.1_194-S	5730593N15Rik	NM_175263	hypothetical protein LOC77583	NM_175263
scl41016.3_521-S	Dlx3	NM_010055	distal-less homeobox 3	NM_010055
scl0001526.1_22-S	Irf1	NM_008390	interferon regulatory factor 1	NM_008390
scl16260.3.61_18-S	Elf3	NM_007921	E74-like factor 3	NM_007921
scl0230868.3_5-S	BC055811	NM_198610	immunoglobin superfamily, member 21	NM_198610
scl31869.6.1_15-S	Tnni2	NM_009405	troponin I, skeletal, fast 2	NM_009405
scl000238.1_112-S	Stard10	NM_019990	START domain containing 10	NM_019990
scl27140.1_204-S	Cldn3	NM_009902	claudin 3	NM_009902
GI_46909570-S	Gata6	NM_010258	GATA binding protein 6	NM_010258
scl31868.19.1_21-S	Tnnt3	NM_011620	troponin T3, skeletal, fast	NM_011620
GI_6753645-S	Dlx2	NM_010054	distal-less homeobox 2	NM_010054
scl40193.3.1_118-S	Hand1	NM_008213	heart and neural crest derivatives expressed transcript 1	NM_008213

Eleph EnR up				
Gene Name	Common	Genbank	Product	RefSeq
scl0004206.1_0-S	BC034507	XM_131888	claudin 12	
scl40193.3.1_118-S	Hand1	NM_008213	heart and neural crest derivatives expressed transcript 1	NM_008213
scl28780.8_93-S	Cyp26b1	NM_175475	cytochrome P450, family 26, subfamily b, polypeptide 1	NM_175475
scl41016.3_521-S	Dlx3	NM_010055	distal-less homeobox 3	NM_010055
Eleph EnR down				
Gene Name	Common	Genbank	Product	RefSeq
scl39344.12.1_117-S	Fdxr	NM_007997	ferredoxin reductase	NM_007997
scl0001008.1_94-S	Il24	NM_053095	interleukin 24	NM_053095
scl17644.2.1_29-S	Twist2	NM_007855	twist homolog 2	NM_007855
scl40483.14.1041_16-	Meis1	NM_010789	myeloid ecotropic viral integration site 1	NM_010789
scl023886.1_155-S	Gdf15	NM_011819	growth differentiation factor 15	NM_011819
scl0330941.1_271-S	AI593442	NM_177907	hypothetical protein LOC330941 isoform 1	NM_177907
scl00237221.1_81-S	BC023488	NM_146238	hypothetical protein LOC237221	NM_146238
scl27354.15.1_3-S		XM_355680		
scl54438.4.1_134-S	2010001H14Rik	NM_027227	hypothetical protein LOC69824	NM_027227
scl25839.9.1_1-S	BC019731	NM_144914	hypothetical protein LOC231832	NM_144914
scl25883.15_188-S	Slc12a9	NM_031406	solute carrier family 12 (potassium/chloride transporters), member 9	NM_031406
scl42842.7.1_0-S	D12Ertd647e	NM_026790	hypothetical protein LOC52668 isoform 1	NM_026790
scl0013222.1_296-S	Defcr-rs2	NM_007847	defensin related cryptdin, related sequence 2	NM_007847
scl40626.1.1_194-S	5730593N15Rik	NM_175263	hypothetical protein LOC77583	NM_175263

Platy VP16 up				
Gene Name	Common	Genbank	Product	RefSeq
scl0004206.1_0-S	BC034507	XM_131888	claudin 12	
scl40901.16_69-S	Stat5a	NM_011488	signal transducer and activator of transcription 5A	NM_011488
scl013627.1_210-S	Eef1a1	XM_203909	eukaryotic translation elongation factor 1 alpha 1	
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl33661.14.3_4-S	Tom1	NM_011622	target of myb1 homolog	NM_011622
scl000319.1_7-S	Acin1	NM_019567	apoptotic chromatin condensation inducer 1 isoform 1	NM_019567
scl0012514.2_91-S	Cd68	NM_009853	CD68 antigen	NM_009853
scl0001298.1_25-S	1300013J15Rik	NM_026183	hypothetical protein LOC67473	NM_026183
scl41083.21_226-S	Mtmr4	NM_133215	myotubularin related protein 4	NM_133215
scl45568.9.1_28-S	D14Ertd500e	NM_145462	hypothetical protein LOC219072	NM_145462
scl0319939.2_99-S	Tens1	XM_109868	tensin 3	XM_109868
scl24282.10_504-S	Ltb4dh	NM_025968	leukotriene B4 12-hydroxydehydrogenase	NM_025968
scl0003932.1_84-S	Si	NM_021882	silver	NM_021882
scl0003150.1_14-S		NM_133852		
scl50973.6_371-S	2310051D06Rik	NM_028009	RNA pseudouridylate synthase domain containing 1	NM_028009
scl083456.10_6-S	Mov10l1	NM_031260	Moloney leukemia virus 10-like 1	NM_031260
scl015574.1_13-S	Hus1	NM_008316	Hus1 homolog	NM_008316
scl0003648.1_67-S	Zfp346	NM_012017	zinc finger protein 346	NM_012017
scl0020771.1_0-S	Spt2	NM_009268	salivary protein 2	NM_009268
scl15995.6_195-S	C130085G02Rik	XM_136364	dual specificity phosphatase 27 (putative)	
scl074763.7_75-S	1200013P24Rik	NM_029090	hypothetical protein LOC74763	NM_029090
scl0018752.1_62-S	Prkcc	NM_011102	protein kinase C, gamma	NM_011102
scl0001399.1_22-S	Prpsap2	NM_144806	phosphoribosyl pyrophosphate synthetase-associated protein 2	NM_144806
scl000720.1_16-S	Hmgb2l1	NM_178017	high mobility group box 2-like 1	NM_178017
scl0001603.1_1171-S	Arl6ip2	NM_178050	ADP-ribosylation factor-like 6 interacting protein 2 isoform 1	NM_019717
scl20206.6_190-S	Dstn	NM_019771	destrin	NM_019771
scl0003023.1_84-S	Fnbp4	NM_018828	formin binding protein 4	NM_018828
scl49753.2.1_144-S	Pspn	NM_008954	persephin	NM_008954
scl0004106.1_50-S	Usp46	NM_177561	ubiquitin specific protease 46	NM_177561
scl47614.23.1_38-S	Shank3	NM_021423	SH3/ankyrin domain gene 3	NM_021423
scl17185.20.1_57-S	Fmn2	NM_019445	formin 2	NM_019445
scl0067291.1_12-S	3110023B02Rik	NM_152807	hypothetical protein LOC67291	NM_152807

Platy VP16 down				
Gene Name	Common	Genbank	Product	RefSeq
scl35124.5_47-S	Efnb2	NM_010111	ephrin B2	NM_010111
scl30469.7_1-S	Igf2	NM_010514	insulin-like growth factor 2	NM_010514
scl21948.4.1_81-S	Rga	NM_009057	recombination activating gene 1 gene activation	NM_009057
scl33841.11_453-S	Irf2	NM_008391	interferon regulatory factor 2	NM_008391
scl0114128.8_210-S	Laptm4b	NM_033521	lysosomal-associated protein transmembrane 4B	NM_033521
scl0001660.1_2-S	Jmjd2b	NM_172132	jumonji domain containing 2B	NM_172132
scl0011658.2_92-S	Alcam	NM_009655	activated leukocyte cell adhesion molecule	NM_009655
scl0013078.2_259-S	Cyp1b1	NM_009994	cytochrome P450, family 1, subfamily b, polypeptide 1	NM_009994
scl0014972.1_210-S	H2-K1	NM_001001E	histocompatibility 2, K1, K region	NM_001001E
scl40203.12_25-S	Sparc	NM_009242	secreted acidic cysteine rich glycoprotein	NM_009242
scl0381413.1_190-S	Gm1012	NM_201367	G protein-coupled receptor 176	NM_201367
scl0081003.2_293-S	Trim23	NM_030731	tripartite motif protein 23	NM_030731
scl023886.1_155-S	Gdf15	NM_011819	growth differentiation factor 15	NM_011819
scl43090.8_531-S	Rhoj	NM_023275	ras homolog gene family, member J	NM_023275
gi_7305154_ref_NM_0	Hprt1	NM_013556	hypoxanthine guanine phosphoribosyl transferase 1	NM_013556
scl38210.4_147-S	Rab32	NM_026405	RAB32	NM_026405
scl38108.3_14-S	C030003D03Rik	XM_282904	hypothetical protein LOC77220	
scl26371.6_22-S	Cxcl10	NM_021274	chemokine (C-X-C motif) ligand 10	NM_021274
scl35919.5_337-S	Tagln	NM_011526	transgelin	NM_011526
scl37307.8.1_29-S	Mmp3	NM_010809	matrix metalloproteinase 3	NM_010809
scl30211.10.1_117-S	Stra8	NM_009292	stimulated by retinoic acid gene 8	NM_009292
scl32431.24_246-S	Nox4	NM_015760	NADPH oxidase 4	NM_015760
scl34187.26.1_95-S	Nup133	NM_172288	nucleoporin 133	NM_172288
scl0012837.1_129-S	Col8a1	NM_007739	procollagen, type VIII, alpha 1	NM_007739
scl39611.13_705-S	9930017A07Rik	NM_172564	C-terminal tensin-like	NM_172564
scl27567.9_262-S	Ccng2	NM_007635	cyclin G2	NM_007635
scl31061.7_464-S	2310046G15Rik	NM_029614	protease, serine, 23	NM_029614
scl24923.6.1_23-S	1810017F10Rik	NM_025452	beta-casein-like protein	NM_025452
GI_46909570-S	Gata6	NM_010258	GATA binding protein 6	NM_010258
GI_6753645-S	Dlx2	NM_010054	distal-less homeobox 2	NM_010054
scl41016.3_521-S	Dlx3	NM_010055	distal-less homeobox 3	NM_010055
scl053606.2_17-S	G1p2	NM_015783	interferon, alpha-inducible protein	NM_015783
scl44946.1_175-S	Foxq1	NM_008239	forkhead box Q1	NM_008239
scl34033.37.1_91-S	Myom2	NM_008664	myomesin 2	NM_008664
scl47854.5_552-S	Wisp1	NM_018865	WNT1 inducible signaling pathway protein 1	NM_018865
scl020296.2_11-S	Ccl2	NM_011333	chemokine (C-C motif) ligand 2	NM_011333
scl24919.4.1_260-S	Sync	NM_023485	syncoilin	NM_023485
scl0268903.1_0-S	Nrip1	NM_173440	nuclear receptor interacting protein 1	NM_173440
scl33556.12_71-S	Gpt2	NM_173866	glutamic pyruvate transaminase (alanine aminotransferase) 2	NM_173866
scl18860.2_108-S	Grem1	NM_011824	gremlin 1	NM_011824
scl0002109.1_724-S	Dnajb4	NM_025926	DnaJ (Hsp40) homolog, subfamily B, member 4	NM_025926
scl35747.13.1_171-S	B230114P05Rik	NM_172444	hypothetical protein LOC207596	NM_172444
scl40193.3.1_118-S	Hand1	NM_008213	heart and neural crest derivatives expressed transcript 1	NM_008213
scl42842.7.1_0-S	D12Ertd647e	NM_026790	hypothetical protein LOC52668 isoform 1	NM_026790
scl34884.5.1_26-S	Frg1	NM_013522	FSHD region gene 1	NM_013522
scl52545.7.1_87-S		NM_181587		
scl45267.5.1_68-S	1190002H23Rik	NM_025427	response gene to complement 32	NM_025427
scl000050.1_4-S	Gpr124	NM_054044	G protein-coupled receptor 124	NM_054044
scl45278.7.1_11-S	Dnajd1	NM_025384	DnaJ (Hsp40) homolog, subfamily D, member 1	NM_025384
scl29554.6.1_30-S	Usp18	NM_011909	ubiquitin specific protease 18	NM_011909
scl00399591.1_43-S	4930488E11Rik	NM_207267	hypothetical protein LOC399591	NM_207267

Platy EnR up				
Gene Name	Common	Genbank	Product	RefSeq
scl0004206.1_0-S	BC034507	XM_131888	claudin 12	
scl013627.1_210-S	Eef1a1	XM_203909	eukaryotic translation elongation factor 1 alpha 1	
scl0018044.2_32-S	Nfya	NM_010913	nuclear transcription factor-Y alpha	NM_010913
scl018715.2_20-S	Pim2	NM_138606	serine-threonine protein kinase pim-2 isoform 1	NM_138606
scl0001399.1_22-S	Prpsap2	NM_144806	phosphoribosyl pyrophosphate synthetase-associated protein 2	NM_144806
scl50149.11.1_28-S	Pdip	XM_128552	protein disulfide isomerase associated 2	XM_128552
scl39640.3.1_127-S	1700001P01Rik	XM_126645	hypothetical protein LOC72215	
scl15995.6_195-S	C130085G02Rik	XM_136364	dual specificity phosphatase 27 (putative)	
scl0004065.1_58-S	Abhd1	NM_021304	abhydrolase domain containing 1	NM_021304
scl47614.23.1_38-S	Shank3	NM_021423	SH3/ankyrin domain gene 3	NM_021423
scl0016619.1_79-S	Klk27	NM_020268	kallikrein 27	NM_020268
scl28889.10.1_70-S	MGC59076	NM_178413	hypothetical protein LOC232078	NM_0010339
scl0019240.2_329-S	Tmsb10	NM_025284	thymosin, beta 10	NM_025284
scl0002466.1_124-S	Fbxl6	NM_013909	F-box and leucine-rich repeat protein 6	NM_013909
scl00217219.2_216-S	BC025575	NM_199200	hypothetical protein LOC217219	NM_199200
scl0001287.1_52-S	Ascc2	NM_029291	ASC-1 complex subunit P100	NM_029291
scl0000117.1_15-S	Taf6	NM_009315	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associat	NM_009315
scl067106.7_0-S	Arch	NM_025970	zinc finger and BTB domain containing 8 opposite strand	NM_025970
scl0003902.1_2-S	Snx3	NM_017472	sorting nexin 3	NM_017472
scl000319.1_7-S	Acin1	NM_019567	apoptotic chromatin condensation inducer 1 isoform 1	NM_019567
Platy EnR down				
Gene Name	Common	Genbank	Product	RefSeq
scl42956.7_401-S	Jundm2	NM_030887	Jun dimerization protein 2	NM_030887
scl000304.1_6-S	Isgf3g	NM_008394	interferon dependent positive acting transcription factor 3 gamma	NM_008394
scl0066643.1_330-S	Lix1	NM_025681	limb expression 1 homolog	NM_025681
scl40441.14.1_29-S	0610010F05Rik	XM_203572	RIKEN cDNA 0610010F05	
scl43090.8_531-S	Rhoj	NM_023275	ras homolog gene family, member J	NM_023275
scl18860.2_108-S	Grem1	NM_011824	gremlin 1	NM_011824
scl18732.66_476-S	Fbn1	NM_007993	fibrillin 1	NM_007993
scl0016998.2_125-S	Ltbp3	NM_008520	latent transforming growth factor beta binding protein 3	NM_008520
scl0003.1_30-S	Smpd1	NM_011421	sphingomyelin phosphodiesterase 1, acid lysosomal	NM_011421
scl0109113.3_5-S	Uhrf2	NM_144873	Np95-like ring finger protein	NM_144873
scl42842.7.1_0-S	D12Ertd647e	NM_026790	hypothetical protein LOC52668 isoform 1	NM_026790
scl0001034.1_561-S	Crbn	NM_021449	cereblon isoform 1	NM_021449
scl0067845.2_211-S	Zfp364	NM_026406	Rabring 7	NM_026406
scl29554.6.1_30-S	Usp18	NM_011909	ubiquitin specific protease 18	NM_011909
scl36160.67.1_2-S	Col5a3	NM_016919	procollagen, type V, alpha 3	NM_016919
scl0001526.1_22-S	Irf1	NM_008390	interferon regulatory factor 1	NM_008390
scl019073.1_109-S	Prg1	NM_011157	proteoglycan 1, secretory granule	NM_011157

Appendix E

Mouse iPS Cell Protocol

Required Materials:

Please prepare the below reagents before starting the protocol.

(A) Viral packaging cells (Plat-E)

Prepare FP medium with the following components:

Media components	Amount for 500ml
10% FBS	50ml
50U and 50mg ml^{-1} Pen/Strep	2.5ml
DMEM containing 4.5gl^{-1} glucose	Fill to 500ml

Blasticidin S hydrochloride

Dissolve in distilled water at 10 mg ml^{-1} and sterilize through a 0.22µm filter. Aliquot and store at -20°C.

Puromycin

Dissolve in distilled water at 10 mg ml^{-1} and sterilize through a 0.22µm filter. Aliquot and store at -20°C.

Polybrene (Hexadimethrine bromide)

Dissolve 0.8g of polybrene in 10ml of distilled water for a 10X stock (80mg ml^{-1}). Dilute 1ml of 10X stock solution with 9ml of distilled water, filter with a 0.22µm filter. Store at 4°C.

~~~~~~
~~~~~~

(B) Fibroblasts (balb/c)

Prepare Mef Medium with the following components:

Media components	Amount for 500ml
10% FBS	50ml
50U and 50mg ml^{-1} Pen/Strep	2.5ml
L-glutamine	5ml
DMEM containing 4.5gl^{-1} glucose	Fill to 500ml

~~~~~~

(C) ES colonies

Prepare ES medium with the following components:

Media components	Amount for 500ml
15% FBS	75ml
50U and 50mg ml^{-1} Pen/Strep	2.5ml
L-glutamine	5ml
NEAA	5ml
2-mercaptoenthanol	1ml
DMEM containing 4.5gl^{-1} glucose	Fill to 500ml
LIF	2ml

~~~~~~

Induction of Pluripotent Stem Cells from Fibroblast Cells (Modified from Tahira’s protocol)

A. Plat-E Production

This procedure takes around 3 days, depending on the cell number required.

Thawing Plat-E (Platinum-E) cells

1. Prepare 10ml of FP medium in a 15-ml tube. Prepare a 15-cm tissue culture dish (no need to gelatin-coat).
2. Remove vial of frozen Plat-E stocks from the liquid nitrogen tank and put the vial in a 37°C water bath until most (but not all, a small portion still thawing) cells are thawed. Note that each tube contains **6 million** Plat-E cells.
3. Wipe the vial with ethanol and transfer the cells to the 15-ml tube with FP medium.
4. Centrifuge at 180g for 5min and remove supernatant.
5. Resuspend the cells with 10ml of FP medium and transfer to the 15-cm plate. Incubate the cells in a 37°C, 5% CO_2 incubator.
6. The next day, replace the medium with new FP medium supplemented with puromycin and blasticidin S hydrochloride. For a 20ml FP medium, add 20µl of 1mg ml^{-1} puromycin stock and 20µl of 10mg ml^{-1} .(Note: Add the puromycin and blasticidin freshly to the medium for each time usage.)

Passaging Plat-E cells

1. Aspirate the spent medium and add 20ml of PBS. Rinse the surface of the cells with PBS and aspirate. Add 4ml of 0.05% trypsin and incubate for 3 min in the 37°C incubator.

2. Detach cells from the flask by tapping and inactivate trypsin with 20ml of FP medium and break cells into single cell suspension by pipetting up and down several times. Seed them into new 15cm dishes (Up to 1:4 ratio).
3. Passage Plat-E cells until sufficient cell number is produced. Note that each confluent 15cm dish contains approx. **20 million** live cells.

B. Retrovirus Production

This procedure takes about four days.

Day ONE: Seeding the appropriate number of Plat-E cells

Note: FP culture does not contain puromycin or blasticidin. These antibiotics will not be used from this point onwards.

1. Aspirate the spent medium and wash the cells with 20ml of PBS. Aspirate the PBS and add 0.05% trypsin and incubate for 3 min in the 37°C incubator. Prepare a number of 10-cm tissue culture dishes as required.
2. After incubation add 20ml of FP medium and dislodge the cells into single cell suspension. Transfer the cells into a 50ml tube.
3. Centrifuge the cells at 180g for 5 min.
4. Discard the supernatant and break the pellet by finger tapping and add appropriate volume of FP medium.
5. Count the number of cells and seed cells at **8 million** cells (in 10ml of FP medium) per 10cm dish and incubate overnight at 37°C, 5% CO_2 incubator.

(Note: At least one Plat-E dish should be prepared for one pMX plasmid DNA. Eg: If you have four pMXs plasmid DNA (encoding Oct3/4, Sox2, Klf4 and c-Myc), then you should prepare a minimum of four plates of Plat-E cells for transfection.)

DAY TWO: Transfection of pMXs plasmid DNA into Plat-E

1. Transfer 0.3ml of DMEM into a 1.5ml eppendorf tube (Alternatively you can prepare a master mix in a 15-cm tube).
2. Add 27µl of Fugene 6 transfection reagent per 0.3ml of DMEM. Incubate for 5min at room temperature.
3. Add 9µg of pMXs plasmid DNA (encoding Oct3/4, Sox2, Klf4 and c-Myc) drop-by-drop into the Fugene 6/DMEM- containing tube, mix gently by finger tapping and incubate for 15mins.
4. Add the DNA/Fugene 6 complex dropwise into the Plat-E dish and incubate overnight at 37°C, 5% CO_2 incubator. **Also, transfect with a suitable control eg. empty vector. Having a control is critical.**

Also, on this very day, thaw inactivated MEFs onto gelatin coated plates:

1. Coat 6cm dishes with 10ml of gelatin. Incubate for 30mins at room temperature.
2. Prepare 10ml of MEF medium on a 15ml tube.
3. Remove vial of inactivated MEF (frozen down at 2.0×10^6) from liquid nitrogen stock and place it onto the 37°C water bath until most (but not all, a small portion still thawing) cells are thawed.

4. Transfer cells to the tube with MEF medium and centrifuge at 160g for 5mins.
5. Aspirate the supernatant and add appropriate medium to seed cells. Each iMEF tube can used to seed six 6cm dishes.

DAY THREE: Changing spent medium from the Plat-E plates

Note: From this point onwards, standard virus handling procedures are to be followed. The supernatant in Plat-E plates contain retroviruses. Remember to immerse used labware and unwanted cultures in bleach separately before disposal.

1. Aspirate the transfection reagent-containing medium. (Aspirate separately using the pipettor into bleach beaker. DO NOT USE VACUUM SUCTION!)
2. Add 6ml of fresh FP medium and return cells to the incubator.

Also, on this very day, prepare BL6 fibroblasts which will be re-programmed into pluripotent stem cells.

1. MEFs used for re-programming should be of passage <3. Thaw the cells using the MEF medium. Note that each tube contains **3 million** cells.
2. Centrifuge at 160g for 5mins. Aspirate the supernatant and add appropriate volume of MEF medium.
3. Seed approximately **267,000** cells onto each 6cm dish of the inactivated MEFs that were prepared the day before.
4. Incubate the dish overnight at 37°C, 5% CO_2 incubator.

DAY FOUR: Harvesting the viruses and infecting the BL6 fibroblasts

Note: Remember to follow standard virus handling procedures.

1. Collect the medium for the Plat-E cells (~6ml) by using a 10ml sterile disposable syringe, filtering it through a 0.45µm pore size cellulose acetate filter, transferring into a 15ml tube, from each of the pMX plasmid DNA plate.
2. Add 5µl of 8 mg ml^{-1} polybrene solution into the filtered virus-containing medium. Mix gently by pipetting up and down.
3. Make a mixture of equal parts of the medium containing Oct3/4, Sox2, Klf4 and c-Myc retroviruses. **Retroviruses should be used freshly. Do not freeze/thaw the retroviruses as it will decrease the titer of the retrovirus.**
4. Aspirate the medium from the BL6 dishes and add appropriate amounts of the polybrene/virus containing medium. Typically, 1ml of each factor is added to each 6cm dish (ie. 4 factors = 4ml total per dish). For the "Empty vector only control" plates, add 4ml of empty vector virus-containing medium. For the "No Infection control" plates, simply add fresh FP medium. Incubate the cells from 4hrs to overnight at 37°C, 5% CO_2 incubator.

C. Monitoring iPS Progress

DAY FIVE: Changing the spent medium on the BL6 plates

Note: Remember to follow standard virus handling procedures.

1. Aspirate the medium and add fresh FP medium.

DAY SIX – TWENTY: Changing the spent medium on the BL6 plates

Note: No need for virus handling procedure from this point onwards.

1. Aspirate the medium and switch to fresh ES medium. Change medium daily.
2. One day before picking the colonies, thaw inactivated MEFs onto a 24-well plate.

D. Handling iPS Cells

Picking up the iPS colonies

1. Aliquot 20 µl of 0.25% trypsin per well of a 96-well plate.
2. Remove the spent medium from the fibroblast dish and add 10ml of PBS.
3. Aspirate the PBS and add 5ml pf PBS.
4. Pick colonies from the dish using a glass Pasteur pipette and transfer the colonies using a pipetman into the 96-well plate with trypsin. Incubate for 15mins at at 37°C.
5. Add 180 µl of ES medium to each well, and pipette up and down to break up the colony to single cells.
6. Transfer the cell suspension into the well of a 24-well plate with inactivated MEFs. Add 300 µl ES medium and incubate until cells reach 80-90% confluency. At this point they should be passaged into 6-well plates. 6 well plates with inactivated MEFs should be ready a day before the passage.

Expansion of iPS cells

1. Aspirate the medium and wash the cells with 1 ml PBS.
2. Remove PBS completely and add 0.1ml of 0.25% trypsin and incubate at 37°C for 10min.
3. Add 0.4ml of the ES medium and suspend the cells by pipetting up and down to single cell suspension.
4. Transfer the cell suspension to a 6- well plate and add 1.5ml ES medium and incubate at 37°C, 5% CO_2 incubator until cells reach 80-90% confluency in the 6 well plates. At this point, prepare frozen stock of the cells.