ABSTRACT

Title of Document: INVESTIGATING THE ROLES OF OCT4 AND CDX2 IN DIRECTING BOVINE TROPHECTODERM LINEAGE DEVELOPMENT

Andrew Thomas Schiffmacher, Ph.D., 2010

Directed By: Associate Professor Carol L Keefer, Department of Animal and Avian Sciences

Domestic animal embryo technologies would benefit from a better understanding of the molecular mechanisms that direct early embryonic development. Failure to establish a normal transcriptional regulatory program in the early trophectoderm of nuclear transfer or in vitro derived bovine embryos has been implicated as an underlying cause of placental abnormalities and fetal death. Misexpression of trophoblast-specific genes in these embryos has been identified, but the functions and roles of these genes remain poorly understood. The main focus of this study was to study genes involved in bovine trophectoderm lineage development using the bovine trophectoderm derived CT-1 cell line as a genetic model. Specifically, we investigated the roles of the regulatory transcription factors OCT4 and CDX2 in directing the developmental program of the early bovine trophectoderm via gene regulation of other trophectoderm-specific transcription factors. First, we overcame certain technical limitations of CT-1 cells by improving nucleic acid transfection, CT-1 cell dispersal, and culture protocols,

demonstrating for the first time that overexpression assays using Lipitoid are feasible in the hard-to-transfect CT-1 cell line. We expanded the list of trophoblast genes known to be expressed in CT-1 cells and determined that the expression profile was similar to that of the ovoid stage of bovine pre-attachment embryogenesis. We measured relative levels of these genes in response to *OCT4* and *CDX2* overexpression and knockdown. Our results indicated that CDX2 may be a regulator of transcription of many bovine trophoblast genes and should be the focus of further study. We identified a novel *OCT4* retrocopy transcribed into both sense and natural antisense transcripts, which may have a role in post-transcriptionally regulating *OCT4* expression within the early bovine trophoblast. Together, these studies validate the CT-1 cell line as an appropriate genetic model for studying gene regulation in the bovine trophectoderm.

# INVESTIGATING THE ROLES OF OCT4 AND CDX2 IN DIRECTING BOVINE TROPHECTODERM LINEAGE DEVELOPMENT

By

Andrew Thomas Schiffmacher

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Associate Professor Carol L. Keefer, Chair
Professor Ian Mather
Professor Mary Ann Ottinger
Associate Professor Iqbal Hamza
Associate Professor Eric Haag

# Dedication

To Emily- my wife, best friend, and former Biometrics 601 tutor:

Thank you for your unconditional love and support.

To my parents, Paul and Marilyn:

Thank you for providing me the nature and nurture that allowed me to achieve what I have achieved.

# Acknowledgments

First and foremost, I wish to thank my Mentor, Dr. Carol Keefer, for this invaluable experience. I sincerely appreciate Dr. Keefer's constant guidance, direction, patience and understanding through both the good and the bad times of my research and personal life. I am also thankful for Dr. Keefer's continual support and confidence in my abilities as well as imparting to me a model philosophy for conducting research, teaching, and mentoring others.

I wish to express deep appreciation to my Ph.D. committee, Drs. Ian Mather, Mary Ann Ottinger, Iqbal Hamza, and Eric Haag, for their unwavering dedication to develop and improve the scientist within me. I am also very thankful for their "Open Door policies" for both advice and equipment.

I wish to thank members of the Keefer laboratory, most notably Lei Li, Disha Pant, and Shuyang He, for their assistance and wisdom over the years. I also have great appreciation for my non-laboratory peers, Dr. Scott Severence, Laura Ellestad, and Jason Sinclair. They were my hallway consultants as well as my friends.

I am indebted to the Department staff for all of their help and time: Janice Barber, Clare Capotosto, Sheryl Grey, Gary Lapanne, Kim Montague-Smith, Sandra Nola, and (last but not least) Edith Silvious.

This project could not have been possible without the services and guidance of Dr. Neil Talbot, the creator of the CT-1 cell line, and Dr. Ronald Zuckermann, whom generously provided me with many Lipitoid samples.

Many thanks to my Mother, Father, brothers Paul and Todd, sister Karen, and all my nieces and nephews for always being there and providing such a strong foundation of love and support. Many thanks to my new family, Susan and Charles Devillier, for all of their love and support as well.

Finally, I wish to express deep gratitude to my loving wife Emily. It's a wonder that so much grew from our friendship in Biometrics 601 back in 2004. Thank you so much.

# Table of Contents

# List of Tables

# List of Figures

**Chapter 1: Literature Review** **Page**



**Chapter 2.**

**Chapter 3.**



**Chapter 4.**

**Chapter 5: Future directions.**



**Appendix:**

# List of Abbreviations

| | |
|---|---|
| ACTG1 | Gamma actin 1 |
| ANOVA | Analysis of variance |
| BCA | Bicinchoninic acid |
| BFF | Bovine fetal fibroblast |
| bHLH | Basic Helix-Loop-Helix |
| BT-1 | Bovine trophoblast cell line (Ushizawa) |
| CDX2 | Caudal type homeobox transcription factor |
| Ct | Cycle Threshold |
| CT-1 | Cow trophoblast cell line 1 (Talbot) |
| DLX3 | Distal-less homeobox 3 transcription factor |
| DMEM | Dulbeccco's Modified Eagle Medium |
| DMSO | Dimethyl Sulphoxide |
| DPBS | Dulbecco's phosphate buffered saline |
| dpc | Days post conception |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetate acid |
| EF | Elongation factor |
| ELF5 | E74-like factor 5 (ETS domain transcription factor) |
| EOMES | Eomesodermin transcription factor |
| EPC | Ectoplacental cone |
| ERRB | Estrogen related receptor beta |
| ERT | Endonuclease- reverse transcriptase |
| ESC | Embryonic stem cell |
| ETS2 | V-ETS erythroblastosis virus E26 oncogene homolog 2 |
| EXE | Extra-embryonic ectoderm |
| FBS | Fetal bovine serum |
| FGF | Fibroblast growth factor |
| FGFR2 | Fibroblast growth factor receptor 2 |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GATA | GATA binding protein transcription factor |
| GFP | Green fluorescent protein |
| HA | Hemagglutinin epitope |
| HAND1 | Heart and neural crest cell derivative 1 transcription factor |
| HEK 293 | Human embryonic kidney 293 cells |
| H2B-GFP | Histone 2B fused to GFP |
| HT1080 | Human sarcoma cell line |
| ICC | Immunocytochemistry |
| ICM | Inner cell mass |

| | |
|---|---|
| ID2 | Inhibitor of DNA binding 2 transcription factor |
| IFNt | Interferon tau |
| IVP | In vitro-produced |
| LEF1 | Lymphoid enhancer-binding factor 1 |
| LINE | Long interspersed nuclear element |
| LTR | Long terminal repeat |
| LV | Lentivirus |
| MASH2 | Achaete-scute complex-like protein 2 transcription factor |
| MEF | Mouse embryonic fibroblast |
| MOI | Multiplicity of infection |
| MYC | Myelocytomatosis oncogene epitope |
| nGFP | Nuclear localized green fluorescent protein |
| NTERA | Human embryonal carcinoma cell line |
| ORF | Open reading frame |
| OCT4 | Octamer binding protein 4 |
| PAG | Pregnancy associated glycoprotein |
| PCV | Packed cell volume |
| PMSF | Phenylmethanesulfonylfluoride |
| PS | Penicillin-Streptomycin |
| RACE | Rapid amplification of cDNA ends |
| RE | Retroelement |
| RFP | Red fluorescent protein |
| SALL4 | Sal-like 4 transcription factor |
| SCNT | Somatic cell nuclear transfer |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SINE | Short interspersed nuclear element |
| SNP | Single nucleotide polymorphism |
| SOLD1 | Bovine secreted protein of Ly-6 domain 1 |
| SOX | SRY (sex determining region Y)-box transcription factor |
| SSLP1 | Secreted seminal vesicle Ly-6 protein 1 |
| STO | Sandos inbred mice 6-Thioguanine and Ouabain resistant |
| TBST | Tris buffered saline with Tween 20 |
| tcRNA | Total cellular RNA |
| TE | Trophectoderm |
| TEAD4 | TEA domain family member 4 transcription factor |
| TSC | Trophoblast stem cell |
| WCL | Whole cell lysate |
| WNT3a | Wingless-type 3A |
| XEN | Extra-embryonic endoderm stem cell |
| ZO-1 | Zona occludens -1 |

# Chapter 1: Literature Review

During the twentieth century, advances in assisted reproductive technologies (ART) progressed rapidly from cryopreservation of gametes and artificial insemination to in vitro embryo production (IVP), transgenic animal production and cloning. Notable achievements in cloning technologies of domestic animals include the demonstration of somatic cell nuclear transfer (SCNT) in sheep (Wilmut et al., 1997), creation of transgenic goats for biopharming (Keefer 2004), and production of transgenic disease resistant cattle for agriculture (Wall et al, 2005). Yet today, these ART methods remain substantially cost inefficient for producing large populations of viable transgenic offspring. This dilemma stems from the suboptimal conditions of the current ART protocols themselves, especially within IVP and SCNT generation of ungulate embryos. Multiple studies have demonstrated that IVP embryos transferred back into surrogate recipients incur higher rates of embryonic death than artificial insemination or transferred in vivo-produced embryos (Farin et al., 2006). Another study found that 82% of SCNT-derived embryo transfers resulted in fetal death between days 30-90 of gestation with many of the cloned fetuses exhibiting poor placental development (Hill et al., 2000). Placental abnormalities observed in SCNT-produced ruminant conceptuses include decreased number of placentomes, avascularization or hypovascularization, hydorallantois, enlarged placentomes, and overall increase in fetal membrane mass. When SCNT and IVP neonates do develop to term, they can exhibit a wide variety of congenital defects commonly referred to as 'large offspring syndrome' or 'abnormal offspring syndrome' (Farin et al., 2006).

The compromised embryonic developmental potential of SCNT-derived embryos has been suggested to result from incomplete reprogramming leading to changes in gene expression. This association has been verified in studies comparing gene expression profiles between IVP, SCNT or in vivo-produced bovine embryos (Aston et al., 2010; Daniels et al., 2001; de A Camargo et al., 2005; Hall et al., 2005; Jang et al., 2005, Rodríguez-Alvarez et al., 2010, Wrenzycki et al., 2004). SCNT embryos exhibited abnormal expression of trophectoderm (TE) or preplacental lineage-specific genes, including *CDX2, ERR2, IFNT* and *EOMES* (Rodríguez-Alvarez et al., 2010; Hall et al., 2005). Abnormal expression levels of additional TE lineage-specific factors *MASH2* and *HAND1* were detected as early as 17 days post conception (17 dpc) and remained significantly elevated in developing cotyledons at 40 dpc. The elevated levels of *MASH2* expression did not result from failure to imprint *MASH2*, as *MASH2* was expressed from both alleles at 17 dpc and expressed from a single allele at 40 dpc as observed in in vivo-derived embryos. Therefore, imprinting of *MASH2* in SCNT-derived embryos occurred, and the elevated *MASH2* transcript levels were expressed from a single allele. These findings were correlated with a reduced number of binucleate cells in cotyledons at 40 dpc, suggesting that early misexpression of trophoblast regulators later contributes to detrimental alterations in placental formation (Arnold et al., 2006).

Even though aberrant gene expression during early embryonic development has been identified as a contributing causal link to the gestational losses following ART, the primary developmental mechanisms regulating gene expression during early gestation remain poorly understood. The challenge is to identify the key developmental regulators and their roles during preplacental tissue lineage differentiation. This would allow

accurate interpretation of the abnormal TE gene expression patterns that are associated with subsequent placental defects.

**Regulative cues in mammalian development**

The period of preimplantation encompasses a series of developmental stages that transform a single-cell zygote into a cavitated embryo called a blastocyst. Even though all placental mammals progress through these stages, there are many inherent species differences in the biological details comprising each stage. Much of what is known about this period of preimplantation stems from mouse model-based research, and this foundation of knowledge provides insight into development of other species. The stages of preimplantation will be reviewed here in the context of the mouse model and our primary focus, the bovine model.

It is widely recognized that mammalian preimplantation development is predominantly regulative, in that an embryonic cell's fate is constantly influenced by external cues from neighboring cells and the environment. After fertilization the mammalian zygote undergoes a series of cleavages to produce an aggregation of cells or blastomeres called the morula (Fig. 1.1). Dividing mouse blastomeres have very short G2 and M phases, resulting in static net embryonic growth and a serial decrease in the size of individual blastomeres. It is widely accepted by many laboratories that the first few cleavages are not synchronous within all blastomeres and occur in random planes so that daughter cells are identical with regard to developmental potential (Hiiragi T. et al., 2006).

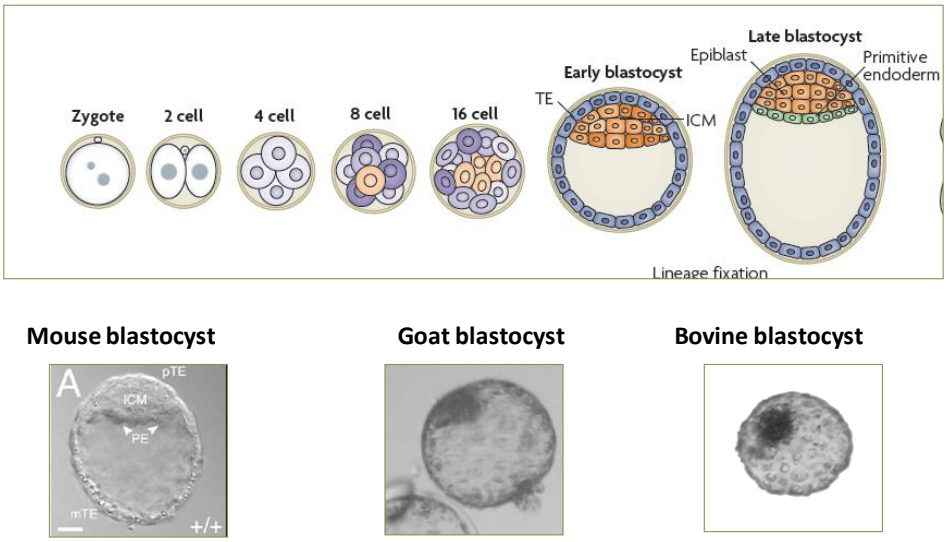



**Figure 1.1**. Schematic representation of placental embryogenesis from the zygote to the blastocyst stage (cartoon from Hemberger et al., 2009). The DIC images show that the blastocyst embryonic structure is conserved between mice and ruminants (images from Degrelle et al., 2005 and Strump et al., 2005).

Other laboratories believe this solely regulative model may be too simplistic based on their findings showing nonrandom tendencies or biases occurring during the initial cleavage stages (Plusa et al., 2005a; Zernicka-Goetz, 2006). Their results provide evidence that asynchronous blastomere divisions tend to follow a pattern. At the two-cell stage, the “leading blastomere” is the first cell to divide. It tends to inherit less volume and undergoes a cascade of divisions to predominantly populate the embryonic pole of the blastocyst, which includes the inner cell mass (ICM). The remaining blastomere of the two-cell stage, called the lagging blastomere, has been observed to be larger and primarily gives rise to the mural TE lineage.

While the presence of a bias during the first few cleavage stages remains under debate, another hypothesis is proposed. According to the "Inside-Outside hypothesis," the specification of the first two cell lineages (ICM and TE) is thought to be determined by cellular positioning within the morula. In context to position, “outer cells” will tend to allocate to the TE lineage, while “inner” blastomeres, having more intercellular contact, often allocate to the ICM line (Tarkowski et al, 1967; Betteridge and Flechon, 1988). The most recent test of this hypothesis involved disaggregation and then random re-aggregation and intermingling of cells originally positioned in the inside or outside of morulae. While many blastomeres sorted and migrated back to their original position, a few cells were observed to ‘reprogram’ based on their new position. (Suwińska et al., 2008). Thus, lineage allocation and formation of the embryonic-abembryonic axis formation may be affected by intrinsic biases and be influenced by external cues.

**Overview of preimplantation mammalian development**

Upon successful conversion to an independent functioning embryonic transcriptome, embryos continue through additional cleavage steps to the first morphogenetic transition called compaction (Johnson and McConnell, 2004). In the mouse, eight blastomeres compose the early morula. Each blastomere is roughly spherical, symmetrical, and equally covered with microvilli over their cortex. By the end of the eight-cell stage, blastomeres undergo apical polarization and flattening to achieve maximal cellular contact between cells axially from the morula center to the surface. Proteins that are involved in establishing cell polarity, such the mouse homologues of the C. elegans PAR (partitioning defective) genes and atypical protein kinase C, are asymmetrically localized during compaction (Plusa et al., 2005b; Vinot et al. 2005). Morphological polarization is also accompanied by cytocortical, cytoskeletal, and cytoplasmic polarizations that contribute to eventual epithelialization of the outer cells. E-cadherin and β-catenin become restricted to basolateral membranes within the polar cells. Zona occludens 1 (ZO-1) protein, a component of apicolateral tight junction complexes, is detected at the 8-cell embryo during compaction and plays a role in the morula-blastocyst transitional process (Wang H. et al., 2008). In bovine embryos, ZO-1 was detected later at the morula stage, but in a similar manner to mice, it was fully localized to the apical membranes of TE cells by the blastocyst stage (Barcroft et al., 1998). Microvilli shift from the basolateral membrane and become restricted to the apical surface membranes. Upon maximized basolateral cell contact, there is evidence of gap, tight and intracellular adherens junctions. In the bovine model, compaction occurs two cleavage stages later in the 32-blastomere morula and is thought to occur by similar

mechanisms as in the mouse (Betteridge and Flechon, 1988).

Following compaction in mouse embryos, cell polarity influences allocation of cells into the ICM or TE tissues (Johnson and Ziomek, 1981). Apolar inner morula cells originate from asymmetric divisions during the fourth and fifth cleavage cycles and are destined to populate the ICM. These cells are devoid of microvilli, are covered in adherens junctions, and serve as an embryonic progenitor population for the epiblast (embryo proper) and hypoblast (primitive endoderm) lineages. Alternatively, the resulting surface-exposed, polarized blastomeres form an epithelial layer and are restricted to become the TE. Yet these cellular changes do not concretely dictate lineage allocation. Many studies have demonstrated that totipotency can still exist in polarized, post-compaction blastomeres from 8-cell embryos (Johnson and McConnell, 2004).

To determine the stage at which blastomeres lose totipotency, 16 and 32-cell morulae and early blastocysts were disaggregated and reaggregated such that cells sharing the same previous position and polarity were used to form new embryos (Suwińska et al., 2008). All 16-cell aggregates composed of only ‘outside’ blastomeres, ‘inside’ blastomeres, or a mixture of both developed into fertile mice. This provides strong evidence that mouse 16-cell blastomeres still maintain totipotency. Furthermore, they found that expression of *Cdx2*, a required transcription factor for the TE lineage, downregulated within 20 hours in ‘outside’ blastomeres converting to ‘inside’ blastomeres, and upregulated within 20 hours within “inside” blastomeres transitioning to “outside” cells. However, it was noted that the implantation rate for aggregates containing “inside only’ cells was less efficient, suggesting that these aggregates need to delay implantation in order to polarize their outer layer or shift lineage specification

programs. At the 32-cell aggregate stage, aggregates composed of "inside only" blastomeres formed blastocysts, but cavitation and blastocoel formation was compromised. None of these embryos developed to term. Aggregates composed of 'outside only' 32-cell stage blastomeres could develop into blastocysts with discernable ICMs only if the outside cells were originally disaggregated immediately following the fifth cleavage stage but before the cells initiated production of blastocoel fluid (Suwińska et al., 2008). These results suggest that totipotency within outside mouse blastomeres is lost when $Na^+/K^+$ ion pumps activate and blastocysts accumulate blastocoel fluid. In regard to bovine embryogenesis, it remains unclear if these blastomere aggregation-re-aggregation experiments would produce similar results as compaction does not occur until the 32-cell stage. The TE tissue lineage is fully formed upon epithelialization of the outer cells and formation of the blastocoel, which are dependent on proper ZO-1 and ZO-2 expression (Sheth et al., 2008; Wang H. et al, 2008) and zygotic expression of E-cadherin (Larue et al., 1994; Ohsugi et al., 1997). The requirement for zygotic E-cadherin expression at this stage is quite specific, as E-cadherin $^{-/-}$ embryos expressing "knocked in" N-cadherin could not rescue the E-cadherin $^{-/-}$ phenotype. E-cadherin $^{-/-}$ embryos can only rely on maternal E-cadherin for adhesion and signaling up until the morula stage (Kan et al., 2007).

Beyond the blastocyst stage, mouse embryo morphology and progenitor tissue lineages become quite unique in comparison with primates and ruminants. While the mouse ICM develops into the "egg cylinder", the ICMs of other species morph into germinal disks (Fig. 1.2 and 1.3).

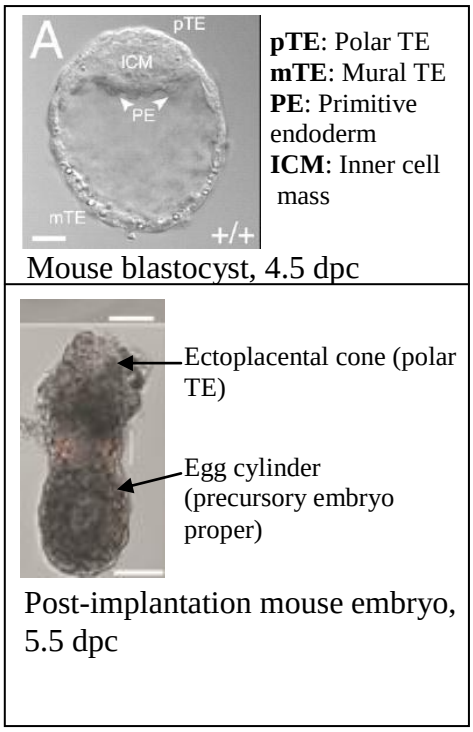



**Figure 1.2.** The top image is a DIC image of mouse blastocyst at 4.5 dpc. The scale bar is 20μm. The bottom image is a composite DIC confocal image of a post implantation mouse embryo at 5.5 dpc. The egg cylinder is the cup-like structure on the bottom that developed from the inner cell mass. The structures above the egg cylinder are the extra-embryonic ectoderm and the ectoplacental cone. The scale bar is 30 μm. (Images from Strumpf et al., 2005).

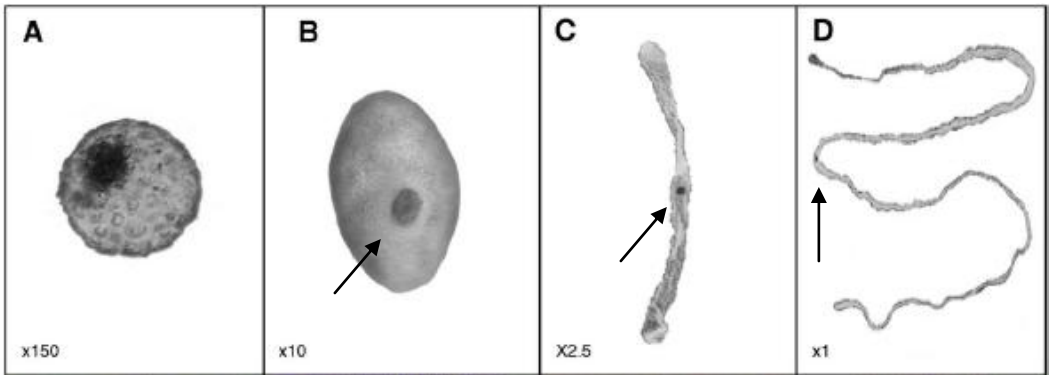



**Figure 1.3.** Pre-implantation bovine embryo morphology at different stages: **A**. Spheroid hatched blastocyst at 7.5 dpc. **B**. Ovoid blastocyst (1-12 mm) at 12 dpc. **C.** Tubular blastocyst (50-60 mm) at 15 dpc. **D.** Filamentous blastocyst (140-160 mm) at 17 dpc. Note the structure of the germinal disk (arrows) which develops into the embryo proper, as compared to the mouse egg cylinder in Fig. 1. (Degrelle et al., 2005).

The mouse TE progenitor population can be separated into polar and mural trophoblastic lineages. The polar trophoblast is restricted to the embryonic region that lies adjacent to the ICM. It contains the trophoblastic stem cell population which proliferates via FGF4/FGFR2-mediated paracrine signaling from the ICM (Chai, et al., 1998; Tanaka, et al., 1998; Haffner-Krausz et al., 1999; Nichols et al., 1998). Proliferation of these cells results in lateral expansion downward to populate the mural TE (Gardner, 2000). The polar TE also gives rise to the extra-embryonic ectoderm (EXE) lineage and formation of the ectoplacental cone (EPC). Both the EXE and the EPC contribute to differentiated placental tissue lineages. The abembryonic mural trophoblast differentiates into primary giant cells, which are essential for implantation and formation of the yolk sac (Armant, 2005).

The mouse blastocyst hatches from the zona pellucida around 3.5 dpc and rapidly implants into the uterine lining around 4.5 dpc. In contrast, bovine embryos hatch around 8.5 dpc and exhibit an extended pre-attachment period until about 21 dpc during which they undergo extensive elongation. Expanded bovine blastocysts initiate alterations in tissue morphology and expand into an ovoid structure by 12 dpc. Around this time, the bovine polar trophoblast, or "Rauber's layer," slowly disappears while the underlying ICM surfaces as the germinal disk (Viebahn, 1999). Unlike the mouse polar TE, the Rauber's layer does not maintain a trophoblast stem cell population which is responsible for TE proliferation, nor does it differentiate into EXE or EPC lineages. In contrast, the mural trophoblast in ruminants is responsible for the considerable embryonic elongation (Betteridge and Flechon, 1988). Before the ovoid embryo has further elongated into a filamentous type structure by dpc 15, it is composed of non-differentiated, IFN-t-

secreting mononucleate cells. After dpc 15, a subset of the mononucleate population differentiate into binucleated cells, until approximately 20% of placental cells are binucleated. These binucleated cells are homologous to mouse primary giant cells and undergo endoreduplication to become multinucleated. They can secrete a multitude of proteins and hormones such as placental lactogen. Around 21 dpc, the bovine embryo has grown to about 30 cm in length and is ready for uterine attachment. Binucleate ruminant cells cross the placentome interface and invade the maternal caruncular epithelia to fuse with maternal cells and form syncytia.

The regulatory mechanisms responsible for directing elongation and attachment of the bovine TE are not well known (see review, Blomberg et al. 2008). In vivo studies in adult ewes lacking proper endometrial gland density show that survival and proper development at the tubular stage and beyond is dependent on endometrial secretions (Gray et al., 2001; Gray et al., 2002). Whether mononucleate cell proliferation and increasing protein synthesis are dependent on maternal factors during the initial period of elongation is not known. Unlike the mouse polar trophoblast, bovine trophoblast proliferation does not appear to be driven by FGF-mediated signaling, although it has been shown that uterine FGF-2-mediated signaling stimulates IFNT production (Michael et al., 2006) and a correlation between uterine IFNT concentration and embryo size was observed (Robinson et al., 2006).

**Key transcription factors and their roles in lineage segregation**

Assessment of protein levels of a few key transcription factors is often regarded as sufficient for designating embryonic cells as committed to TE or ICM lineages. For example, mouse blastomeres are considered committed to an ICM fate if they highly express the transcription factors OCT4 (Rosner et al., 1990, Scholer et al., 1990) and NANOG (Chambers et al., 2003; Mitsui et al., 2003), but express low levels of CDX2 (Strumpf et al., 2005). Designation to TE fate is appointed when the patterns of "lineage selector" expression of OCT4, NANOG and CDX2 are reversed. Expression of mouse OCT4, a POU domain transcription factor encoded by the *Pou5f1* gene, is detected in the nuclei of oocytes and throughout all stages of early cleavage, indicating that early OCT4 expression originates from maternal deposits but is later produced via zygotic gene expression. In early mouse development, OCT4 is expressed in both ICM and TE cells but is markedly downregulated in the TE lineage and restricted to the ICM during blastocyst formation (Dietrich and Hiiragi, 2007). In subsequent differentiation of the ICM, OCT4 expression was upregulated in the primitive endoderm while being maintained at constant levels in the epiblast (Palmieri et al, 1994). *Oct4* $^{-/-}$ embryos fail to maintain a pluripotent ICM and loss of OCT4 induction of FGF4 in the ICM results in the loss of polar TE proliferation (Nichols et al., 1998). Discovery of OCT4 expression in undifferentiated embryonic stem cells (ESC), germ cells, and mouse and human embryonic carcinoma (EC) lines also distinguished OCT4 as a major regulator involved in the maintenance of pluripotency and self renewal (Okamoto et al 1990; Scholer et al., 1989; Wang L. et al, 1996). As a result, OCT4 expression was considered a “molecular lock” against differentiation into trophoblast lineages. Later evidence indicated that

OCT4 expression acts more like a "gatekeeper" to many paths of lineage differentiation (Pesce and Scholer, 2001).

Maintenance of pluripotency in the developing embryo also requires the activity of the transcription factor SOX2 (SRY-related HMG box) which contains a high mobility group (HMG) box DNA binding domain. SOX2 is thought to have multiple roles in regulating different embryonic lineages. In the ICM, SOX2 partners with OCT4 to form a heterodimer that induces FGF4 expression that is required for EXE lineage maintenance (Ambrosetti et al., 1997 and 2000). Mouse *Sox2* transcripts are first detected at the morula stage and expression is maintained in the inner cell mass and well into the epiblast stage. Expression is later restricted to neuroectoderm cells and also becomes reactivated in the EXE. SOX2 protein expression overlaps with OCT4 in embryos and germ cells. However, *Sox2* knockout $^{-/-}$ embryos survive longer than *Oct4* $^{-/-}$ embryos and manage to develop into the peri-implantation stage due to persisting maternally expressed SOX2 protein (Avilion et al., 2003). *Sox2* $^{-/-}$ embryos that implanted lacked a developed egg cylinder or epiblast but possessed extraembryonic endoderm and trophoblast giant cells (TGCs). Similarly, explanted *Sox2* $^{-/-}$ ICMs differentiated into TGCs and differentiation could not be inhibited by FGF4 stimulation as observed in trophoblast stem cells (Avilion et al., 2003; Tanaka et al., 1998). Doxycycline-inducible *Sox2* $^{-/-}$ mouse ESC also differentiate into trophectoderm-like cells (Misui et al., 2007). These findings demonstrate the requirement for SOX2 in restricting differentiation in the ICM.

NANOG was also discovered to play a pivotal role in maintaining pluripotency in mouse ESC and ICM (Chambers et al. 2003, Mitsui et al. 2003). NANOG is a unique homeoprotein transcription factor that is only 50% similar in amino acid sequence to

homeodomains of members of the NK2 transcription factor family and shares no other conserved motifs. Mouse *Nanog* mRNA expression is first detected in the compact morula, maintained in the ICM, and rapidly down-regulated in the TE (Chambers et al., 2003). In expanded mouse blastocysts, *Nanog* mRNA expression is restricted to the epiblast and expressed well into post implantation at day 7.5 dpc (Hatano et al., 2005). Mouse NANOG protein expression closely patterns mRNA expression and is localized to the nuclei. *Nanog* $^{-/-}$ knockout mice are indistinguishable from wildtype around the time of ICM/ TE formation 3.5 dpc, suggesting that NANOG is not absolutely essential for initial mouse blastocyst formation. *Nanog* $^{-/-}$ embryos observed at 5.5 dpc have abnormal development of extraembryonic tissues, and the epiblast is undetectable. Heterozygous mice, however, are viable and indistinguishable from wildtype (Mitsui et al 2003). Thus, NANOG plays a definitive role in the second "embryonic cell fate specification" when the ICM becomes epiblast and hypoblast (Cavaleri and Scholer, 2003).

SALL4, a zinc finger transcription factor of the *spalt*-like family, is the most recently discovered key regulator of pluripotency. During mouse embryogenesis *Sall4* expression is similar to *Oct4* expression, starting with abundant maternal transcripts within the single-cell zygote. Following successive cleavages zygotic expression of SALL4 soon becomes restricted to ICM blastomeres (Hamatani et al., 2004) and is later restricted to the primitive endoderm (Zhang et al., 2006). When functional *Sall4* siRNA was injected into single-cell embryos, significant reductions not only in Sall4 levels but in Oct4 expression were observed. In addition, the downregulated SALL4 embryos mimicked *Oct4* null $^{-/-}$ embryos which exhibit loss of OCT4 expression and ectopic CDX2 expression in the ICM (Zhang et al., 2006). These results prove that SALL4 is a

major regulator of *Oct4* during early embryogenesis. Using ESC as a genetic model, the importance of SALL4's role in maintaining pluripotency has been further elucidated (Lim et al., 2008; Wu et al., 2006; Yang et al., 2008).

It was originally thought that TE lineage formation occurred by default, primarily via down-regulation of OCT4, SOX2 and NANOG expression. Without this triumvirate of transcription factors, blastomeres become alternatively specified and eventually differentiate into the TE lineage. In 2005, however, it was discovered that TE formation is dependent on expression of CDX2, a homeodomain protein homologous to *D. melanogaster caudal* (Strumpf et al, 2005). During preimplantation, *Cdx2* mRNA is zygotically expressed with protein readily detected by the eight-cell stage (Ralston and Rossant, 2008). A cross section pixel analysis of relative CDX2 antibody-labeled nuclear protein levels was performed in z stacks of confocal images spanning entire mouse morula. Relative to ubiquitous levels in compacted 8-cell blastomeres, a threefold difference was detected in the outer cells of the 16-cell morula, while a twofold amount was observed in the inner cell population. In 32-cell morulae, the difference was 4-fold greater in outer cells (Ralston and Rossant, 2008). *Cdx2* $^{-/-}$ knockout embryos developed immature collapsing blastocoels, and by the late blastocyst stage, epithelial integrity was compromised through mislocalization of E-cadherin and other tight junction proteins (Strumpf et al, 2005). Further compromise of TE lineage maintenance in *Cdx2* $^{-/-}$ embryos occurred as trophoblastic markers were downregulated and trophoblast giant cells never formed. Interestingly, OCT4 and NANOG were ectopically expressed in *Cdx2* $^{-/-}$ TE cells, indicating for the first time the presence of an in vivo inhibitory feedback interaction between OCT4, NANOG and CDX2 (Strumpf et al, 2005).

Similar to the role SALL4 in the induction of zygotically expressed OCT4, the TEA DNA binding domain containing transcription factor TEAD4 is responsible for *Cdx2* induction at the 8-cell stage (Nishioka et al., 2008; Yagi et al., 2007). *Tead4* $^{-/-}$ mutants displayed normal E-cadherin mediated cell polarity in the outside cells but these cells did not express any TE lineage regulators including CDX2 and EOMES. In addition, blastocoels failed to form and OCT4 and NANOG were expressed in the outer cells. Since TEAD4 expression is detected earlier in the 4-cell stage following zygotic genome activation at the 2-cell stage, this transcription factor is the earliest expressing key regulator of the TE lineage specification. The biochemical mechanism of TEAD4 mediated *Cdx2* induction remains unknown, as enhancer elements within the *Cdx2* promoter remain to be identified.

**Key regulators in transcriptional networks**

The transcriptional regulatory network that maintains pluripotency in ESC and ICM is managed by key regulators OCT4, SOX2, NANOG and SALL4. These transcription factors serve as the nucleus for multiple "network motifs" (Lee et al., 2002) that facilitate expression programs directing pluripotency and self-renewal. OCT4 and SOX2 commonly heterodimerize on adjacent cis elements to co-function as one master regulator (Remenyi et al., 2003) and together they regulate specific expression of FGF4 (Ambrosetti et al., 1997 and 2000), Osteopontin (OPN1, Botquin et al., 1998) UTF1 (Nishimoto et al., 1999) FBX15 (Tokuzawa Y et al., 2003) and NANOG (Kuroda et al., 2005; Rodda et al., 2005). NANOG forms a heterodimer with SALL4 and functions in a similar manner (Wu et al., 2006). The OCT4-SOX2 and SALL4-NANOG heterodimers

also autoregulate each of the individual monomers (Okumura-Nakanishi et al., 2005; Catena et al., 2004; Chew JL et al., 2005; Lim et al., 2008). Furthermore, SOX2 positively regulates *Oct4* by repressing the *Oct4* negative regulator NR2F2 (nuclear receptor 2f2, or COUPTF II) and activating the *Oct4* positive regulator NR5A2 (nuclear receptor 5a2, or LRH-1) (Masui et al., 2007).

With OCT4 and SOX2 cis activating *NANOG*, a feedforward loop is assembled that has been shown to regulate over 353 protein encoding genes and at least 2 microRNA genes in pluripotent human ESC. Chromatin-immunoprecipitation (ChIP) and DNA microarray analyses performed in human ESC not only confirmed the central role for OCT4 and SOX2, but also found that greater than 905 of bound *OCT4* and *SOX2* promoters were also bound by NANOG (Boyer et al., 2005). A similar genome-wide exploration in mouse ESC found a comparable number of promoter binding sites, and identified 32 protein coding genes that were also regulated in the human ESC study (Loh et al., 2006). Further analyses in mouse ESC confirmed that SALL4 co-occupies hundreds of gene promoters with OCT4, SOX2, and NANOG (Lim et al., 2008; Yang et al., 2008).

Research into these autoregulatory, inhibitory and feedforward loops provide clarification into how pluripotency is maintained in ESC and blastomeres. This interacting network provides increased temporal stability of *Oct4*/*Sox2*/*Nanog* gene expression, particularly in response to pro-differentiating external stimuli. Yet the feedforward loops must be regulated. Guangjin Pan and colleagues explored the interaction of NANOG and OCT4 in greater detail and found that OCT4 activated *Nanog* when expressed at or below steady state levels. When OCT4 was overexpressed, *Nanog*

promoter reporter activity was repressed. OCT4 at high levels also exhibited a negative feedback loop on itself (Pan et al., 2006). Together these findings contend that levels of OCT4 protein determine whether OCT4 participates in feedforward loop to maintain levels of itself, SOX2, and NANOG, or acts as the instigator to shut down the feedforward loop and allow differentiation to occur.

Reciprocal inhibition between NANOG and transcription factor GATA6 may be involved in the specification of ICM to either embryonic (epiblastic) or primitive endodermal (hypoblastic) fates. Evidence for this includes similarities in endodermal phenotypes shared by *Nanog* $^{-/-}$ mouse ESC (Mitsui et al. 2003) and mouse ESC overexpressing GATA6 (Fujikura et al., 2002). Interaction between these lineage selector transcription factors is evident in vivo, as immunocytochemical analysis of the ICM exhibits a "salt and pepper" effect in expressing either NANOG or GATA6 (Chazoud et al., 2006). Unlike the TE and ICM lineage divergence stage, segregation of epiblastic and hypoblastic lineages is not due to any positional effect. Rather, the fate of a cell to primarily express NANOG and become specified to the epiblastic lineage or express GATA6 and become specified to the hypoblast is dependent on the GRB-2/MAPK signaling cascade, possibly via FGF4. Cells expressing GATA6 in turn express GATA6-activated downstream adhesion molecules, thereby initiating physical sorting of blastomeres into two lineages with different adhesion properties. Cell sorting of GATA6 expressing cells from deep within the ICM to the surface is also facilitated by WNT9A signaling (Meilhac et al., 2009). NANOG is also involved in an indirect inhibitory feedback loop with T box transcription factor BRACHYURY in deciding mouse ESC fate to mesoderm (Susuki et al., 2006a). Transient imbalances between these regulators

results in a biphasic, reversible switch of identity between a pluripotent state and the specified state of a mesodermal progenitor cell (Susuki et al., 2006b).

The transcriptional regulatory mechanisms governing differentiation of the TE lineage involve a master regulatory inhibitory loop between CDX2 and OCT4. Interaction between these transcription factors was observed in mouse ESC, where knockdown of OCT4 and overexpression of CDX2 both induce mouse ESC differentiation into trophoblastic lineages (Niwa et al., 2005). Furthermore, knockdown of SALL4, the primary activator of *Oct4* expression, indirectly results in the increase of CDX2 and permits differentiation towards the trophoblast fate (Zhang et al., 2006). Biochemically, CDX2 and OCT4 proteins physically interact to form a complex that reciprocally represses transcriptional activities of both regulators. This mutual inhibitory loop involves both repression of transcription and repression of function by physical interaction, therefore resulting in loss of target transcription and autoregulation. CDX2 has been shown to noncompetitively bind to the *Oct4* autoregulatory element of the distal enhancer of *Oct4* and repress any *Oct4* autoregulatory stimulus (Niwa et al., 2005). This study was first to describe a model for the segregation of TE and ICM that depends on a critical balance of OCT4 and CDX2 (Fig. 1.4).

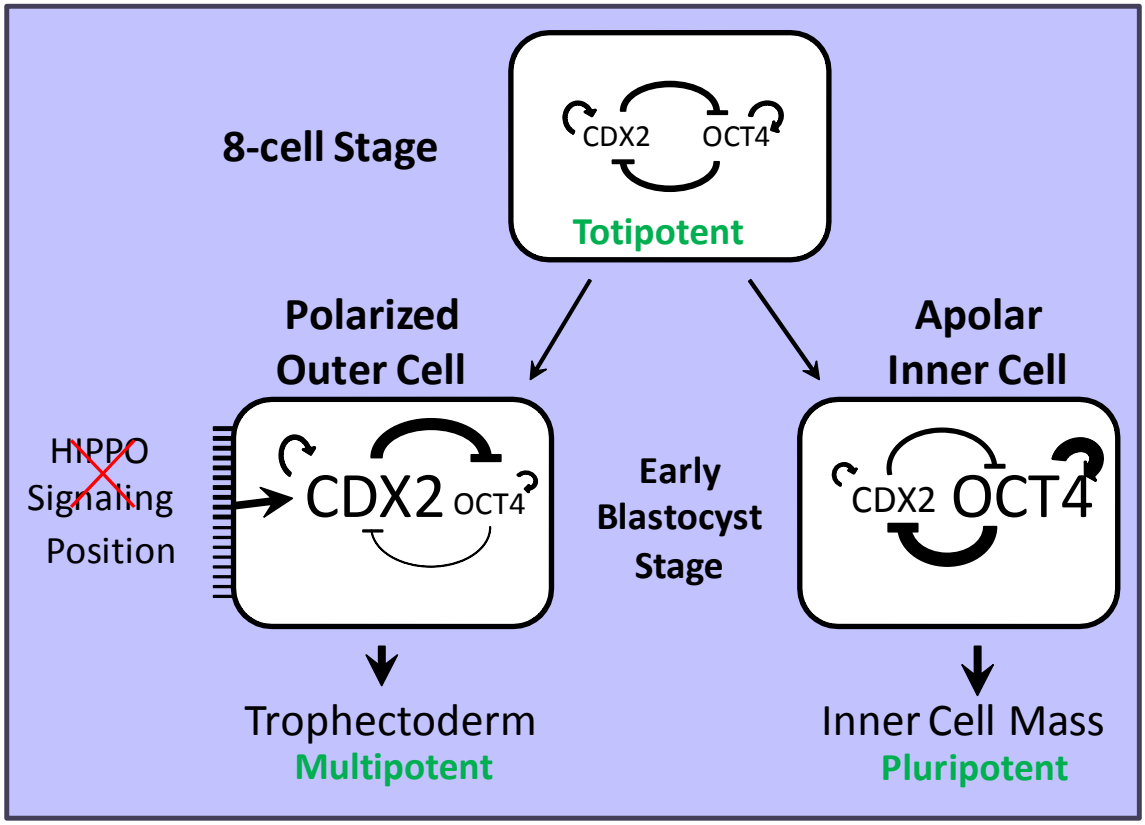
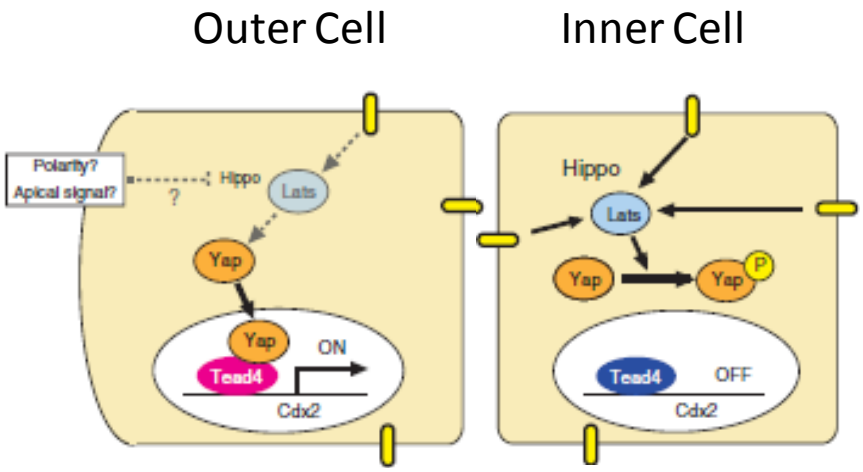



**Figure 1.4.** CDX2 and OCT4 participate in an inhibitory feedback loop that is responsible for ICM and TE lineage segregation in the mouse blastocyst. TEAD4 induction of *Cdx2* as well as other unknown positional cues cause CDX2 levels to increase and suppress OCT4 levels in polar outside cells, eventually leading to *Oct4* repression in the TE lineage. Inner apolar cells reciprocally show an imbalance in the CDX2/OCT4 inhibitory feedback loop in favor of OCT4, and CDX2 expression is eventually silenced (figure adapted from Niwa et al., 2005).

Recently, the external cue that tips the OCT4-CDX2 equilibrium in favor of CDX2 and specifies the outer cells towards the TE lineage was determined (Fig. 1.5A). HIPPO signal transduction occurs within the cells inheriting the inner position of the morula. This leads to phosphorylation of an essential TEAD4 co-activator, Yes–associated protein (YAP1), which is then consequentially shuttled to the cytoplasm where it remains nonfunctional. Without YAP1 present, TEAD4 is unable to induce high levels of CDX2 in the inner cells. Within outer cells, YAP1 is not phosphorylated and remains nuclear, which then permits TEAD4 induction of *Cdx2*. This mechanism allows CDX2 levels to overcome and eventually suppress OCT4 levels in the outer cells. In addition, this study also explains how cell position and polarity influence Cdx2 expression, as Hippo signaling is a conserved pathway that regulates cell-cell contact during proliferation. It is thought that maximized cell membrane contact shared between inner apolar cells, possibly facilitated by polarization of outer cells, initiates Hippo signal transduction (Nishioka et al., 2009; Sasaki et al., 2010).

Once CDX2 levels have reached a threshold whereby positive autoregulation occurs and OCT4 expression is suppressed and eventually silenced, TEAD4 and CDX2 initiate a feedforward loop to induce and maintain expression of EOMES and ELF5 (Hg et al., 2008; Niwa et al,. 2005; Sasaki et al., 2010). CDX2, EOMES and ELF5 expression are required for trophoblast stem cell maintenance and are regarded as “gatekeepers” of the mouse trophoblast lineage (Fig. 1.5B).

A.




**B.**

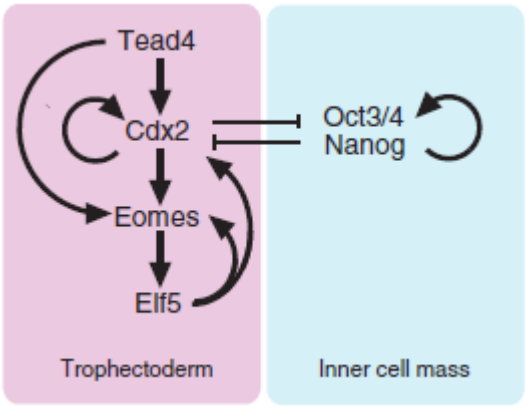



**Figure 1.5. A.** HIPPO signal transduction in inner, apolar cells results in YAP1 protein phosphorylation by LATS kinase. Phosphorylated TAP1 is restricted from the nucleus and prevents TEAD4-mediated induction of *Cdx2*. In outer polarized cells, HIPPO signal transduction is reduced and YAP1/TEAD4 increases *Cdx2* expression and subsequent negative feedback on *Oct4* expression (diagrams from Sasaki et al., 2010). **B**. CDX2, EOMES and ELF5 make up the core of the mouse trophoblast stem cell transcriptional hierarchy.

**Transcriptional networks in bovine trophectoderm**

Do OCT4 and CDX2 play roles in the central transcriptional architecture that specify, differentiate and maintain TE lineages in ungulates? Based on mRNA and protein expression profiles at different stages of domestic animal embryonic development, it seems likely that the transcriptional regulatory loops are involved in directing TE identity (Fig. 1.6). However, there are discrepancies between mouse and ungulate data. Pluripotency specific transcription factors OCT4 and NANOG are not entirely lineage restricted to the ICM at the expanded blastocyst stage. In contrast to mouse TE, OCT4 expression is not rapidly downregulated in the TE of bovine, porcine and caprine blastocysts (He et al., 2006; Keefer et al., 2007; Kirchhof et al., 2000; Pant et al., 2009; Van Eijk et al., 1999). Degrelle and colleagues demonstrated by in situ hybridization and RT-PCR that *OCT4* transcripts were readily detectable in the TE up to the ovoid stage of bovine preimplantation at 12 days post insemination (dpi) (Degrelle et al., 2005). However, in other studies using immunocytochemistry, OCT4 protein appeared to be restricted to the epiblast in 12 days post conception (dpc) hatched blastocysts and was undetectable in the TE (Vejlsted et al., 2005). Some of the discrepancies in timing of OCT4 downregulation observed in TE of ruminant species may be due to the different sources (in vitro vs. in vivo) and culture systems used.

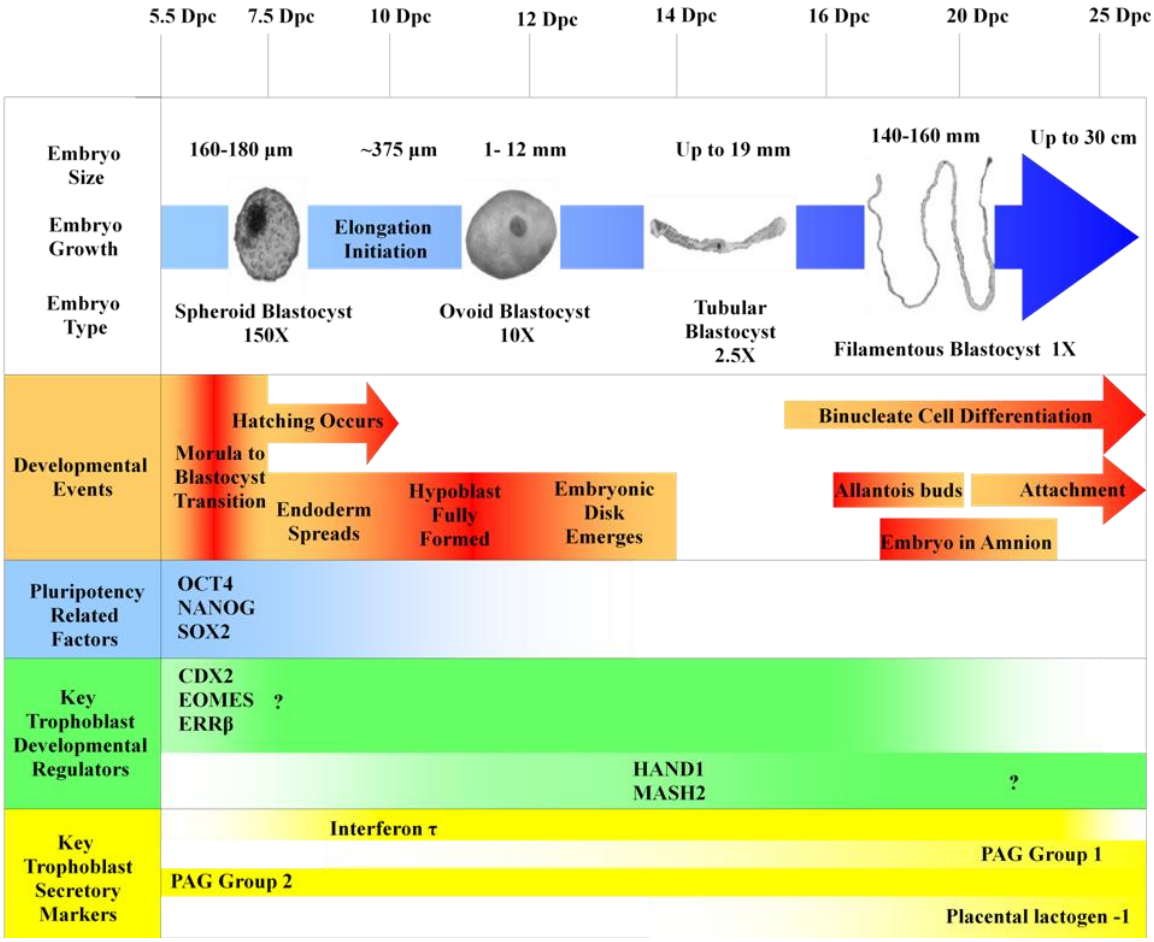



**Figure 1.6.** Schematic representation of early embryonic developmental regulators and markers in the bovine TE. (Images from Degrelle et al., 2005.)

While there are discrepancies as to the exact timing of OCT4 protein downregulation in the bovine TE, it is clear than OCT4 expression remains active much longer than that observed in the mouse. Thus, it is hypothesized that OCT4 may play an alternate role in the development of the elongating trophoblast that is separate from its role in the ICM.

To date, there are only a few published reports profiling NANOG protein expression in ruminants (He et al., 2006; Pant et al., 2009). *NANOG* mRNA expression in the caprine TE was detectable, although 54 fold lower than ICM expression levels (He et al., 2006). It was also shown that NANOG protein expression in bovine and caprine pre-implantation embryos is considerably different from the protein pattern observed in mice. Immunostaining with a validated anti-human NANOG antibody revealed protein localization in the nucleus from the 8 to 16-cell morula to the compacted morula stage. NANOG was also intensely localized to the nucleoli of TE cells, which was confirmed by positive co-localization with NUCLEOLIN, a nucleolar restricted protein. As blastocysts developed and reached their viability threshold of expansion under standard embryo culture conditions, NANOG protein expression tended to decrease in both nuclei and nucleolar ultrastructures. The mechanism for localization of NANOG to the nucleolus is unknown, but it is hypothesized that nucleolar localization correlates with downregulation of NANOG in TE specified cells. Therefore, this shift in localization of NANOG from the nucleus to the nucleolus may signify a posttranslational regulatory mechanism to decrease or inactivate NANOG function within those cells. This specific sequestering mechanism was previously found to be used in regulating nucleostemin

protein in stem cells (Tsai and McKay, 2002) and the transcription factor HAND1 in the rat choriocarcinoma (Rcho-1) trophoblast cell line (Martindill et al., 2007).

In the mouse, all three initial lineages of the preimplantation mouse blastocyst have been successfully derived into stem cell lines for use in studying regulatory networks (Ralston and Rossant, 2005). Pluripotent mouse embryonic stem cells (ESC), multipotent mouse trophoblast stem cells (TSCs) and endodermal stem cells (XEN) can be passaged indefinitely and be further manipulated to differentiate into cell types resembling those observed in vivo. There is no equivalent range of cell lines available for studying bovine development: validated bovine ESC lines have not been established for use as genetic models to investigate transcription factor regulatory networks (Keefer et al., 2007). However, endodermal and trophectodermal cell lines which are similar to XEN and TSC lines in mouse, have been successfully derived from bovine blastocysts. The Cow Trophectoderm-1 (CT-1) bovine TE cell line was derived by Neil Talbot and colleagues at the USDA from hatched day 10-11 bovine blastocyst explants, and then cultured continuously for over 2 years and 76 passages (Talbot et al., 2000). CT-1 cells grow as a monolayer on mitotically inactive mouse embryonic fibroblast feeder layers, polymerized collagen, Matrigel basement membrane matrix (Michael et al., 2006) and Corning Cellbind™ plastic (Schiffmacher and Keefer, 2008). Morphologically, they are epithelial and very similar in appearance to primary bovine TE outgrowths. At the EM level TE specific ultrastructures such as tight junctions, desmosomes, and microvilli were observed (Talbot et al., 2000). After reaching confluency, CT-1 cells mimic blastocoel expansion by budding off spheroid structures resulting from fluid influx underneath the monolayer. They are very robust in producing IFNT and have been previously used as an

in vitro model for understanding stimulatory signals of IFNT secretion (Michael et al, 2006).

Proteomic or transcriptional profiling of developmental regulators in CT-1 cells has been limited to mostly transcription factors that regulate IFNT expression including GATA2, GATA3, GATA6, DLX3, ETS2 and CDX2 (Ealy et al., 2009). A limited impression of the CT-1 mRNA transcriptome can be extrapolated from the gene expression profile of another bovine TE cell line called bovine trophectoderm-1 (BT-1) (Ushizawa et al, 2005.) This cell line was derived in a similar fashion to CT-1 cells (Shimada et al., 2001). The custom cDNA microarray used to characterize BT-1 cells was developed from utero-placental annotated genes and not bovine embryonic annotated genes, and thus fails to supply information on key transcriptional regulator expression. However, Ushizawa's data do indicate that BT-1 cells have similar gene expression patterns to the elongating pre-attachment bovine trophoblast. Furthermore, RT-PCR analysis verified positive expression of *OCT4* transcripts in BT-1 cells, which is similar to our findings in CT-1 cells (Schiffmacher and Keefer, 2008).

Expression of both *OCT4* and *CDX2* transcripts in CT-1 cells suggest this cell line is an appropriate model for elucidating their roles in directing bovine trophoblast maintenance and elongation in vivo. However, CT-1 cells are well known to be resistant to standard transfection techniques, unable to be passaged by single cell suspension and difficult to culture (Talbot et al., 2000; Ezashi et al., 2008). A popular experimental strategy for overcoming these limitations is to switch to human choriocarcinoma cell lines for experiments involving genetic manipulation. The human choriocarcinoma JEG-3 and JAR cell lines are amenable to trypsin-based passaging, are not difficult to transfect,

and are suitable as a bovine trophoblast molecular model when the human orthologs of interest are expressed (Ezashi et al., 2008). However, they would be unsuitable for RNAi studies or other experiments where manipulation of endogenous protein levels is required.

The major focus of this dissertation was to determine to what degree the CDX2/OCT4 inhibitory feedback loop within the TE transcriptional regulatory network maintains the early bovine TE in a proliferative, differentiation-delayed state. This was to be accomplished by testing the hypothesis that disruption of the OCT4/CDX2 equilibrium in CT-1 cells would significantly alter gene expression patterns of many TE lineage specific transcription factors, and alter the early bovine TE phenotype as well (Fig. 1.7). Specifically, we predicted a modest increase in OCT4 or decrease in CDX2 would result in significant downregulation of TE specific genes implicated in directing trophoblast elongation such as ETS2, MASH2, or ELF5. Alternatively, a modest increase in CDX2 or knockdown of OCT4 would have a reverse effect on target gene expression and also induce differentiation into later TE cell phenotypes such as binucleate cells. Results from this study would provide insight into the roles of OCT4 and CDX2 in the TE during bovine post-blastocyst development.

Human choriocarcinoma cell lines are an unsuitable model for addressing this aim. Therefore, our first objective was to establish a successful CT-1 transfection protocol as well as improve upon CT-1 cell passaging and culture techniques, as these prerequisites were paramount for utilizing CT-1 cells to address the main focus. These results are found in Chapter 1.

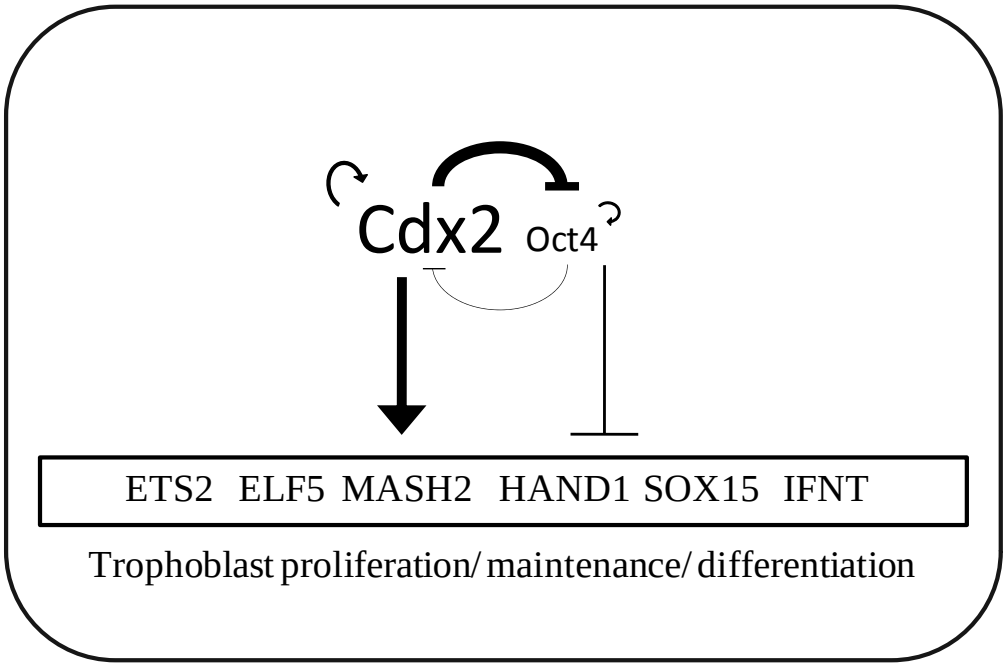



**Figure 1.7**. Model of how CDX2 and OCT4 regulate gene expression of common downstream trophoblast regulators in the early elongating bovine trophoblast. Unlike OCT4 restriction to the ICM in expanded mouse blastocysts, it is hypothesized that OCT4 negatively regulates expression of trophoblast regulators to prevent abnormal elevated expression and ensure proper proliferation and differentiation rates during this critical embryonic period.

Next, it was critical to further characterize the CT-1 cell line as a suitable model for bovine trophoblast development in which to study the roles of OCT4 and CDX2. The second objective was to expand on the expression patterns of candidate bovine trophoblast regulators. This assessment was done by testing the hypothesis that CT-1 cells possess a similar transcriptome of developmental regulator genes to the early TE lineage between the spheroid stage and the ovoid stage. Our confirmations of candidate gene expression, as well as the studies involving manipulation of *OCT4* and *CDX2* gene expression are found in Chapter 2. In this chapter it was important to fully characterize *OCT4* mRNA and protein expression as the recent discoveries of human OCT4 isoforms with different functions and confounding *OCT4* pseudogene expression have stressed the need for more thorough analysis of OCT4 in cell lines and tissues (Wang et al., 2010). Our findings of a transcribed *OCT4* retrocopy in CT-1 cells led to development of a third objective and chapter. We hypothesized that this transcribed *OCT4* retrocopy may produce a functional protein or RNA that either possesses a unique function or regulates "parental" OCT4 gene expression in the early bovine TE. Our characterization of this OCT4 retrocopy is found in Chapter 3.

# Chapter 2: Optimization of nucleic acid transfection methods for bovine trophectoderm CT-1 cells

**Abstract**

Studies investigating the molecular mechanisms of embryonic lineage segregation and maintenance rely heavily on the use of embryonic derived cell lines as models. While the bovine trophectoderm derived CT-1 cell line has been used mainly to study interferon tau (IFNT) gene regulation, it possesses potential as a model for mechanisms controlling trophoblast lineage differentiation and development. However, current use as a model is limited as CT-1 cells are recalcitrant to standard transfection methods. In addition, typical colony passaging methods are inefficient and involve physically shearing of colonies into large clumps. The focus of this study was to test several methods for nucleic acid delivery into CT-1 cells and improve upon current cell dispersal and culture techniques. Commercial cationic liposomal reagents, electroporation, biolistic bombardment, and lentiviral transduction all failed to successfully transfect plasmid DNA into CT-1 cells at an acceptable rate. However, optimization of Lipitoid-based transfection of plasmid DNA improved transfection efficiency to 9% across entire CT-1 colonies. It was also determined that successful siRNA-mediated gene knockdown can be achieved through optimization of Lipitoid-based oligonucleotide transfection. A successful enzymatic passaging protocol was also established using collagenase I and dispase that allowed for colony seeding with as little as 10 cells per seed aggregate. CT-1 cells were also successfully adapted to substrate-free culture for over 20 passages, eliminating the need to manually coat cultureware with substrate. Together these optimizations greatly

increase the utility of the CT-1 cell line as a bovine trophectoderm model amenable to genetic manipulation.

**Introduction**

While embryonic stem cell, trophoblastic stem cell and extraembryonic endoderm cell lines have proven to be valuable tools for studying developmental mechanisms in the mouse, the availability of similar cell lines in other species is much more limited. Two bovine cell lines, BT-1 (Shimada et al., 2002) and CT-1 (Talbot et al., 2000), have been utilized to study mechanisms that regulate gene expression of the bovine pregnancy recognition factor interferon tau (IFNT) (Michael et al, 2006a, 2006b; Ezashi et al., 2008; Bai et al., 2009; Sakurai et al., 2009), direct binucleate cell differentiation (Nakano et al., 2002, 2005), and characterize bovine trophoblast specific proteins (Walker et al., 2009). The CT-1 bovine trophectoderm cell line was derived following 2 years of continuous culture of cells obtained from hatched day 10-11 bovine blastocyst explants (Talbot et al., 2000). However, issues with CT-1 cell culture and standard transfection methods limit their functional capacity as a trophoblast model cell line. They are recalcitrant to standard transfection methods which are used in functional experiments including transcriptional reporter analysis, protein overexpression, and RNA interference assays. Only one study has reported successful knockdown of a gene in CT-1 cells. In that study the Lipitoid reagent (DMPE(NaeNmpeNmpe)$_3$) which had been demonstrated to deliver short interfering RNAs into hard-to-transfect IMR-90 primary cells (Utku et al., 2006) was used to reduce DLX3 protein expression in CT-1 cells (Ezashi et al., 2008). While Lipitoid has also proven to be successful for the transfection of plasmid DNA into other cell types (Huang et al., 1998), this reagent has not been used in any published studies to

deliver plasmid DNA into CT-1 cells. To date, there have not been any reports demonstrating success in creating a stable transgenic ruminant trophoblastic cell line in vitro. While cell lines can be derived from transgenic embryos produced by somatic cell nuclear transfer, derivation of transgenic lines by this method is costly and labor intensive (Talbot et al., 2007).

Another limitation of CT-1 cells is their inability to be dispersed into a single cell suspension. CT-1 cells do not survive enzymatic dispersal with trypsin (Talbot et al., 2000). Therefore, typical dispersal methods involve the physical shearing of colonies into large clumps or the collection of floating multicellular vesicles that can only propagate into one new colony per newly adhered clump or vesicle. Dispersal into single cells, or at least smaller clumps, is needed for selection of transgenic lines following in vitro transfection.

In this study, we attempted to resolve these challenges and improve upon transfection, cell dispersal, and culture methods. We were successful in adapting CT-1 culture to substrate-free Cellbind cultureware, and optimized an enzymatic passaging method using dispase and collagenase that increased colony dissociation and reduced clump size for new colony seeding. Our optimization efforts focused on increasing transfection efficiency of both plasmid DNA and short interfering oligonucleotides into CT-1 cells adapted to our feeder-cell free and substrate-free culture. While many conventional transfection techniques failed to increase transfection efficiency to significant levels, we found Lipitoid-based transfection to be the most powerful transfection reagent.

## Methods and Materials

### Cell Culture

CT-1 cells were received as a gift from Dr. Neil Talbot (U.S. Department of Agriculture, Beltsville, MD). Cells were initially expanded on mitotically inactive, gamma-irradiated mouse STO mouse embryonic fibroblast (MEF) feeder layers (CRL-1503, ATCC, Manassas, VA) plated into cell-culture-treated 60 mm plates or 25 ml cell-culture-treated flasks at a density of 40,000 cells per $cm^2$. CT-1 and STO MEF co-cultures were grown in CT-1 medium consisting of DMEM, containing 4.5g per L glucose and 4 mM L-glutamine (11965-084, Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS) (SH 30070.03; Hyclone, Logan, UT), 1 mM sodium pyruvate (11360-070, Invitrogen) and 50 U per ml penicillin-streptomycin solution (15070-063, Invitrogen). Cultures were incubated at 37 °C in 5% $CO_2$ for 4 to 5 days, in which CT-1 cells proliferated over feeder cells to form radial epithelial sheets. Colonies lying adjacent to each other grew until boundary edges touched, resulting in contact inhibition. Cultures were then passaged when either the majority of colonies were 4 to 5 mm in diameter or reached 50% confluency. Routine passaging consisted of mechanical dissociation by gentle scraping with a cell scraper. Cell clumps in suspension were then transferred to a 50 ml conical tube and the suspension volume was adjusted to 10 ml. Clumps were further dissociated by being twice drawn and expelled through a 22-gauge needle attached to a 20-ml syringe. The suspension was then diluted 3- to 5-fold and plated onto new gamma-irradiated STO MEF feeder layers or cryopreserved in cryopreservation medium (80% CT-1 media, 10% FBS and 10% DMSO). A bovine fetal fibroblast (BFF) cell line cloned from primary fetal hip muscle fibroblasts was also

received as a gift from Dr. Neil Talbot (U.S. Department of Agriculture, Beltsville, MD). The creation of this cell line is described elsewhere (Powell et al., 2004). BFFs and HEK-293 cells (CRL-1573, ATCC) were cultured in CT-1 medium at 37 °C in 5% $CO_2$. Upon reaching near confluency, cells were passaged by removing cell culture medium and adding trypsin-EDTA (25200-056, Invitrogen). After a five min incubation at 37°C, dissociated cells in suspension were collected, centrifuged at 175 x g for 3 min at room temperature (RT) in a table-top centrifuge. Supernatant was removed, fresh medium was added to cell pellets to a desired dilution and cells were dispersed.

**Enzymatic passaging of CT-1 cells**

All enzyme stock solutions were made fresh in DPBS and sterile filtered through a 0.2 µm SFCA syringe filter before use. CT-1 cells grown in penicillin-streptomycin (PS) free CT-1 medium were first washed with DPBS and incubated in 37 °C prewarmed dispase-collagenase solution (2.4 U per ml dispase (17105, Invitrogen) and 200 U per ml collagenase type 1A (C2574, Sigma Aldrich, St. Louis, MO) for 15 min at 37°C to detach colonies. The colony suspension was transferred to a 50 ml tube and centrifuged in a table-top centrifuge at 175 x g for 3 min at RT. The supernatant was aspirated and the pellet was resuspended in collagenase type 1A solution (200 U per ml collagenase in PS free CT-1 media). The suspension was incubated at 37°C for 1.5 hours and then lightly pipetted up and down with a 10 ml transfer pipet to complete colony dissociation. The cell suspension was centrifuged once again in a table-top centrifuge at 175 x g for 3 min at RT. Supernatant was removed and replaced with fresh CT-1 media.

**Adapting CT-1 cells to feeder-free and substrate-free culture**

CT-1 and STO MEF co-cultures were passaged at 1:2 and 1:3 into collagen coated cultureware prepared in the following manner: Collagen solution (pH 7.4; 40% PureCol collagen (5005, Advanced Biomatrix, San Diego, CA), 5% 10X DPBS (14200-075, Invitrogen), and 5% 0.1 M NaOH, and 50% CT-1 medium) was added to cultureware to coat the entire bottom and then aspirated to leave a thin collagen film. Coated flasks or plates were then incubated at 37 °C for one hour before air drying overnight at RT. Just before use, dried collagen coated wells were rinsed with CT-1 medium. CT-1 cells were considered successfully adapted and free of STO MEF contamination following a minimum of 8 passages. CT-1 cultures adapted to growing on a collagen matrix were then similarly passaged into untreated Cellbind cultureware (Corning, NY). CT-1 cultures growing on a collagen substrate or on Cellbind were maintained at 37°C and 8.5% $CO_2$ and subsequently passaged every 7 to 10 days.

**Lipofectamine 2000 transfection of adhered CT-1 cells**

DNA plasmids phEFnGFP (Wells et al, 2000) and H2B-GFP ("Addgene plasmid 11680", www.Addgene.org) were used for all plasmid DNA transfection optimization experiments. Both plasmids were previously tested in bovine fetal fibroblasts. The plasmid phEFnGFP constitutively produces high levels of nuclear localized GFP via induction of the human elongation factor (EF) promoter. Plasmid H2B-GFP constitutively produced human histone subunit 2B/ GFP fusion protein. The Invitrogen Lipofectamine 2000 protocol for optimizing plasmid DNA transfection was utilized as a starting point to which all subsequent modifications in the protocol were made. All dilution and plating volumes remained unchanged. CT-1 cells were grown in PS-free CT-

1 medium in collagen-coated 12-wells. HEK293 cells and BFFs, were also transfected with identical treatments to serve as high and moderate transfection level controls. Cells were washed twice with DPBS and once with OPTI-MEM prior to adding transfection cocktails. Cells were transfected with either phEFnGFP or H2B-GFP in ratios of 1:1, 1:2.5, and 1:3.75 (µg DNA/ µl Lipofectamine 2000). At each ratio, 1.6, 2.4, 2.8, 3.2, 3.6 and 4.8 µg plasmid were tested. Plasmid and Lipofectamine 2000 were diluted to final 100 µl volumes in OPTI-MEM before being mixed together and incubated for 10 min at RT. 200 µl DNA/Lipofectamine 2000 cocktails were then added to 12-wells containing 1 ml OPTI-MEM plating medium. Treated cells were then incubated overnight. Lipofectamine 2000 cocktails without DNA were also performed to assess Lipofectamine 2000 toxicity.

**Lipofectamine 2000 transfection of suspended CT-1 cells**

A 2- to 9-fold increase in Lipofectamine 2000 mediated transfection efficiency in difficult-to-transfect cell lines was achieved by transfecting suspended cells in undiluted transfection cocktail as opposed to following the standard protocol of transfecting adhered cells (Zhang et al, 2007). This method was tested on CT-1 cells. A fixed plasmid DNA/ Lipofectamine 2000 ratio of 1:2.5 was used for all treatments. 100 µl transfection cocktails containing 0.8 µg, 1.2 µg, and 4.8 µg of phEFnGFP plasmid and Lipofectamine 2000 were made according to the manufacturer's instructions. Each cocktail treatment was tested at different transfection incubation periods: 15 min, 60 min, 120 min, and overnight at 37°C. CT-1 cells grown on Cellbind in PS-free medium were dissociated using dispase-collagenase and centrifuged at 175 x g for 3 min in a table-top centrifuge at RT. Cells were then homogenously suspended and aliquoted into equal volumes per time

treatment. Aliquots were centrifuged once again to form individual pellets. 100 µl of transfection mixture was added to pellets and incubated at RT except for the overnight incubations, which were performed at 37°C. Following the incubation, 300 µl of PS free CT-1 medium was added and 400 µl cell suspensions were plated into Cellbind 6-well plates. Verification of transfection was determined 48 hours following passaging to 6-well plates.

**Lipofectamine 2000 and Nupherin transfection**

CT-1 cells were enzymatically passaged using the dispase-collagenase procedure and plated into collagen coated 24-well plates. Colonies were permitted to grow until over 50% of the well surface area was covered. Initial transfections were performed at DNA Lipofectamine 2000 ratios of 1:2 and 1:3.75 (µg/µl) with 1.2 µg plasmid H2B-GFP and with increasing levels of Nupherin reagent at 9, 30, 45, and 60 µg Nupherin per µg DNA (BML-SE225-0075, Enzo Life Sciences, Plymouth Meeting, PA). Lipofectamine 2000 transfection controls without Nupherin were also performed. Each transfection cocktail was assembled in the following manner: In tube A, plasmid DNA and Nupherin were added to OPTI-MEM to a total volume of 150 µl. In tube B, Lipofectamine 2000 was added to OPTI-MEM to a total volume of 150 µl. Both tubes were incubated at RT for 15 min, mixed together, and incubated at RT for an additional 40 min. CT-1 cells were then washed twice with DPBS, washed once with OPTI-MEM, and overlayed with the 300 µl cocktail. Twelve-well plates were then centrifuged at RT at 100 x g for 5 min before being placed into an incubator for 4 hours or overnight. Control treatments containing Lipofectamine 2000 and Nupherin but no DNA were performed to assess any potential reagent cytotoxicity. This protocol was tested a second time on new cells with a

few modifications. All transfection incubations were reduced to 4 hours, and all transfections were performed with fixed Nupherin concentrations at 30 µg Nupherin per µg DNA.

**Biolistic delivery of plasmid DNA into CT-1 cells**

Biolistic transformations were performed with a Model PDS-1000/He Biolistic Particle Delivery System with Hepta Adapter (Bio-Rad, Hercules, CA). CT-1 cells were grown to 70% confluency in collagen-coated 10 cm plates, Cellbind plates, or on 10 cm plates previously plated with a gamma- irradiated STO MEF feeder layer at a density of 50,000 cells per $cm^2$. Bovine fetal fibroblasts and HEK 293 cells were also plated and grown to 90% confluency to serve as comparisons for transfection rates. Plates containing CT-1 colonies grown on a feeder layer were overlayed with a thin layer of 0.5% agar. Sterile 0.5% agar was heated until completely liquid and allowed to cool to 37°C. Three ml of 0.5% agar was then pipetted on top of CT-1 and STO co-culture plates and quickly aspirated to produce a thin overlay. CT-1 medium was then added back to the plates. Gold microcarriers (1.6 µm diameter, 165-2262, Bio-Rad) were prepared according to Bio-Rad instructions and suspended in 50% sterile glycerol at 60 mg per ml. In a siliconized 1.5 ml tube, 50 µl of Afe I linearized, purified phEFnGFP plasmid (20 µg) was added to 100 µl (6 mg) of microcarriers and then vigorously vortexed for 1 min. Next, 150 µl 2.5 M $CaCl_2$ was added, and the suspension was then vortexed again for 1 min before adding 60 µl of 0.1 M spermidine. The suspension was then constantly vortexed for another 5 min before gently centrifuging the sample and removing the supernatant. The coated microcarrier pellet was then washed once with 300 µl 70% ethanol, vigorously vortexed, and centrifuged. The supernatant was removed once again,

and 170 µl of 100% ethanol was added. The suspension was then continuously vortexed for 10 min. During this step, the particle delivery system was sterilized and set up according to Bio-Rad instructions. An 1100 psi rupture disk (165-2329, Bio-Rad) was loaded into the rupture disk retaining cap. A mock 'test fire' was performed to verify disk rupture at around 1100 psi. A new 1100 psi rupture disk was reloaded and 25 µl of DNA-coated microcarriers were pipetted onto 7 separate sterile loaded macrocarriers (165-2335, Bio-Rad) previously loaded into the Hepta Adaptor. Once the assembly was complete, an uncovered cell culture plate was positioned at the 6 cm stage level and bombardment was performed. Cell culture plates were then immediately covered with CT-1 medium and placed back into the incubator.

**Electroporation of adhered CT-1 cells**

CT-1 cells were grown on gamma-irradiated STO MEFs in 6-well plates until colonies covered 75% of the feeder layer. HEK 293 cells were grown to near confluency and served as a transfection control. Electroporation was performed using a Petri Pulser electrode (PP35-2P) and ECM 2001 Electroporator (BTX, San Diego, CA) according to the manufacturer's instructions. All electroporation treatments were performed in duplicate. EcoR V linearized, purified phEFnGFP plasmid (10 µg) was diluted in 0.5 ml OPTI-MEM or DPBS per electroporation treatment tested. Prior to adding DNA to wells, cells were washed twice in OPTI-MEM and then overlayed with the 0.5 ml DNA solution. Electroporation treatments consisted of a single pulse at 100V, 120 V, 140 V, 160 V or 180V with each duration set at 10 ms and 35 ms. Between treatments, the Petri Pulser electrode was washed according to the manufacturer's protocol. Electroporated

cells were allowed to recover for 5 min at RT before adding additional CT-1 medium and incubating the plates.

**Lentiviral transduction of CT-1 cells**

Transduction optimization was performed using processed and purified LentiMax Lenti-RFP (LV-301RFP) and Lenti-GFP (LV-173GFP) lentiviral particles (Lentigen, Baltimore, MD). These stock vectors both utilize a simian CMV promoter to express high levels of fluorescent reporter. While the use of transduction enhancing reagent Polybrene is recommended, the cationic Lipofectamine 2000 was tested for transduction enhancement based on previous optimization efforts reported elsewhere (Syda et al., 2006). CT-1 cells were passaged into collagen-coated 96-well plates and cultured up to 9 days to achieve substantial surface area coverage by a few colonies. Cells treated with Lentivirus and Lipofectamine 2000 were first pre-incubated in OPTI-MEM containing Lipofectamine 2000 before LV particles were added to achieve the designated multiplicity of infection (MOI) ratios. MOI is defined as the ratio of transduction units per cell. BFFs and HEK-293 cells were also passaged into 96-well plates and cultured to over 90% confluency. As a positive transduction control, HEK 293 cells were transduced at 3 MOIs (5, 10, and 20). Each MOI treatment was additionally tested with or without Lipofectamine 2000 (1 µl/ ml) and with or without plate centrifugation (600 x g for 1 hr at 32 °C) prior to the overnight transduction incubation at 37°C and 8.5% $CO_2$. BFF and CT-1 cells were treated with increasing MOI (10, 20, 50, 100, and 200), with each treatment performed with or without Lipofectamine 2000 (1 µl/ml and 2 µl/ml). Plates that were not subjected to the plate centrifugation step were rocked by hand for 15-20

seconds every 30 min during the first few hours of the overnight incubation. Cells were assessed for fluorescent reporter expression every 24 hours for 5 days.

**Lipitoid-based plasmid transfection optimization**

The Lipitoid reagent was received as a gift from Dr. Ronald Zuckermann of the Molecular Foundry of the Lawrence Berkeley National Laboratory, Berkeley, CA. Efforts to optimize plasmid DNA transfection efficiency were adapted from the original report in which Lipitoid was used to transfect HT1080 cells (Huang et al., 1998). Based on the conversion factor that double stranded DNA has a mass of 660 grams per mol per base pair, 5 µg of the 6972 base pair phEFnGFP plasmid calculated to 15.2 nmol of negative charge equivalent. To determine which Lipitoid/ DNA charge (+/-) ratio (2/1, 3/1, or 4/1) resulted in the highest transfection efficiency, 5 µg of phEFnGFP was used per 6-well treatment, which is comparable in mass to other transfection reagent protocols at the 6-well plate scale, including Xfect (Clontech, Mountain View, CA) and Lipofectamine 2000 (Invitrogen). Each transfection reagent-plasmid treatment was tested in both collagen-coated 6-well plates and Cellbind 6-well plates. Prior to adding Lipitoid transfection cocktails, CT-1 cells were washed twice with DPBS, once with OPTI-MEM and then 1.6 ml OPTI-MEM were added back to each well. Lipitoid and DNA were separately suspended in OPTI-MEM to a total volume of 250 µl before lightly mixing together. Following 10 min incubation at RT, the 500 µl cocktails were then added to wells drop-wise, resulting in a total treatment volume of 2.1 ml. Plates were incubated (37 °C, 8.5% $CO_2$) for 4 hours or overnight followed by complete medium aspiration and replacement with CT-1 medium. Cells were assessed for nGFP reporter expression 40 hours later.

**CT-1 colony fixation and DNA staining**

Following set transfection incubations, cells were rinsed twice with DPBS and fixed with 4% formaldehyde for 30 min at RT. Fixed CT-1 colonies were then incubated in 10 µg/ ml Hoechst 33342 in DPBS for 10 min. Wells were rinsed twice in DPBS. Fixed colonies were then teased off of well bottoms using a fine paint brush (3/0, 9718, IMEX) and delicately applied to slides so that colonies did not have folded edges or tears. Following a brief drying period, slides were mounted with Fluoromount G (0100-01, Southern Biotech, Birmingham, AL) and cover slipped. Images of finished slides were captured using a Zeiss Axio Observer Z1 Inverted Microscope with Axiovision software (Carl Zeiss Inc., Thornwood, New York).

**Assessment of plasmid transfection efficiency**

Transfection efficiency for delivery of phEFnGFP was quantified by calculating the percentage of nGFP-positive, Hoechst 33342-stained nuclei to total Hoechst 33342-stained nuclei. For each transfection treatment, 4 to 6 images (20X magnification) were taken across the diameter of a representative colony. Over 50% of the nuclei in each image were counted (the range was between 379 and 1072 nuclei per image). This range is indicative of differences within monolayer cell density, with lower dense cell density occurring at the periphery. The trajectory of captured images across the colony diameter was also designated at random, as some regions of the colony were observed to possess higher transfection rates than other areas comparable in cell density and compact epithelial monolayer morphology. To determine whether CT-1 cells expressing nuclear GFP were alive and without compromised membrane integrity, transfected CT-1 cells were incubated in propidium iodide diluted in DPBS (2 µg per ml) and incubated for 5-10

min. Live cell images were then captured using a Leica DM IL inverted microscope (Leica, Bannockburn, IL) fitted with a Sony Cybershot 5.0 megapixel digital camera with a Martin Microscope MM19 adapter (Martin Microscope Company, Easley, SC).

**Short interfering RNA transfection optimization**

Optimization of Lipitoid based transfection in CT-1 cells is based on the original Lipitoid optimization methods with few modifications (Utku et al., 2006). CT-1 cells were grown in collagen-coated or Cellbind 6-well or 12-well plates for 7 to 10 days to allow CT-1 colonies to proliferate over 50% of the well surface area. Prior to the addition of Lipitoid/ double stranded RNA oligonucleotide mixtures, wells were washed twice with DPBS and once with OPTI-MEM. One ml OPTI-MEM was then added back to each well for 6-well plates (400 µl for 12-well). Lipitoid and RNA were both diluted in OPTI-MEM to total volumes of 100 µl for each 6-well (40 µl for 12-well) and gently vortexed and centrifuged before combining. Lipitoid/RNA suspensions were gently mixed and then incubated for 10 min at RT before being added drop-wise to each well. Treated CT-1 cells were then placed back into an incubator (37 °C, 8.5% $CO_2$) for 4 hours. Following incubation, FBS was added to each well so that final serum concentrations were at 10%. Cells were then placed back into the incubator overnight until morning, when the transfection mixture was completely replaced by CT-1 media. For Lipitoid/ BLOCK-iT fluorescent oligomer transfections (2013, Invitrogen), live cell assessment of fluorescein conjugated double stranded oligomer delivery into CT-1 nuclei was performed 24 hours later using the Leica DM IL inverted microscope fitted with a Sony Cybershot 5.0 megapixel digital camera. Lipitoid/ Stealth siRNA transfected cells were incubated for 48 hours following medium replacement and then colonies were lysed for RNA extraction.

**RNA interference and quantitative RT-PCR**

Short-interfering RNA oligonucleotide-mediated knockdown of both *OCT4* and *CDX2* transcript levels was performed using custom Stealth RNAi siRNA duplex oligonucleotides (Invitrogen). A bovine *OCT4* (Pou5f1) 25-nucleotide duplex siRNA (3 2) was designed using Invitrogen's BLOCK-IT RNAi designer. A bovine *CDX2* 25-nucleotide duplex siRNA was modified from a previously reported functional 19-nucleotide siRNA sequence (siRNA2; Sakurai et al., 2009) using Invitrogen's siRNA to Stealth RNAi siRNA converter. A nonspecific scrambled 25-nucleotide duplex siRNA was also designed to serve as a negative RNA interference control. Similar to BLOCK-iT fluorescent 21-mer oligonucleotides, Stealth siRNAs possess a net charge of -50 and were used at a 20 µM stock concentration. Therefore, any modifications to Lipitoid and Lipofectamine 2000 transfection protocols were unnecessary. CT-1 colonies were quickly washed twice with DPBS and lysed within the well with RNeasy lysis buffer (Qiagen, Valencia, CA) 48 hours after the end of the 4 hour transfection incubation. RNA was further extracted, DNase I treated, and purified using the RNeasy Miniprep kit (74124, Qiagen). Purified RNA was quantified using the Quant-iT RiboGreen RNA Assay Kit (R11490, Invitrogen) and each sample was diluted to a final concentration of 250 ng per µl. One µg of total cellular RNA (tcRNA) from each sample was then used to synthesize first strand cDNA using an AffinityScript Multiple Temperature cDNA Synthesis Kit (200436, Stratagene, La Jolla, CA). First strand cDNAs were synthesized using oligo(dT) primers according to the manufacturer's instructions with one modification: during synthesis, cDNA reactions were incubated at 55°C for 1 hour. Completed 20 µl cDNA reactions were then diluted 10-fold with nuclease-free water before using as template for

quantitative RT-PCR (qRT-PCR). To account for potential genomic DNA contamination amplification during RT-PCR, 1µg of pooled tcRNAs (3-4 combined equivalents per treatment) were also used as template for cDNA synthesis, but reverse transcriptase was omitted from these reactions.

Sense and antisense primers for qRT-PCR of *OCT4* and *CDX2* were developed using IDT Real Time PCR software (Integrated DNA Technologies, http://www.idtdna.com/Scitools/ Applications/RealTimePCR) and designed to span across exon-intron splice junctions (Table 2.1). *OCT4* primers were based on published sequence (Accession NM_174580.1) and verified to specifically amplify *OCT4* mRNA and not transcribed *OCT4* pseudogene transcripts. *CDX2* primers were also designed from published sequence (Accession XM_871005.3). Primers designed to gamma–actin (*ACTG1*, Accession NM_001033618.1) were utilized for qRT-PCR amplification normalization. *CDX2* and *ACTG1* qRT-PCR reactions consisted of 7.5 µl 2X iQ SYBR Green Super mix (170-8882, Bio-Rad), 0.6 µl of both sense and antisense 10 µM stock primers, 5.3 µl nuclease-free $H_2O$, and 1 µl diluted cDNA or 'RT-' pooled control. The 2X iQ SYBR green mix contains fluorescein as an internal reference to normalize well-to-well optical variation. qRT-PCR was performed using the MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad) programmed with the following steps: initial denaturation at 95°C for 3 min and 40 repeated steps consisting of 15 seconds at 95°C and extension at 60°C for 1 min. After 40 cycles, a dissociation (melt) curve was performed. Each PCR reaction was performed on 3 independently treated cDNA replicates, and each cDNA was amplified in duplicate. All cDNA amplifications produced single peaks in the melt curve analysis, and PCR products were further

analyzed by gel electrophoresis for correct size. Single amplicon bands were purified from agarose gels using a MinElute PCR Purification Kit (28004, Qiagen) and sequenced at the University of Maryland Biotechnology Institute: Center for Biosystems Research (College park, MD). The data output of *OCT*4 and *CDX2* Ct (threshold cycle) values, representing fixed threshold crossing points where all samples are undergoing logarithmic amplification, were first normalized to *ACTG1* Ct values using the formula: $\Delta Ct = (Ct_{\text{`no RT'}} - Ct_{\text{sample}})_{\text{target}} - (Ct_{\text{`no RT'}} - Ct)_{ACTG1}$. $\Delta Ct$ values were then transformed ($2^{\Delta Ct}$) and data were represented as relative abundance of target mRNA normalized to *ACTG1* levels.

**Statistical Analysis**

Rates of DNA plasmid transfection were presented as percentages of nGFP positive cells to the total number of cells as determined by Hoechst 33342 staining. Transfection percentages were analyzed by ANOVA using the PROC MIXED model in SAS statistical software (SAS Institute, Cary, NC). Lipitoid/ DNA (+/-) charge ratio and incubation length were included as independent sources of variation. Data transformation for a binomial distribution was unnecessary as assumptions for variance heterogeneity and normality were met. As a significant difference in transfection paradigms was detected ($P<0.05$), the main effects of (+/-) charge ratio and incubation were compared. While a significant difference in incubation was detected ($P< 0.05$), no significant differences between (+/-) charge ratios or interactions between main effects were detected. Therefore, differences in individual transfection paradigms were analyzed using the PDIFF procedure with a Tukey's adjustment. Data are presented as treatment means ± pooled SE as calculated by the PROC MIXED procedure. Levels of gene expression were

calculated as a ratio of treatment ΔCt values (described in Methods and Materials) relative to untransfected control ΔCt values of at least 3 independent experimental replicates. Presented data are the means and SE of these ratios. Data were first log transformed to meet assumptions for variance heterogeneity and normality and analyzed by ANOVA using the PROC MIXED model in SAS. Lipitoid (+/-) charge ratio, siRNA concentration, and siRNA treatment (untransfected, target, control) were classified as independent variables. As a significant difference in transfection paradigms was detected ($P<0.05$), individual treatments were compared to the untransfected control using the PDIFF procedure with a Dunnett's adjustment. Independent contrasts were also made to compare targeted siRNA effects to control siRNA effects within individual transfection treatments. For all statistical analysis, differences are considered significant at $P<0.05$.

## Table 2.1. siRNA Stealth Oligomer and Quantitative RT-PCR Primer information.

| siRNA | Sequence (5'-3') | Length (nt) | Target Genbank Accession # | Location | %GC Content | Design |
|---|---|---|---|---|---|---|
| Bovine OCT4 | GAGAGGUGUUGAGCAGUCUCUAGGA CUCUCCACAACUCGUCAGAGAUCCU | 25 | NM_174580.1 | 5' UTR | 52 | Invitrogen BLOCK-IT RNAi designer |
| Bovine CDX2 | CUCUCAGAGAGGCAGGUUAAAAUUU GAGAGUCUCUCCGUCCAAUUUUAAA | 25 | XM_871005.3 | ORF | 40 | Adapted from siRNA#2 (Sakurai et al., 2009) |
| Negative Control | GAGGUGGUUCGAUGACUCAUAGGGA CUCCACCAAGCUACUGAGUAUCCCU | 25 | No target | | 52 | Invitrogen BLOCK-IT RNAi designer |

| QPCR Primers | Sequence (5'-3') | Length (nt) | Genbank Accession # | Location | Amplicon Length (nt) | Annealing Temp. (°C) |
|---|---|---|---|---|---|---|
| Bovine OCT4 (FWD) | CAAATTAGCCACATCGCC | 18 | NM_174580.1 | Exon 4 | 126 | 60 |
| Bovine OCT4 (REV) | AGCCTCAAAATCCTCACG | 18 | | Exon 5 | | |
| Bovine CDX2 (FWD) | AGTCGTTATATCACCATCC | 19 | XM_871005.3 | Exon 2 | 104 | 60 |
| Bovine CDX2 (REV) | CTTTCCTTTGCTCTGCG | 17 | | Exon 3 | | |
| Bovine ACTG1 (FWD) | TTGCTGACAGGATGCAGAAG | 20 | NM_001033618.1 | Exon 4 | 145 | 60 |
| Bovine ACTG1 (REV) | TGATCCACATCTGCTGGAAG | 20 | | Exon 5 | | |

**Results**

We were successful in adapting CT-1 cell culture to a substrate-free, nonbiological Corning Cellbind surface and have passaged CT-1 cells on Cellbind for over 20 passages. We observed no differences in colony seeding, expansion and growth rate between CT-1 cells grown on collagen or Cellbind. However, minor differences in colony morphology were noted. Colonies grown on Cellbind appear slightly flattened and cells located at the colony periphery are less compacted. Also, colonies grown on Cellbind are more prone to detachment following physical disturbance resulting from various transfection protocols (Table 2.2), indicating that culturing CT-1 cells grown on a substrate may be beneficial for some experimental procedures. No extra consideration is needed for passaging CT-1 cells between collagen and Cellbind surfaces, and vice versa.

We also developed an enzymatic cell dispersal protocol using dispase and collagenase that was comparable to mechanical passaging in cell-aggregate-colony seeding rates and colony recovery. However, this method allowed for cell aggregates as small as 10 cells to adhere and grow into typical colonies.

We tested 7 techniques to deliver DNA into CT-1 cells. As CT-1 cells exhibit autofluorescence that may confound confirmation of a cytoplasmically localized GFP, plasmids that constitutively produce high levels of nuclear localized GFP were used in these experiments. Lipitoid-mediated transfection resulted in the greatest numbers of transfected nGFP positive CT-1 cells (8.9%). Transfection efficiencies for the remaining 6 techniques were never above 1%. Cationic liposomal reagents such as Lipofectamine 2000 (Dalby et al., 2003), which have been used successfully on a multitude of cell lines, had almost no effect on CT-1 cells despite extensive optimization efforts. No obvious

increases in transfection were observed when amounts of DNA or Lipofectamine 2000 were increased at elevating ratios of 1:1, 1:2.5, and 1:3.75. Nupherin was added to Lipofectamine 2000-based treatments to aid nuclear delivery of transfected plasmid DNA. Only treatments with high levels of Nupherin (45 and 60 µg Nupherin per µg DNA) resulted in slight but insignificant increases in nGFP positive cells: However, these levels were toxic and damaged colonies. A cell suspension method for hard-to–transfect cell lines (Zhang et al., 2007) was also tested. Basal and lateral membranes may be exposed during suspension and, therefore, available for DNA/ liposome uptake. All suspension transfection treatments (fixed 1:2.5 DNA/ Lipofectamine 2000 (µg/µl) ratio; 0.8 µg, 1.2 µg, and 4.8µg plasmid) at all incubation periods (15 min, 60 min, 120 min, and overnight at 37°C) produced a high density of seeded colonies, yet very few nGFP positive cells were observed.

Similar adherent colony transfection results were obtained using other cationic lipid reagents such as TransFectin (Bio-Rad) and Metafectene (Biontex, San Diego, CA), or the polymer Xfect (Clontech), which were all used according to manufacturer's instructions. Noncationic liposomal transfection methods including electroporation, biolistic delivery, and lentiviral transduction were also attempted. Colonies electroporated at all voltages (100-180 V) resulted in cell damage. At higher voltages (140-180 V) the damage was more extensive and whole colonies detached. Growing CT-1 colonies on mitotically-inactive STO MEFs stabilized colonies from detaching and improved recovery. However, even the best set of parameters (100 V, 35 ms) produced less than 10 clusters of clonally expanding, nGFP positive cells per 6-well. Results for biolistic delivery were similar.

**Table 2.2. Summary of attempted transfection methods for delivery of DNA into CT-1 cells.**

| Method Description | Rationale | CT-1 Cell Line Results | Cell Line Control Results |
|---|---|---|---|
| Lipofectamine 2000 transfection of adhered colonies | Optimization of the standard Lipofectamine 2000 protocol. DNA and reagent concentrations were increased to determine if there is a resistance threshold to transfection. | No significant observable increase in transfection with increasing µg plasmid or DNA/ lipofectamine 2000 (µg/ µl) ratio. (TE < 1%.) | HEK 293 cells transfected at high levels (TE >80%). BFFs transfected at moderate levels (TE= 50-70%). |
| Lipofectamine 2000 transfection of suspended cell clumps | Determine if transfection complex delivery is blocked at apical membrane only. Clumped cells in suspension should have some exposed lateral and basal membranes. | No significant observable increase in transfection with increasing µg plasmid or DNA/ lipofectamine 2000 (µg/ µl) ratio. (TE < 1%.) | No control cell lines were tested. |
| Lipofectamine 2000 transfection with Nupherin supplementation | Determine if transfection resistance is due to inability of transfection complex to be imported into nucleus. | Overnight incubation in Nupherin transfection complex was toxic. High levels of nupherin (45 and 60 µg/ µg DNA) slightly increased TE. | No control cell lines were tested. |
| Biolistic delivery of plasmid DNA into adhered colonies | Bypass any transfection complex issues occurring at the cell and nuclear membranes by physically delivering DNA into nucleus. | Successful optimization of bombardment velocity to decrease cell damage and colony detachment. TE rates are low and confined to bombardment radius (TE < 1%). | BFFs transfected at low levels within bombardment radius (TE <10%) with and without the agar overlay. |
| Electroporation of adhered colonies | Determine if electroporation results in higher TE than other methods. The Petri Pulser electrode electroporates adhered colonies. | A single pulse (100 V, 35 mS) produced the greatest number of transfected cell clumps near edges of colony/electrode contact. TE is not acceptable for transient transfection. | Adhered, undamaged HEK 293 cells were transfected at high levels (TE >80%). |
| Lentiviral transduction of adhered colonies | Determine if lentiviral transduction results in higher TE than other methods. Commercial lentiviral vectors were used for the optmization process. | No transduction was observed at any MOI level (10-200). Lipofectamine supplementation and brief centrifugation had no effect for any tested cell lines. | HEK 293 cells were transduced at moderate rates (TE = 50%) at 10X MOI level. Very few transduced BFFs observed at 200X MOI level. |

Our optimization efforts increased transfection efficiency in BFFs but not CT-1 cells. Lentiviral transduction was also performed using GFP and RFP lentiviral vectors. While HEK 293's showed moderate levels of transduction at a MOI of 10, success in bovine cell lines CT-1 and BFF was limited to a few transductions in BFFs at the highest MOI tested (200).

Lipitoid/ DNA transfection treatment cocktails were assembled according to Lipitoid/DNA (+/-) charge ratios of 2/1, 3/1 and 4/1, and each ratio-based transfection cocktail was tested at 4 hour and overnight incubations (Fig. 2.1). Overnight incubations resulted in significantly higher transfection efficiencies than 4 hour incubations ($p < .05$). CT-1 cells incubated overnight in the 2/1 Lipitoid/DNA (+/-) charge ratio transfection cocktail produced the highest degree of transfection efficiency (8.9%). CT-1 cells incubated in the least effective transfection cocktail (4/1 Lipitoid/DNA (+/-) charge ratio for only 4 hours) were still transfected at higher rates (2.7%) than all non-Lipitoid transfection methods tested (Table 2.2). However, for all overnight incubations, including "Lipitoid only" controls, obvious signs of cell damage and toxicity were observed as the Lipitoid amount increased. Colonies incubated overnight in the 4/1 (+/-) charge ratio transfection were not included in Fig. 1 as colony damage was too extensive. Colonies incubated overnight in 2/1 and 3/1 (+/-) charge ratio transfections recovered from Lipitoid toxicity within 24-48 hours and were stained with propidium iodide to determine cell viability (Fig. 2.2). Propidium iodide stained nuclei, which indicated dead cells, did not co-localize with nuclear GFP+ nuclei, demonstrating that transfected cells were viable.

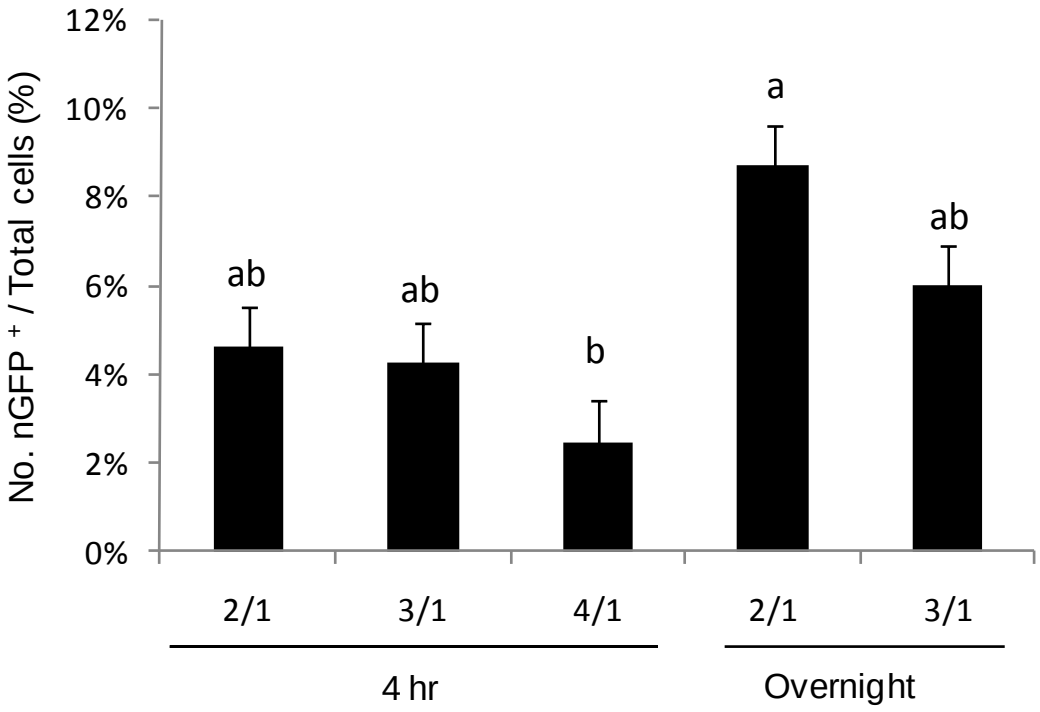



**Figure 2.1.** Optimization of plasmid DNA transfection. 5 µg phEFnGFP plasmid was transfected into CT-1 cells at three Lipitoid/ DNA (+/-) charge ratios (3/1, 4/1, 2/1). Each ratio-specific treatment was compared at two incubation periods (4 hr vs. overnight.) Cells were cultured for an additional 48 hours after transfection, fixed and stained with Hoechst 33342. Graphed values represent the mean and pooled SE of the total number of nuclear GFP expressing cells over total nuclei (n=2). Means that share letter superscripts are not significantly different ($P < 0.05$).

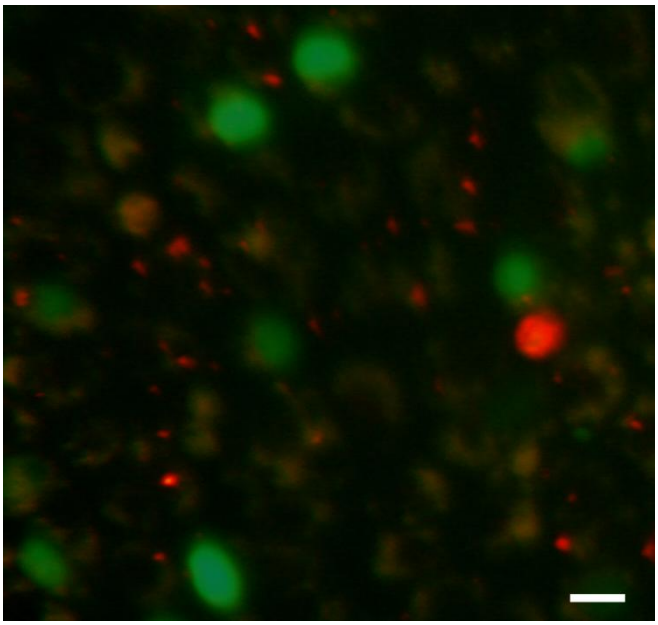


**Figure 2.2.** Visualization of dead cells within phEFnGFP transfected CT-1 colonies. Propidium Iodide (red), a nucleic acid stain that intercalates DNA of non-viable cells, does not stain nuclei of cells expressing nuclear localized GFP (green). Bar = 10 μm.

Based on our success with Lipitoid-based plasmid transfection, we next optimized Lipitoid-based oligonucleotide transfection. A visual reporter-based approach was first used to qualitatively optimize siRNA duplex transfection in CT-1 before proceeding with the testing of functional siRNA assays using qRT-PCR. BLOCK-iT is a fluorescein tagged RNA duplex oligonucleotide that possesses a similar net negative charge to short-interfering double stranded nucleotides. Transfection cocktails containing either 50 nM or 100 nM BLOCK-iT fluorescent oligomer were each tested at Lipitoid/DNA (+/-) charge ratios of 2/1, 3/1 and 4/1. At the 50 nM BLOCK-iT, only the transfection performed at 4/1 (+/-) charge ratio resulted in faint but identifiable nuclear fluorescence (Fig. 2.3a). All transfections performed with 100 nM BLOCK-iT fluorescent oligomer resulted in a high degree of transfection, with CT-1 cells exhibiting high levels of nuclear fluorescence (Fig. 2.3b and c). The 100 nM treatment at the 4/1 (+/-) charge ratio had the most abundant uptake (Fig. 2.3c) while no obvious differences were observed between 100 nM treatments at 3/1 and 2/1 (+/-) charge ratios. Based on these results, 3/1 and 4/1 (+/-) charge ratios were chosen as treatment ratios in further optimization with functional siRNA. Although the 50 nM BLOCK-IT transfection at the 3/1 (+/-) charge ratio resulted in virtually undetectable levels of fluorescence in our tests, this treatment was included since it had been shown previously to decrease levels of DLX3 protein in CT-1 cells (Ezashi et al., 2008).

Functional siRNAs were designed to suppress expression of transcription factors CDX2 and OCT, which are expressed in CT-1 cells (Schiffmacher and Keefer, 2008). Each siRNA was transfected into CT-1 cells at 50 nM and 100 nM concentrations and each concentration was mixed with Lipitoid to produce 3/1 and 4/1 (+/-) charge ratios.

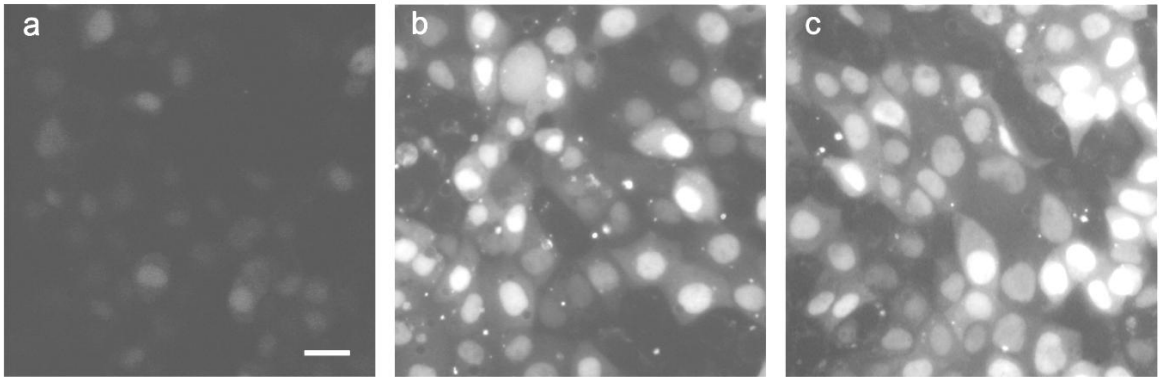



**Figure 2.3.** Visualization of fluorescent tagged BLOCK-IT RNA duplexes in live CT-1 cells according to transfection parameters: a) 50 nM oligomer transfected at 4/1 Lipitoid/ DNA (+/-) charge ratio, b) 100 nM oligomer transfected at 3/1 (+/-) charge ratio, and c) 100 nM oligomer transfected at 4/1 (+/-) charge ratio. Images were taken 48 hours after the end of the transfection incubation. All images are at the same magnification. Bar =20 µm.

Transfection efficiency was determined by evaluating the degree of mRNA knockdown as determined by qRT-PCR (Fig. 2.4). All transfection parameters tested exhibited the similar degree of *CDX2* transcript downregulation (59-61%) and were significantly different from untransfected CT-1 levels and negative siRNA controls (Fig. 2.4a, $p < 0.05$). In addition, none of the siRNA treatments, irrespective of concentration or charge ratio, resulted in nonspecific alterations to *CDX2* expression levels. Therefore, significant *CDX2* specific targeting and knockdown was achievable at both siRNA concentrations (50 nM vs. 100 nM) and Lipitoid/DNA (+/-) charge ratios (3/1 vs. 4/1). Specific siRNA targeting of *OCT4* mRNA resulted in different rates of downregulation between transfection treatments (Fig. 2.4b). *OCT4* expression was significantly down-regulated by 58% using 50 nM *OCT4* siRNA transfected at a 4/1 Lipitoid/DNA (+/-) charge ratio and was downregulated by 55% using 100 nM *OCT4* siRNA transfected at a 3/1 Lipitoid/DNA (+/-) charge ratio. However, the difference between *OCT4* siRNA and negative control transfected at 100 nM was not significant. Unlike the results obtained from the *CDX2* siRNA optimization, there was an indication that transfections of 100 nM siRNA may exhibit a nonspecific effect on *OCT4* expression as quantified by qRT-PCR. Therefore, 50 nM *OCT4* siRNA transfected at a 4/1 Lipitoid/DNA (+/-) charge ratio was the most effective treatment for suppression of OCT4 levels.

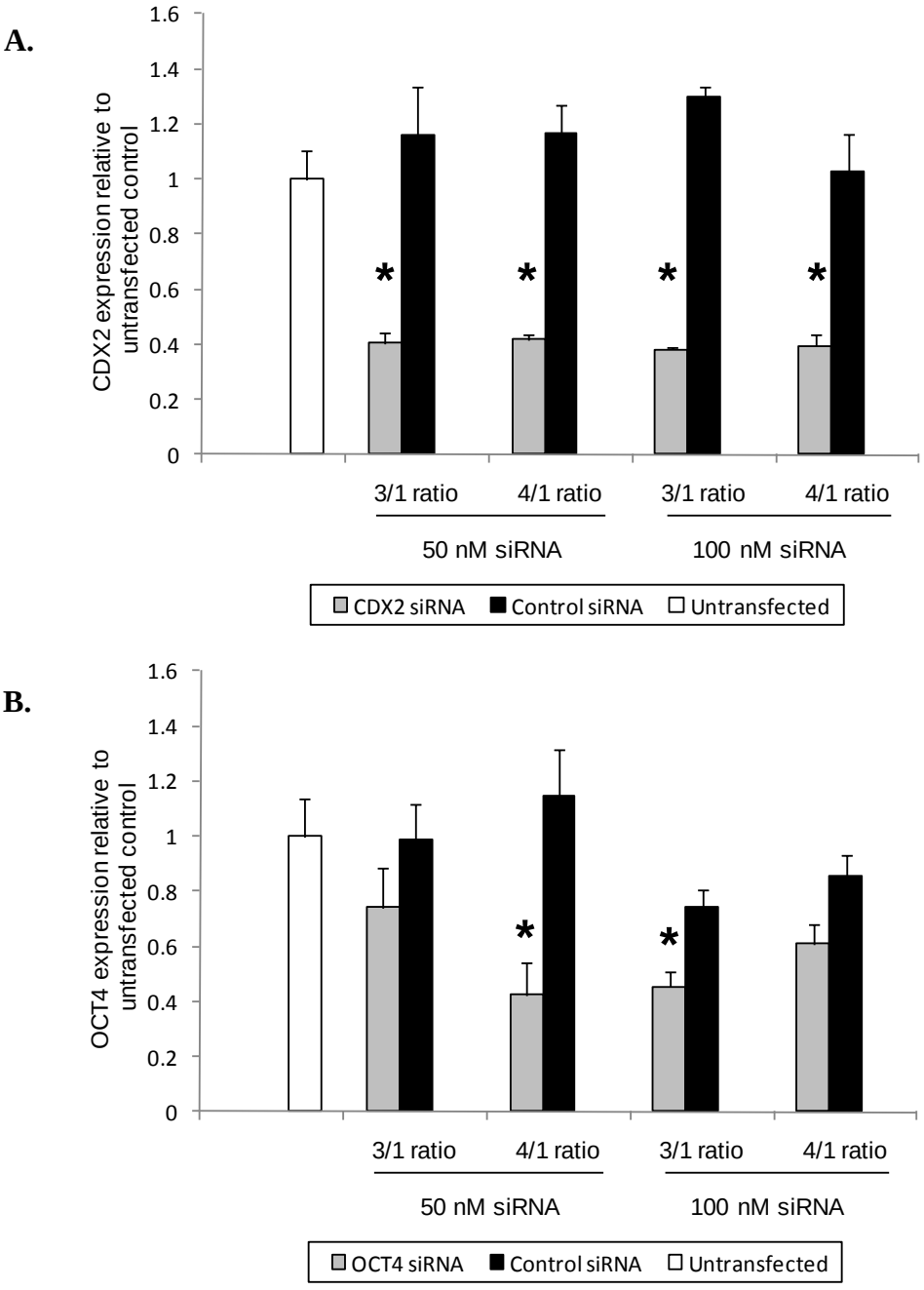



**Figure 2.4.** mRNA levels of transcription factors *CDX2* (a) and *OCT4* (b) following siRNA knockdown. Two siRNA concentrations (50 nM vs. 100 nM) were tested at two Lipitoid/ DNA (+/-) charge ratios (3/1 vs. 4/1). For each treatment, a nonspecific control siRNA was also transfected to serve as a negative control. Each transfection treatment is expressed as the mean and SE (n=3) of the siRNA target relative expression level normalized to treatment control levels. An asterisk (*) denotes a significant difference between treatment effect and untransfected control ($p < .05$).

**Discussion**

For the CT-1 cell line to be useful as a model for studying gene regulation, it must be amenable to genetic manipulation through exogenous DNA or RNA uptake. A battery of well-characterized nucleic acid delivery methods did not even remotely achieve transfection levels in CT-1 cells comparable to those observed in the control cell lines (Table 2.2). In contrast, Lipitoid transfection resulted in a 9% rate for overall colony transfection with 5 µg phEFnGFP incubated overnight at a 2/1 Lipitoid/DNA (+/-) charge ratio. Transfection rates up to 15% were observed in peripheral regions of these colonies. Slight toxicity at the 2/1 ratio was also observed, although colonies quickly recovered and proliferation of transfected cells was observed 4 days following transfection. While a 9% transfection rate may not be sufficient for many experiments, it should prove acceptable for generating stable transfected CT-1 cell lines by clonal patching of transfected mononucleate cell clusters.

In our hands cationic Lipofectamine 2000 was completely ineffective up to 3 times the amount of DNA used for typical transfections in 24-well plates. Furthermore, Lipofectamine 2000 based transfection of small CT-1 aggregates in suspension or supplemented with Nupherin did not improve rates.

Successful electroporation of DNA using a Petri Pulser electrode (BTX) can be accomplished in HEK 293 cells but not CT-1 cells. Biolistic delivery of DNA was found to be the most inconsistent technique and produced the lowest transfection rates in confluent BFF. Lentiviral transduction is a powerful technique and has been used successfully to reprogram porcine fetal fibroblasts into a pluripotent state (Ezashi et al, 2009). However, attempts to transduce CT-1 cells with lentiviral fluorescent reporter

vectors were unsuccessful, including the testing of other previously optimized protocols using Lipofectamine 2000 supplementation (Syda et al., 2006).

The Lipitoid reagent also proved effective in our optimization efforts to deliver oligonucleotides. Using two representative siRNAs to target transcription factors *CDX2* and *OCT4*, we demonstrated that 50 nM siRNA transfected at a 4.1 Lipitoid/DNA (+/-) charge ratio reduced both *CDX2* and *OCT4* transcripts by almost 60%. However, for *CDX2* knockdown, 50 nM siRNA transfected at a 3/1 Lipitoid/DNA (+/-) charge ratio was just as effective. This confirms similar findings wherein 50 nM *DLX3* siRNA was used at the 3/1 (+/-) charge ratio to efficiently reduce DLX3 expression (Ezashi et al., 2008). With further optimization of siRNA sequences, Lipitoid-based siRNA transfection may be even more effective in CT-1 cells.

The reasons why CT-1 cells are hard to transfect remain unknown, but the major block in nucleic acid delivery most likely occurs at the membrane. DNA complexed to cationic liposomes composed of Lipofectamine 2000 form heterogenous complex shapes and may not sufficiently mask the DNA negative charge to overcome electrostatic repulsion of the cell membrane and extracellular matrix. This repulsion issue may also explain the lack of success in electroporation. Another reason for the lack of transfection includes insufficient protection from possible DNases secreted from the CT-1 cells into the transfection medium. Lastly, the lipid chemistry of the CT-1 apical membrane may be quite unique and incompatible with the cationic lipids to promote membrane/ liposome fusion. The transfection success obtained with Lipitoid reagent may result from the unique interaction with Lipitoid and DNA. Electron microscope images of complexes reveal uniform spherical structures that presumably consist of plasmid enveloped inside

Lipitoid spheres (Huang et al. 1998). These spheres provide adequate protection from DNases and their condensed structure may be more effective in overcoming electrostatic cell membrane repulsion or membrane fusion and nucleic acid release into the cytoplasm.

In summary, our efforts demonstrate that the technical issues in CT-1 culture surfaces, passaging, and nucleic acid transfection can be successfully addressed and further optimization should increase the utility of CT-1 cells as model cell line for studying trophoblast development.

# Chapter 3: The roles of OCT4 and CDX2 in directing early bovine trophoblast development.

**Abstract**

Misregulation of genes in the early trophectoderm (TE) has been implicated to play a causal role in abnormal placental formation and fetal loss of cloned and in vitro-produced embryos. In mice, the transcription factor CDX2 is a key regulator that is required to repress the pluripotency specific transcription factor OCT4 during TE lineage segregation and direct trophoblast maintenance. However, during bovine embryogenesis CDX2 does not completely repress OCT4 and both genes are co-expressed during early trophoblast elongation. To elucidate the roles of CDX2 and OCT4 in the trophoblast transcriptional hierarchy, we utilized the bovine trophectoderm-derived CT-1 cell line as a genetic model. RT-PCR analysis showed that CT-1 cells expressed transcripts for TE lineage associated transcription factors *CDX2, ETS2, ERRB, ID2, SOX15, ELF5, HAND1, MASH2* and *GATA6*. Expression of pluripotency-associated transcription factors *NANOG* and *SOX2* were undetectable, while *OCT4* expression was evident. *CDX2* knockdown reduced levels of *IFNT, HAND1, MASH2*, and *SOX15*. Overexpression of CDX2 in CT-1 cells had a reciprocal effect on these genes, in addition to increasing *ELF5* transcript levels. CDX2 also reduced *OCT4* levels. Both overexpression and knockdown of *CDX2* increased *ETS2* levels. Overexpression and knockdown of *OCT4* had no effect on any candidate gene levels with the exception of positive feedback on itself. These results implicate CDX2 as a regulator of OCT4 and multiple downstream bovine trophoblast genes and demonstrate that CT-1 cells are useful as a genetic model for studying the transcriptional networks governing the early bovine trophoblast.

## Introduction

During eutherian embryogenesis, the first tissue lineage segregation event forms the inner cell mass (ICM), which gives rise to the embryo proper, and the trophectoderm (TE), the precursor to placental trophoblast lineages. Embryonic stem cells (ESC) have been successfully utilized as a genetic model for studying differentiation of these early embryonic lineages. Chromatin-immunoprecipitations coupled with DNA microarray analyses performed in both mouse and human ESC reveal that transcription factors OCT4 and NANOG, in concert with transcription factors SOX2 and SALL4, co-regulate hundreds of genes to maintain pluripotency (Boyer et al., 2005; (Loh et al., 2006; Lim et al., 2008; Yang et al., 2008). Downregulation of OCT4 below a threshold level required for pluripotency maintenance permits induction of TE specific genes and consequential differentiation into trophoblast cells (Niwa et al., 2000; Niwa et al 2005). In the mouse, the caudal type transcription factor CDX2 is a primary regulator at the core of the TE lineage hierarchy based on its ability to suppress *Oct4* and *Nanog* expression. While critical reduction of OCT4 by RNA interference only permitted differentiation into trophoblast cells, reduction of OCT4 by CDX2 overexpression not only forced differentiation but maintained trophoblast stem cell self renewal (Niwa et al., 2005). CDX2 is involved in inhibitory feedback loops with OCT4 and NANOG as CDX2 represses *Oct4* and *Nanog* transcription and vice versa (Chen et al., 2009; Niwa et al., 2005). Mouse embryo knockout models have confirmed that this network functions in vivo as predicted by ESC studies. OCT4 and NANOG are inappropriately expressed in the TE of *Cdx2* $^{-/-}$ mutant blastocysts (Strumpf et al., 2005) and this expression is autonomous as demonstrated by OCT4 expression in *Cdx2* $^{-/-}$ cells positioned in the TE of

chimeric (*Cdx2* $^{-/-}$ /WT) embryos (Ralston and Rossant, 2008). Conversely, ectopic expression of CDX2 was detected in early *Nanog* $^{-/-}$ mutant ICMs (Chen et al., 2009) and later in the compromised ICMs of *Oct4* $^{-/-}$ mutant blastocysts (Ralston and Rossant, 2010).

Beyond the blastocyst stage, the mouse embryo is not an appropriate model for ungulate embryogenesis. While the mouse blastocyst implants shortly after hatching from the zona pellucida, hatched bovine embryos delay attachment and undergo a different developmental program (Betteridge and Flechon, 1988). Bovine embryos elongate into an ovoid structure around 12 days post conception (dpc), transition into a longer tubular structure, and become filamentous by 15 dpc. At the ovoid stage, the bovine polar TE, or Rauber's layer, slowly degenerates while the underlying ICM surfaces as the germinal disk. Unlike the polar TE in the mouse blastocyst, the Rauber's layer does not maintain a trophoblast stem cell population, nor does it differentiate into extraembryonic ectoderm or ectoplacental cone lineages. The mural TE in ruminants is responsible for the considerable embryonic elongation (Betteridge and Flechon, 1988). Very little is known about initial ungulate trophoblast lineage maintenance and the molecular mechanisms driving trophoblast expansion (Blomberg et al., 2008). Bovine trophoblast proliferation does not appear to be driven by FGF-mediated signaling. However, in vivo studies in adult ewes lacking proper endometrial gland density show that survival and proper development at the tubular stage and beyond is dependent on endometrial secretions (Gray et al., 2001; Gray et al., 2002).

Spatial and temporal expression of CDX2 protein is similar between mice and other species in that protein is restricted to the TE by the blastocyst stage (Douglas et al.,

2009; Kuijk et al., 2008; Sritanaudomchai et al., 2009). However, patterning of OCT4 beyond the blastocyst stage is dissimilar between mice and other species. While OCT4 protein is readily detected in all cells of the early blastocyst in all species, it is rapidly downregulated during blastocyst expansion in mice (Dietrich and Hiiragi, 2007) and primates (Mitalipov et al., 2003; Sritanaudomchai et al., 2009). Contrarily, OCT4 protein expression is not immediately downregulated but maintained in the TE into the ovoid stage of bovine, caprine, and porcine embryos (Nichols et al., 1998; Kirchhof et al., 2000; Roberts et al., 2004; van Eijk et al., 1999; He et al., 2006; Keefer et al., 2007). Localization of OCT4 protein within TE nuclei of ovoid stage embryos of many species may call into question the universal role of OCT4 as 'gatekeeper' during the first lineage segregation (Pesce and Scholer, 2001). However, studies in mouse ESC demonstrate that this 'gatekeeper' function may be protein level dependent (Niwa et al., 2000; Niwa et al 2005) and cofactor dependent (Hemberger et al., 2009). Therefore, *OCT4* downregulation, but not repression, may be sufficient for TE lineage segregation. This evidence is corroborated with in vivo data demonstrating that *OCT4* transcript levels in ICM cells are much greater than *OCT4* mRNA levels in bovine and caprine TE cells (Degrelle et al., 2005; He et al., 2006). It has been hypothesized that OCT4 may play a role in maintaining proliferative TE cells in a "differentiation delayed state" (Degrelle et al. 2005; Kurosaka et al., 2004).

To gain insight into how ungulate trophoblast lineage segregation and maintenance are regulated, many studies have evaluated expression profiles of lineage markers initially identified in the mouse. Degrelle and colleagues demonstrated that transcripts of transcription factors *CDX2, EOMES, GATA6, HAND1* and *ETS2* are

detected throughout all stages of bovine elongation, although it was noted that *EOMES* was not expressed in the filamentous TE (Degrelle et al., 2005). Relative expression levels bHLH transcription factor HAND1 and the pregnancy recognition factor IFNT appeared to be positively correlated with embryonic growth. *ELF5* transcripts were readily detected at the ovoid stage, corroborating mouse data that Elf5 expression is downstream of Eomes and Cdx2 in mouse TE development (Donnison et al., 2005; Ng et al., 2008).

Embryo-derived cell lines derived from bovine blastocysts can be used in the same manner mouse embryonic-derived cell lines are utilized for exploring the roles of OCT4 and CDX2 in directing lineage maintenance and differentiation (Ralston and Rossant, 2005), Our studies were performed in the bovine trophoblast CT-1 cell line which was derived from hatched day 10-11 bovine blastocyst explants and has served as a genetic model for elucidating regulation of *IFNT* expression (Talbot et al., 2000; Ealy and Yang, 2009). However, the capacity of CT-1 cells to be used as a developmental model beyond *IFNT* regulation has not been explored. Therefore, the aims of this study were to further identify which trophoblast specific genes are expressed in CT-1 cells and to determine if they have a similar expression pattern to in vivo early elongating trophoblasts. Furthermore, the effect of OCT4 and CDX2 equilibrium disruption on expression of downstream genes was determined. Together these findings increase our understanding of the roles of OCT4 and CDX2 in maintaining developmental potential in the early bovine trophoblast.

## Methods and Materials

### Cell culture and tissue collection

CT-1 cells were received as a gift from Dr. Neil Talbot (U.S. Department of Agriculture, Beltsville, MD). Cells were initially expanded on mitotically inactive, gamma-irradiated mouse STO mouse embryonic fibroblast (MEF) feeder layers (CRL-1503, ATCC, Manassas, VA) at a density of 40,000 cells per $cm^2$. CT-1 and STO MEF co-cultures were grown in CT-1 medium consisting of DMEM containing 4.5 g per L glucose and 4 mM L-glutamine (11965-084, Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS) (SH30070.03; Hyclone, Logan, UT), 1 mM sodium pyruvate (11360-070, Invitrogen) and 50 U per ml penicillin-streptomycin solution (15070-063, Invitrogen). Cultures were incubated at 37°C in 5% $CO_2$ for 4 to 5 days and passaged when either the majority of colonies reached a threshold 4 to 5 mm in diameter or when CT-1 colonies reached 50% confluency. Routine passaging consisted of mechanical dissociation by gentle scraping with a cell scraper. Clumps were further dissociated by being drawn and expelled two times through a 22-gauge needle attached to a 20 ml syringe. CT-1 and STO MEF co-cultures were passaged at 1:2 to 1:3 (v/v) into collagen-coated cultureware prepared in the following manner: Collagen solution (pH 7.4; 40% PureCol collagen (5005, Advanced Biomatrix, San Diego, CA), 5% 10X DPBS (14200-075, Invitrogen) 5% 0.1 M NaOH, and 50% CT-1 medium) was added to cultureware to coat the entire bottom and then aspirated to leave a thin collagen film. Coated flasks or plates were then incubated at 37 °C for one hour before air drying overnight at room temperature (RT). Right before use, dried collagen-coated wells were rinsed with CT-1 medium. CT-1 cells were considered successfully adapted and free of

STO MEF contamination following a minimum of 8 passages. CT-1 cultures adapted to growing on a collagen matrix were then similarly passaged into untreated Cellbind cultureware (Corning, NY). CT-1 cultures growing on a collagen substrate or on Cellbind were maintained at 37°C and 8.5% $CO_2$ and subsequently passaged every 7 to 10 days. A bovine fetal fibroblast (BFF) cell line cloned from primary fetal hip muscle fibroblasts was also received as a gift from Dr. Neil Talbot (U.S. Department of Agriculture, Beltsville, MD). BFFs, HEK-293 cells (CRL-1573, ATCC), and NTERA2/D1 embryonal carcinoma cells (CRL-1973, ATCC) were cultured in CT-1 medium at 37 °C in 5% $CO_2$. BFFs and HEK 293 cells were maintained in culture by enzymatically passaging with trypsin-EDTA (25200-056, Invitrogen). NTERA cells were mechanically passaged by cell scraping. In vitro produced (IVP) day 6 bovine morula were purchased from Bomed, Inc. (Madison, WI) and cultured in embryo culture medium (G2 version 3, Vitrolife, Englewood,CO). Blastocysts were pooled and lysed in RNeasy lysis buffer (Qiagen, Valencia, CA) and stored at -80 °C until used for RNA extraction. Mid-gestation bovine cotyledonary and caruncular tissues were collected at a local abattoir (Treuth and Sons, Baltimore, MD), snap frozen, and stored in liquid $N_2$ until needed for RNA extraction.

**Transfections**

The Lipitoid reagent was received as a gift from Dr. Ronald Zuckermann of the Molecular Foundry of the Lawrence Berkeley National Laboratory, Berkeley, CA. Lipitoid-based plasmid and siRNA transfections were based on previously described methods (Schiffmacher and Keefer, 2010). All plasmid transfections were performed in 12-well plates with 0.42 pmol plasmid at the 2/1 Lipitoid/DNA (+/-) charge ratio; the amount of Lipitoid used was adjusted based on net charge of each plasmid. CT-1 colonies

were passaged into Cellbind 12-well plates and grown for 7 to 10 days to achieve 50% confluency. Prior to adding Lipitoid transfection cocktails, CT-1 cells were washed twice with DPBS, once with OPTI-MEM, and then 600 µl of OPTI-MEM were added back to each well. Both treatment-specific volumes of Lipitoid and DNA were separately suspended in OPTI-MEM to a total volume of 100 µl before lightly mixing together. Following a 10 min incubation at RT, the 200 µl cocktails were then added to wells drop by drop, resulting in a total treatment volume of 800 µl. Plates were incubated (37 °C, 8.5% $CO_2$) overnight followed by complete medium aspiration and replacement with CT-1 medium. Cells were incubated for an additional 48 hours before being utilized for experiments.

Short-interfering RNA oligonucleotide-mediated knockdown was performed using custom Stealth RNAi siRNA duplex oligonucleotides (Invitrogen), which possess a net charge of -50. All siRNA transfections were performed at 50 nM final oligonucleotide concentration at a fixed 4/1 Lipitoid/DNA (+/-) charge ratio. CT-1 cells were grown until 50% confluent in Cellbind 12-well plates. Prior to the addition of Lipitoid-double stranded RNA oligonucleotide mixtures, CT-1 colonies were washed twice with DPBS, once with OPTI-MEM and then 420 µl OPTI-MEM were added back to each 12-well. Lipitoid and siRNAs were both diluted in OPTI-MEM to total volumes of 40 µl, gently vortexed and centrifuged before combining together. Lipitoid/RNA suspensions were gently mixed and then incubated for 10 min at RT before being added drop-wise to each well. Treated CT-1 cells were then incubated (37 °C, 8.5% $CO_2$) for 4 hours. Following incubation, FBS was added to each well to achieve 10% final serum concentration. Cells were then incubated overnight, after which the transfection mixture was completely

replaced by CT-1 medium. Lipitoid/ Stealth siRNA transfections were incubated for 48 hours after the addition of FBS to wells. Medium was replaced and colonies were lysed for RNA extraction. HEK 293 cells were transfected using Lipofectamine 2000 (11668-027, Invitrogen) according to manufacturer's instructions and scale. Transfection of HEK 293 cells used for western blotting was performed in 10 $cm^2$ plates, while transfections for immunocytochemistry were performed in 4-well (1.8 $cm^2$) Lab-Tek II CC2 chamber slides (12-565-2, Thermo Scientific Nunc) using the 24-well scale protocol. HEK 293 cells were grown in antibiotic-free CT-1 medium to 90% confluency and transfected overnight at 37 °C and 5% $CO_2$. The following day, medium was aspirated and replaced with CT-1 medium or antibiotic free CT-1 medium supplemented with 0.5-100 ng/ml doxycycline. Cells were incubated in doxycycline for 24 to 48 hours before being processed for experiments.

**RNA extraction and RT-PCR**

Cells were quickly washed twice with DPBS and lysed within the well by directly adding RNeasy lysis buffer (Qiagen). Total cellular RNA (tcRNA) from cell lines and bovine blastocysts were extracted, DNase I treated, and purified using the RNeasy Micro kit (74004, Qiagen). RNA extractions from frozen cotyledonary and caruncular tissues were performed by Dounce homogenization in Trizol (15596-026, Invitrogen). Phases were separated using chloroform and nucleic acids were precipitated from the aqueous phase using isopropanol. Nucleic acids were DNase I treated (18068-015, Invitrogen) and further purified by phenol/ chloroform/ ethanol extraction. All purified RNAs were quantified using the Quant-iT RiboGreen RNA Assay Kit (R11490, Invitrogen) and each sample was diluted to a final concentration of 250 ng per µl in nuclease free $H_2O$. Total cellular RNA (tcRNA) from each sample (1 µg) was reverse transcribed using Superscript III reverse transcriptase (18080-051, Invitrogen). First strand cDNAs were synthesized using oligo ($dT_{20}$) primers according to the manufacturer's instructions. Completed 20 µl cDNA reactions were then diluted 8 fold with nuclease-free water before using as template for RT-PCR. To account for potential genomic DNA contamination amplification during RT-PCR, 1µg of tcRNA from each sample was also used as a template for cDNA synthesis, but reverse transcriptase was not added to these reactions. Sense (forward) and anti-sense (reverse) primers for RT-PCR in Table 3.1 were developed using IDT Real Time PCR software (Integrated DNA Technologies, http://www.idtdna.com/analyzer/ Applications/Oligo Analyzer). PCR reactions consisted of 200 µM dNTP, 1.5 mM $MgCl_2$, 0.25 U recombinant TAQ polymerase (10342-053 Invitrogen), 500 nM each primer, 1 µl diluted CT-1 cDNA, and nuclease free $H_2O$. An

additional reaction for each primer set was performed without cDNA to account for any reagent contamination. Prior to amplification of candidate genes, *GAPDH* RT-PCR was performed on all cDNAs serially diluted two-fold to determine the dilution at which amplification was not saturated after 18 cycles. Saturation assessment was based on ethidium bromide staining following gel electrophoresis. Thermocycler conditions were as follows: initial denaturation at 95 °C for 3 min followed by 35 cycles (18 cycles for *GAPDH*, 25 cycles for *PAG2*) consisting of denaturation at 95 °C for 30 sec, annealing along a gradient of temperatures for 30 sec (Table 3.1), and extension at 72 °C for 1 min. Final extension was performed at 72 °C for 10 min. Equal volumes of all cDNAs and replicates were electrophoresed on 1.5% agarose gels to determine expression, band size, and band specificity. Gel images were captured using the Bio-Rad Chemidoc XRS Imaging System and software. Single DNA bands were gel extracted using a QIAquick Gel Extraction Kit (28704, Qiagen) and sequenced at the University of Maryland Biotechnology Institute: Center for Biosystems Research (College Park, MD).

**Table 3.1.** Primer information for RT-PCR.

| QPCR Primers | Sequence | Amplicon Length (nt) | Genbank Accession # | Annealing Temp. (°C) | Reference |
|---|---|---|---|---|---|
| Bovine OCT4 (FWD) | 5'-AGGAGTCCCAGGACATCAA-3' | 429 | NM_174580.2 | 53 | |
| Bovine OCT4 (REV) | 5'-ACACTCGGACCACGTCTTTC-3' | | | | Ushizawa et al., 2005 |
| Bovine NANOG (FWD) | 5'-CAGTCCTGATTCTTCCACAA-3' | 727 | NM_001025344.1 | 56 | |
| Bovine NANOG (REV) | 5'-TTACAAATCTTCAGGCTGTATGTT-3' | | | | |
| Bovine ERRβ (FWD) | 5'-AGCTGGTGCGCAGGTACAAG-3' | 306 | BC111277.1 | 55 | |
| Bovine ERRβ (REV) | 5'-TCTCCAGGAAGAGTTTGTGC-3' | | | | |
| Bovine SOX2 (FWD) | 5'-CCGCATGTACAACATGATGG-3' | 989 | NM_001105463.1 | 58 | |
| Bovine SOX2 (REV) | 5'-CCTCCAGTTCACCGTCCG-3' | | | | |
| Bovine CDX2 (FWD) | 5'-AGTGAAAACCAGGACGAAAGA-3' | 142 | XM_871005.3 | 61 | |
| Bovine CDX2 (REV) | 5'-CTCTGAGAGCCCCAGCGT-3' | | | | |
| Bovine ETS2 (FWD) | 5'-CAGCAGTTACAGAGGGACAC-3' | 958 | NM_001080214.1 | 54 | |
| Bovine ETS2 (REV) | 5'-GTAGTCCTTGAAGGACATGG-3' | | | | |
| Bovine ELF5 (FWD) | 5'-TTGGACTCAGTGACACACAG-3' | 717 | NM_001024569.1 | 53 | |
| Bovine ELF5 (REV) | 5'-CACTAATCTTCGGTCAACCC-3' | | | | |
| Bovine SOX15 (FWD) | 5'-AACGCGTTCATGGTGTGGAG-3' | 380 | XM_582242.3 | 58 | |
| Bovine SOX15 (REV) | 5'-GAAGAGCCATAACTGCCAGG-3' | | | | |
| Bovine ID2 (FWD) | 5'-AGTCCAGTGAGGTCCGTTAG-3' | 502 | NM_001034231.1 | 54 | |
| Bovine ID2 (REV) | 5'-TCCTCCTCCTTGTGAAATGG-3' | | | | |
| Bovine HAND1 (FWD) | 5'-GCTCTCCAAGATCAAGACTCTGC-3' | 221 | NM_001075761.1 | 58 | |
| Bovine HAND1 (REV) | 5'-CGGTGCGTCCTTTAATCCTCTTC-3' | | | | |
| Bovine GATA6 (FWD) | 5'-CTCTACAGCAAGATGAACGG-3' | 449 | XM_001253596.2 | 53 | |
| Bovine GATA6 (REV) | 5'-TGACCTGAGTACTTGAGCTC-3' | | | | |
| Bovine FGF2 (FWD) | 5'-AGCGGCTGTACTGCAAGAAC-3' | 379 | NM_174056.3 | 55 | |
| Bovine FGF2 (REV) | 5'-CAGCTCTTAGCAGACATTGG-3' | | | | |
| Bovine FGFR2 (FWD) | 5'-GTCATCGTTGAATACGCCTC-3' | 367 | XM_001789706.1 | 54 | |
| Bovine FGFR2 (REV) | 5'-TCTGATGGGTGTACACTCTG-3' | | | | |
| Bovine SSLP1 (FWD) | 5'-TCGGTCCTGCCTTCTGTAAG-3' | 1482 | NM_001105478.1 | 56 | |
| Bovine SSLP1 (REV) | 5'-ATGGCAACTCACATGTGCTC-3' | | | | |
| Bovine IFNT (FWD) | 5'-TGTTACCTGTCTGAGAACCACATGCT-3' | 519 | NM_001168275.1 | 60 | Ushizawa et al., 2005 |
| Bovine IFNT (REV) | 5'-TCAAAGTGAGTTCAGATCTCCACC-3' | | | | Ushizawa et al., 2005 |
| Bovine CSH1 (FWD) | 5'-CCAAGGTCATCAACAGCTGC-3' | 378 | NM_181007.2 | 57 | |
| Bovine CSH1 (REV) | 5'-CCCTGTGTAGGCAGTGGAAC-3' | | | | |
| Bovine PAG9 (FWD) | 5'-TCCTTTTGTACCATGCCAGC-3' | 355 | NM_176620.2 | 56 | Arnold et al., 2006 |
| Bovine PAG9 (REV) | 5'-TGCCCTCCTGCTTGTTTTTG-3' | | | | Arnold et al., 2006 |
| Bovine PAG2 (FWD) | 5'-AGGAAAGAAGCATGAAGTGGCT-3' | 1253 | NM_176614.1 | 54 | Modified from |
| Bovine PAG2 (REV) | 5'-AGCACCAAACACAATTCACC-3' | | | | Xie et al., 1991, 1997 |
| Bovine GAPDH (FWD) | 5'-TGTTCCAGTATGATTCCACCC-3' | 841 | NM_001034034.1 | 55 | |
| Bovine GAPDH (REV) | 5'-TCCACCACCCTGTTGCTGTA-3' | | | | |
| Bovine LEF1 (FWD) | 5'-GACGAGATGATCCCCTTCAA-3' | 133 | XM_615475.4 | 58 | |
| Bovine LEF1 (REV) | 5'-GGATGATTTCGGATTCGTTG-3' | | | | |
| Bovine EOMES (FWD) | 5'-GCAGAGGCTCTTATCAGA-3' | 265 | XM_001251929.2 | 55 | Hall et al., 2005 |
| Bovine EOMES (REV) | 5'-GCGTTAATGTCCTCACACTT-3' | | | | Hall et al., 2005 |

**Quantitative RT-PCR**

Purified RNA was quantified using the Quant-iT RiboGreen RNA Assay Kit (CR11490, Invitrogen) and each sample was diluted to a final concentration of 250 ng per µl. One µg of tcRNA from each sample was then used to synthesize first strand cDNA using the AffinityScript Multiple Temperature cDNA Synthesis Kit (200436, Stratagene, La Jolla, CA). First strand cDNAs were synthesized using oligo (dT) primers according to the manufacturer's instructions with one modification: for the synthesis step, cDNA reactions were incubated at 55°C for 1 hour. Completed 20 µl cDNA reactions were then diluted 10-fold with nuclease-free water before using as template for quantitative RT-PCR (qRT-PCR). To account for potential genomic DNA contamination amplification during RT-PCR, 1µg of pooled tcRNAs (3 to 4 combined equivalents per treatment) were also used as template for cDNA synthesis, but reverse transcriptase was not added to these reactions.

Sense and antisense primers were developed using IDT Real Time PCR software (Integrated DNA Technologies, http://www.idtdna.com/Scitools/ Applications /RealTimePCR) and designed to span across exon/ intron splice junctions (Table 3.2). All primers were based on published sequence (http://www.ncbi.nlm.nih.gov/pubmed/). *OCT4* primers were based on published sequence (Accession no. NM_174580.1) and verified to specifically amplify *OCT4* mRNA and not transcribed *OCT4* pseudogene transcripts. Primers designed to *gamma–Actin* (*ACTG1*, Accession no. NM_001033618.1) were utilized for qRT-PCR amplification normalization.

All qRT-PCR reactions (except amplification of *OCT4*) consisted of 7.5 µl 2X iQ SYBR Green Super mix (170-8882, Bio-Rad), 0.6 µl of both sense and antisense 10 µM stock primers, 5.3 µl nuclease-fee $H_2O$, and 1 µl diluted cDNA or 'no RT' pooled

control. *OCT4* qRT-PCR reactions were adjusted to contain 5 µl of cDNA or 'no RT' pooled control. 2X iQ SYBR green mix contains fluorescein as an internal reference to normalize well-to-well optical variation. qRT-PCR was performed using the MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad) programmed with the following steps: initial denaturation at 95°C for 3 min and 40 repeated steps consisting of 15 seconds at 95°C and extension at 60°C for 1 min. After 40 cycles, a dissociation (melt) curve was performed. Each PCR reaction was performed on 3 independently treated cDNA replicates and each cDNA was amplified in duplicate. All cDNA amplifications produced single peaks in the melt curve analysis, and PCR products were further analyzed by gel electrophoresis for correct size. Single amplicon bands were purified from agarose gels using a MinElute PCR Purification Kit (28004, Qiagen) and sequenced at the University of Maryland Biotechnology Institute: Center for Biosystems Research (College Park, MD). The data output of target gene reaction Ct (threshold cycle) values, representing fixed threshold crossing points where all samples are undergoing logarithmic amplification, were first normalized to *ACTG1* Ct values using the formula: $\Delta Ct = (Ct_{\text{'no RT'}} - Ct_{\text{sample}})_{\text{target}} - (Ct_{\text{'no RT'}} - Ct_{\text{sample}})_{ACTG1}$. ΔCt values were then transformed ($2^{\Delta Ct}$) and data were represented as relative abundance of target mRNA normalized to *ACTG1* levels.

**Table 3.2.** Primer information for quantitative RT-PCR (qRT-PCR).

| QPCR Primers | Sequence (5' - 3') | exon | Amplicon Length (nt) | Genbank Accession # |
|---|---|---|---|---|
| OCT4 (FWD) | CAAATTAGCCACATCGCC | 4 | 126 | NM_174580.2 |
| OCT4 (REV) | AGCCTCAAAATCCTCACG | 5 | | |
| 3' UTR OCT4 (FWD) | CTTTCCCTCGGTGTCTG | 5 | 142 | NM_174580.2 |
| 3' UTR OCT4 (REV) | ACTTAATCCCAAAGGCCTG | 5 | | |
| CDX2 (FWD) | AGTCGCTATATCACCATCC | 2 | 104 | XM_871005.3 |
| CDX2 (REV) | CTTTCCTTTGCTCTGCG | 3 | | |
| CDX2 (FWD) | TTAAACCCTACTGTCACCC | 3 | 139 | XM_871005.3 |
| 3' UTR CDX2 (REV) | AGGTCAGCTGGTAAACATTAG | 3 | | |
| IFNT (FWD) * [#] | GATCCTTCTGGAGCTGGYTG | 1 | 100 | NM_001168275.1 |
| IFNT (REV) * | GCCCGAATGAACAGACTCYC | 1 | | |
| ELF5 (FWD) | GCCTGTATCTCTGACTGTG | 1 | 147 | NM_001024569.1 |
| ELF5 (REV) | GGGTAATACTCTTCATTGCTG | 2 | | |
| MASH2 (FWD) | GCTGCTCGACTTCTCCAG | 1 | 126 | NM_001040607.1 |
| MASH2 (REV) | CGGAACGAGGAACACGG | 2 | | |
| HAND1 (FWD) | ACTGAAGAAGGCGGATG | 1 | 150 | NM_001075761.1 |
| HAND1 (REV) | TGGTTTAACTCCAGCGC | 2 | | |
| ID2 (FWD) | CCCTTCTGAGTTAATGTCAAA | 2 | 150 | NM_001034231.1 |
| ID2 (REV) | CTCCTTGTGAAATGGTTGAA | 3 | | |
| SOX15 (FWD) | TGGATATGCAGCCAACC | 1 | 150 | XM_582242.3 |
| SOX15 (REV) | GAGCCTTGGGTTACTCTG | 2 | | |
| GATA2 (FWD) ** | GAGGACTGTAAGCGTAAAGG | 6 | 140 | XM_583307.3 |
| GATA2 (REV)** | AAGAACCAAGTCTCCCCAT | 6 | | |
| GATA3 (FWD)** | ATGAAACCGAAACCCGATG | 5 | 185 | NM_001076804.1 |
| GATA3 (REV)** | TTCACAGCACTAGAGAGACC | 5 | | |
| GATA6 (FWD) | AAGATGCTGACCAGACATCT | 7 | 206 | XM_001253596.2 |
| GATA6 (REV) | AGACCAGCTGCCTGGAAGT | 7 | | |
| ETS2 (FWD) | CCGACCATGTCCTTCAAG | 6 | 150 | NM_001080214.1 |
| ETS2 (REV) | CTTGTCGGAGAGCAATTC | 7+8 | | |
| FGFR2 (FWD) | CTCACACTCACAACCAATG | 18 | 147 | XM_001789706.1 |
| FGFR2 (REV) | AAGGCAGGGTTCGTAAG | 19 | | |
| ETF1 (FWD) | GAGCTACGTTGGAAATTGTC | 9 | 106 | NM_001076254.1 |
| ETF1(REV) | CCCTGGAAATCTACTCGG | 10 | | |
| ACTG1 (FWD) | TTGCTGACAGGATGCAGAAG | 4 | 145 | NM_001033618.1 |
| ACTG1 (REV) | TGATCCACATCTGCTGGAAG | 5 | | |

* Primers from Cooke et al., 2009. ** Primers from Bai et al., 2009

[#] Y is either a C or T nucleotide in IFNT (FWD)

**Short interfering RNA oligonucleotide and plasmid design**

Custom Stealth bovine *OCT4* (*POU5F1*) 25-nucleotide duplex siRNA was designed using Invitrogen's BLOCK-IT RNAi designer. A bovine *CDX2* 25-nucleotide duplex siRNA was modified from a previously reported functional 19-nucleotide siRNA sequence (siRNA #2; Sakurai et al., 2009) using Invitrogen's siRNA to Stealth RNAi siRNA converter. A nonspecific scrambled 25-nucleotide duplex siRNA was also designed to serve as a negative RNA interference control (Schiffmacher and Keefer, 2010). Plasmids pTHE.HA:OCT4 and pTHE.MYC:CDX2 were created using the tetracycline induction system single vector pTHE (Sup. Fig. 2, Jiang et al., 2001). The pTHE (Addgene plasmid 12512) is a single vector containing a repressor cassette that expresses tetracycline repressor (Tetr) protein fused to the mSIN3-binding domain (SID) of MAD protein. This fusion protein not only represses transcription at the tetracycline responsive promoter but conveys extra repression control as the mSIN3-binding domain recruits mSIN3 and histone deacetylases. Primer sets designed bind the 5' and 3' UTRs of the full length bovine *OCT4* open reading frame (ORF) and full length *CDX2* ORF (Table 3.3) were used to amplify the coding sequences from CT-1 cDNA by RT-PCR.

The *OCT4* PCR reaction consisted of 4 µl 5X HF iPROOF buffer (Bio-Rad), 0.2 µl iPROOF polymerase (172-5330, Bio-Rad, Hercules, CA), 1.6 µl 2.5 mM dNTP, 11.2 µl nuclease-free $H_2O$, 1 µl each primer, and 1 µl diluted CT-1 cDNA. This reaction was amplified using the following parameters: initial denaturation at 98°C for 30 seconds (sec), 35 cycles at 98 °C for 10 sec, annealing at 60 °C for 30 sec, and extension at 72 °C for 45 sec, and lastly, a final extension step at 72 °C for 10 min. The *CDX2* ORF PCR reaction consisted of 4 µl 5X GC iPROOF buffer (Bio-Rad), 0.2 µl iPROOF polymerase

(172-5330, Bio-Rad, Hercules, CA), 1.2 µl 2.5 mM dNTP, 11 µl nuclease-free $H_2O$, 1 µl each primer, 0.6 µl DMSO, and 1 µl diluted CT-1 cDNA. This RT-PCR reaction was initially denatured for 30 sec at 98 °C. The next 10 cycles consisted of a denaturation step at 98 °C for 10 sec, annealing for 30 sec, and extension 45 sec. First round of annealing was performed at 73 °C, and at each subsequent cycle the annealing temperature was decreased by 1 °C. The remaining 25 cycles were performed as above at a fixed annealing temperature of 63 °C, before finishing with a final extension period at 72 °C for 10 min. Single bands for OCT4 and CDX2 amplicons were purified by agarose gel electrophoresis (NucleoSpin Extract II kit, 740609.50, Machery-Nagel, Bethlehem, PA). These RT-PCR products were then used as templates for amplifying an *OCT4* ORF fused to a 5' hemagglutinin (*HA*) epitope sequence and a *CDX2* ORF fused to a *MYC* epitope sequence. Both new amplicons were also 5' flanked by a Stu I site prior to the translational start sight and 3' flanked by a Not I site downstream of the ORF stop codon. The forward primer containing the *HA* epitope coded the amino acid sequence MYPYDVPDYAAG that is fused in frame to the *OCT4* ORF lacking the start codon. A similar strategy was used for tagging the *CDX2* sequence with a *MYC* tag that translates into the N terminal amino acid epitope EQKLISEEDL. Both reactions were amplified using the following parameters: initial denaturation at 98 °C for 30 seconds (sec), 35 cycles at 98 °C for 10 sec, annealing at 61 °C for 30 sec, and extension at 72 °C for 45 sec, and lastly, a final extension step at 72 °C for 10 min. Single bands for each amplicon were purified and double digested overnight at 37 °C with 20 U each Stu I and Not I and agarose gel purified. Ten µg of the pTHE vector was also double digested overnight at 37 °C with 20 U each Stu I and Not I. Linearized vector was treated with 1 U calf intestinal

alkaline phosphatase (18009-027, Invitrogen) at 50 °C for 5 min before purification. Ten µl ligations were set up at 1:3 (vector/insert) molar ratios and incubated at 16 °C overnight with T4 DNA ligase (M0202S, New England Biolabs, Ipswich, MA). The ligation reaction was heat inactivated at 65 °C for 20 min and 2 µl were chemically transformed into DH5α *E. coli* (18258-012, Invitrogen) and plated onto LB plates containing 75 µg per ml carbenicillin (10177-012, Invitrogen). True positive transformed colonies were verified by colony PCR and restriction enzyme mapping before mini-preparations of plasmid were produced (Nucleospin, 740588.10, Machery-Nagel). Plasmids were then sequenced at the University of Maryland Biotechnology Institute: Center for Biosystems Research (College Park, MD).

**Table 3.3.** Primer information for *OCT4* ORF, *CDX2* ORF and *OCT4* retrocopy.

| PCR Primers | Sequence | Amplicon Length (nt) | Genbank Accession # | Primer Length (nt) |
|---|---|---|---|---|
| Bovine OCT4 (FWD) | 5'-GGTGTTGAGCAGTCTCTAGG-3' | 1162 | NM_174580.1 | 20 |
| Bovine OCT4 (REV) | 5'-GATCAGGCACCTCAGTTTGC-3' | | | 20 |
| Bovine POU5F1rs1 (FWD) | 5'-GGGTGTGAAGTGGGTTTGTG-3' | 1276 | NW_001495368 | 20 |
| Bovine POU5F1rs1 (REV)** | 5'-CCATTCATCTGCTGATGGAC-3' | | (3804496-3803222) | 20 |
| Bovine CDX2 (FWD) | 5'-TCGCCACCATGTACGTGAGC-3' | 955 | XM_871005.3 | 20 |
| Bovine CDX2 (REV) | 5'-GAATTGCTACTGCAGGCCGC-3' | | | 20 |

** This primer hybridizes to highly redundant sequence (bovine LINE L1).

**Protein extraction and Western blot**

Whole cell lysates (WCL) for western blotting were prepared in the following manner. CT-1 cells were cultured to 80% confluency in a T75 Cellbind flask. Transfected and untransfected HEK 293 cells grown to 80% confluency in 10 $cm^2$ tissue culture plates were washed quickly twice with 10 ml ice cold DPBS (14190-136, Invitrogen) and scraped from the surface. Cell suspensions were transferred to 15 ml conical tubes on ice and centrifuged at 1,850 x g at 4 °C for 3 min. Supernatant was removed and packed cell volume (PCV) was estimated. Pellets were resuspended in 300-500 µl ice cold lysis buffer ( 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.5% Triton X-100) supplemented immediately prior to use with 1 mM phenylmethane-sulfonylfluoride (PMSF) and 1X Protease Inhibitor cocktail Set 3 (539134, EMD chemicals, Darmstadt, Germany). PMSF was solubilized in 100% ethanol to 100 mM stock concentration just before adding to all lysis buffers. Cell suspensions were transferred to 1.5 ml tubes and then subjected to 3 freeze-thaw cycles consisting of an incubation at -80 °C until suspensions were completely frozen (5-10 min), followed by incubation at RT until melted (but no longer). Lysed cells were then centrifuged at 4 °C at 20,000 x g for 15 min, and supernatants were transferred to new 1.5 ml tubes on ice. Nuclear-cytoplasmic fractionation was performed in the following manner. PCVs were obtained as previously noted in WCL preparation. PCV were resuspended in 5X PCV ice cold cytoplasmic lysis buffer (10 mM HEPES, pH 7.4, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol (DTT), 0.1% NP-40, 300 mM sucrose) supplemented immediately prior to use with 1 mM PMSF, 1X Protease Inhibitor cocktail Set 3, 1 mM $Na_3VO_4$, 10 mM NaF, and 20 mM β-glycerophospate. Cell suspensions were centrifuged at 1,850 x g for 3 min at 4 °C. The supernatant was

removed and the PCV was resuspended in 1.5X PCV of complete cytoplasmic lysis buffer. Suspensions were incubated on ice for 10 min to allow cell swelling and then transferred to a chilled Dounce homogenizer. Cells were homogenized slowly but forcefully for 20-50 strokes with a 'loose' pestle, taking care not to create bubbles. Cell lysis was assessed by evaluating a small aliquot mixed with Trypan blue on a hemocytometer. Homogenization was stopped when approximately 90% of cells were lysed. Suspensions were then transferred to a new chilled 1.5 ml tube and centrifuged at 3,300 x g for 15 min at 4 °C. The cytoplasmic fraction was removed and transferred to a new 1.5 ml tube and stored on ice. The packed nuclear volume (PNV) was estimated and resuspended in 0.75X PNV nuclear lysis buffer (50 mM HEPES, pH 7.4, 250 mM KCl, 0.1 mM EDTA, 0.5 mM dithiothreitol (DTT), 0.1% NP-40, 0.1% glycerol) supplemented immediately prior to use with 1mM PMSF, 1X Protease Inhibitor cocktail Set 3, 1 mM $Na_3VO_4$, and 10 mM NaF. Nuclear fraction suspensions were rocked for 30 min at 4 °C. Nuclear and cytoplasmic fractions were then centrifuged at 20,000 x g for 15' at 4 °C. Supernatants were removed and transferred to new tubes. Protein concentrations of lysates were calculated by bicinchoninic acid (BCA) protein assay according to the manufacturer's instructions (23225, Thermo Scientific Pierce, Waltham, MA). Protein samples were then prepared for separation by Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). All buffers and both SDS-PAGE and western blot protocols were performed according to Bio-Rad Mini-PROTEAN 3 Cell instructions (165-2954, Bio-Rad). Lysates (75 µg) were mixed with Laemmli sample buffer and boiled at 100 °C for 3 min and centrifuged at 16,000 x g for 5 min to further remove any insoluble particulate. Supernatants were loaded into 4% stacking-10% resolving

polyacrylamide gels. Electrophoresis was performed at 150 V until the dye front completely ran off the gels (about 1 hr, 15 min). Protein transfer was performed onto nitrocellulose membranes (162-0115, Bio-Rad) in chilled transfer buffer (80% running buffer, 20% methanol, 0.1% SDS) at 75 V for 45 min. Blots were incubated at RT in 5% blotto (nonfat dry milk reconstituted in Tris-buffered saline with 0.05% Tween 20 (TBST)) on a rocker for 1 hr. Blots were incubated in one of the following antibodies diluted in 5% blotto and incubated overnight at 4 °C on a rocker: polyclonal rabbit anti-human OCT4 (1:250, sc-9081, Santa Cruz Biotechnology, Santa Cruz, CA), monoclonal anti-human OCT4 (1:250, sc-5279, Santa Cruz) monoclonal anti-HA tag (1:1250, 05-904, Millipore, Temecula, CA), rabbit anti-human Lamin A/C (1:250, sc-20861, Santa Cruz) or rabbit anti-human Actin (1:250, sc-1615, Santa Cruz). Blots were then thoroughly washed in TBST and incubated on a rocker for 1 hr at RT in peroxidase-conjugated donkey anti-rabbit IgG (711-035-152, Jackson ImmunoResearch, West Grove, PA) or donkey anti-mouse IgG (715-035-150, Jackson ImmunoResearch) diluted 20,000-fold in TBST. Blots were thoroughly washed again in TBST, developed using an Immun-Star WesternC Chemiluminescent Kit (170-5070, Bio-Rad), and Chemidoc XRS Imaging System.

**Immunocytochemistry**

CT-1 cells were grown in 6-well Cellbind plates until colonies reached over 2 mm in diameter. HEK 293 and NTERA cells used for immunocytochemistry were grown in 4-well Lab-Tek II CC2 chamber slides (12-565-2, Thermo Scientific Nunc). Cells were washed twice in DPBS with divalent cations (14040-133, Invitrogen) and fixed in 4% formaldehyde for 30 min at RT. Fixed cells were washed 3 times for 5 min in TBST (pH

7.4, 20mM Tris-HCl, 0.15 M NaCl, 0.05% Tween 20). Cell membranes were then permeabilized in permeabilization buffer (TBST with an additional 0.05% Tween 20 and 0.2% Triton X-100) for 10 min on a rocker at RT. Permeabilization buffer was removed by washing twice in TBST for 5 min. Cells were then blocked in 10% normal donkey serum (017-000-001, Jackson ImmunoResearch) in TBST while rocking at RT for 30 min. Cells were incubated overnight at 4 °C on a rocker in any one of the following antibodies diluted in TBST: rabbit anti-human OCT4 (1:250, sc-9081, Santa Cruz Biotechnology), monoclonal anti-HA tag (1:1250, 05-904, Millipore) and monoclonal anti-human CDX2 (1:200, MU392A-UC, Biogenex, San Ramon, CA). The following day, cells were washed three times in TBST and incubated for 30 min on a rocker at RT in one of the following secondary antibodies (2.5 µg/ ml in TBST): Dylight 549 conjugated donkey anti-rabbit IgG (711-505-152, Jackson ImmunoResearch), Dylight 549 conjugated donkey anti-mouse IgG (715-505-150, Jackson ImmunoResearch), or Alexa Fluor 488 donkey anti-rabbit IgG (A-21206, Invitrogen). Excess secondary antibody was removed by washing 3 times in TBST for 10 min on a rocker at RT. Cells were incubated in 10 µg per ml Hoechst 33342 in DPBS for 10 min. Wells were then rinsed twice in DPBS. Fixed colonies were then teased off of well bottoms using a fine paint brush (3/0, IMEX) and delicately applied to slides so that colonies did not have folded edges or tears. Following a brief drying period, slides were mounted with Fluoromount G (0100-01, Southern Biotech, Birmingham, AL) and cover slipped. Images of finished slides were captured using a Zeiss Axio Observer Z1 Inverted Microscope with Axiovision software (Carl Ziess Inc, Thornwood, New York). Secondary antibody only negative controls were also performed. Double label immunocytochemistry was

performed as above except that colonies were simultaneously incubated in two primary antibodies and then in two secondary antibodies. Negative controls were performed in a similar manner.

**Immunoprecipitation**

Lysates were prepared as previously described. Samples were adjusted to 500 µl of 750 µg total protein and mixed with 1 µl (60 µg) of normal rabbit serum (011-000-001, Jackson ImmunoResearch) and 60 µl of 50% slurry protein A beads (20333, Thermo Scientific Pierce). Immune complexes were then rocked for 1 hr at 4 °C, and centrifuged at 8000 x g for 15 min at 4 °C. Supernatants were transferred to new 1.5 ml tubes, mixed with 1 µg of either antibody polyclonal rabbit anti-human OCT4 (sc-9081, Santa Cruz) or monoclonal anti HA tag (05-904, Millipore) and 40 µl of 50% slurry protein A beads and rocked overnight at 4 °C. The following day, immune complexes were centrifuged at 16,000 x g for 2 min at 4 °C. The supernatant was removed and the immune complexes were washed 5X in the following manner: 500 µl of whole cell lysate buffer was added to the pellet, the bead suspension was lightly vortexed and then centrifuged at 16,000 x g for 2 min at 4 °C. Following the wash steps, the supernatant was removed and 40 µl of Laemmli sample buffer was added. Bead suspensions were vortexed and boiled at 100 °C to dissociate the immune complexes. Suspensions were centrifuged a final time at 16,000 x g for 2 min at 4 °C, and the eluates were transferred to new tubes to be run by SDS-PAGE for western blotting.

**Statistical Analysis**

Levels of gene expression were calculated as mean ratios of treatment $\Delta C_t$ values (described in Methods and Materials) relative to untransfected control $\Delta C_t$ values of at least 3 independent experimental replicates. Presented data are the means and SE of these ratios. All data were first log transformed to meet assumptions for variance heterogeneity and normality. qRT-PCR data with multiple treatments (untransfected, target, control) was analyzed by ANOVA using the PROC MIXED model in SAS statistical software (SAS Institute, Cary, NC). Plasmid type or siRNA type were classified as independent variables. If a significant difference in treatments was detected ($P<0.05$), individual treatments were compared to the untransfected control using the PDIFF procedure. Statistical analyses of overexpression-qRT-PCR experiments comparing transcript levels between treatment and control (empty vector transfection) were performed using T test procedure for tests of independent samples. When the probability for the test of equality of pooled variances was significant ($p< 0.05$), then the Satterthwaite approximate T test for unequal variances was used. For all statistical analysis, differences were considered significant at $P<0.05$.

## Results

### CT-1 cells and ovoid stage embryos express common trophoblast regulators and markers.

To determine if gene expression patterns in CT-1 cells and the elongating bovine trophoblast were similar, a profile of candidate genes was tested (Fig. 3.1). Mid-gestation cotyledonary tissue and pooled IVP bovine blastocysts served as positive controls when appropriate. There were no gross differences in relative expression between replicates of any amplified CT-1 cDNAs grown on collagen or Cellbind for any transcripts tested. Assessment of pluripotency-associated gene expression showed that *OCT4* transcripts, but not *NANOG* and *SOX2* transcripts, were detected in CT-1 cells while all 3 mRNAs were expressed in the blastocyst control. Chromatogram analysis of the *OCT4* sequence revealed 3 single nucleotide polymorphisms (SNPs) that indicated that both *OCT4* (Genbank accession no. NM_174580) and transcribed *OCT4* retrocopy cDNAs (Genbank accession no. XM_001789212.1) were co-amplified. RT-PCR performed with *OCT4* mRNA specific primers and primers unique to the retrocopy sequence located on chromosome 7 confirmed amplification of 2 separate cDNAs (Fig. 3.2). However, the amplified *OCT4* retrocopy sequence obtained using primers flanking the chromosome 7 *OCT4* retrocopy was not identical to the predicted *OCT4* retrocopy transcript sequence (XM_001789212.1), suggesting the prediction is incorrect. While the retrocopy transcript was expressed in CT-1 cells and placental tissues, full length *OCT4* transcripts were only amplified from CT-1 cells (Fig. 3.2). TE lineage transcription factors *CDX2* and Estrogen-Related Receptor Beta (*ERRB*) were both expressed as well. However, *LEF1*, which facilitates WNT3a-mediated differentiation of mouse embryonic stems cells into trophoblast lineages (He et al., 2008) and the TE lineage regulator *EOMES* could not be

detected (data not shown). Other trophoblast specific transcription factors that regulate trophoblast stem cell maintenance and differentiation in mice were also expressed in CT-1 cells. This list includes *ELF5*, *ETS2*, *HAND1*, *SOX15*, and *ID2*. *FGF2* and *FGFR2*, components of FGF signal transduction pathways involved in TE development, were detected as well.

Using previously established primers (Arnold et al., 2006) *PAG-9* was amplified from mid-gestation cotyledon cDNA but not from CT-1 cDNA. Abridged primers designed to conserved sequences of the *PAG* II family were utilized to amplify any expressed *PAG* II family members (Garbayo et al., 2008). PCR products were directly sequenced and *PAG2* was identified in CT-1 cells. Similar to secreted seminal-vesicle Ly-6 protein 1 (*SSLP1*), initially identified as the novel transcript (c12) expressed in both mural and polar trophoblast at the ovoid stage, was also expressed in CT-1 cells (Degrelle et al., 2005).

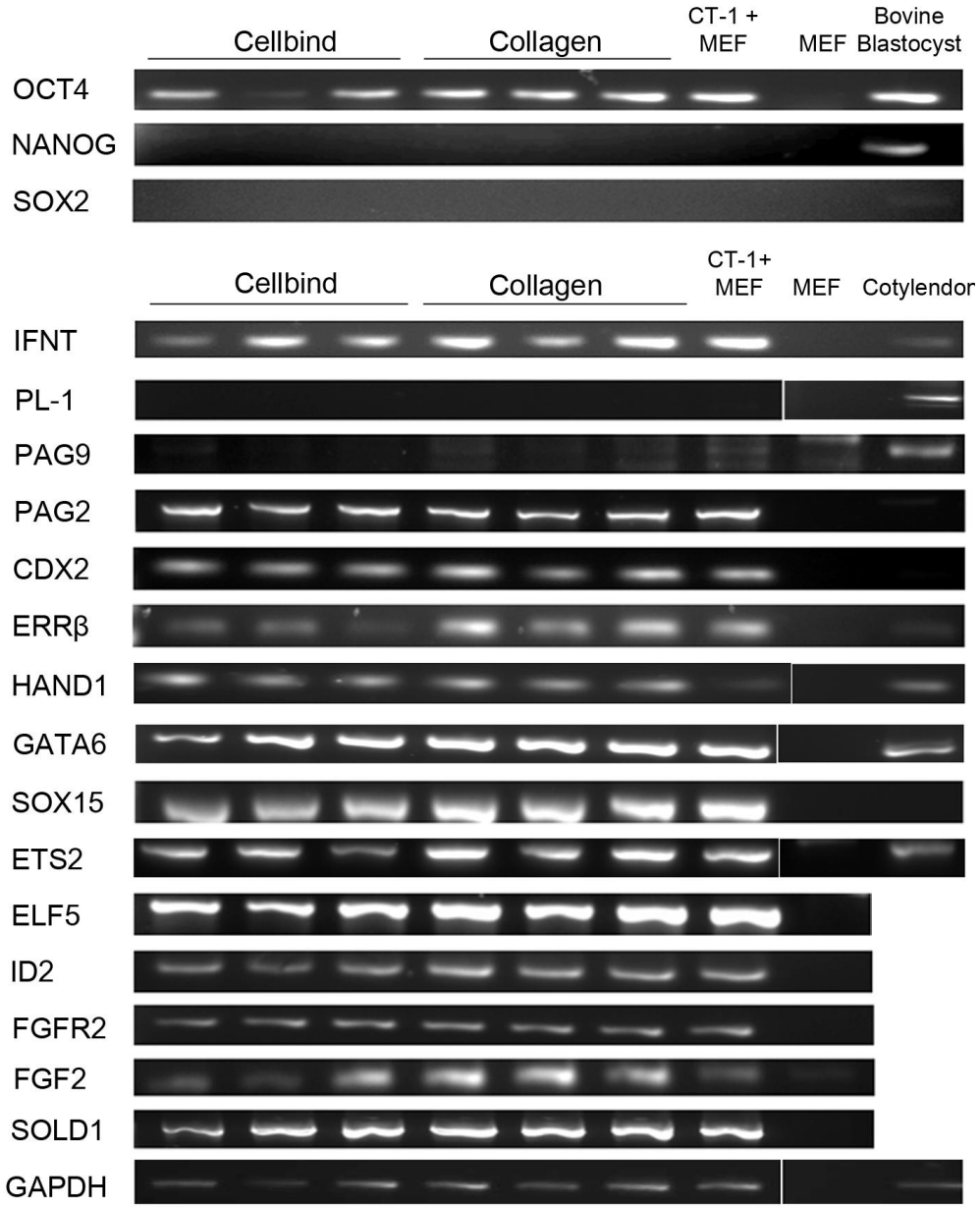



**Figure 3.1.** Transcription profile of selected genes in the CT-1 cell line that are expressed during bovine or mouse trophoblast development. Three replicates of cDNAs from CT-1 cells grown on a collagen substrate or substrate-free Cellbind were tested. cDNAs from a CT-1/MEF co-culture and a MEF culture served as controls for MEF cDNA contamination. cDNAs from mid gestation cotyledon and pooled IVP blastocysts served as positive controls where appropriate.

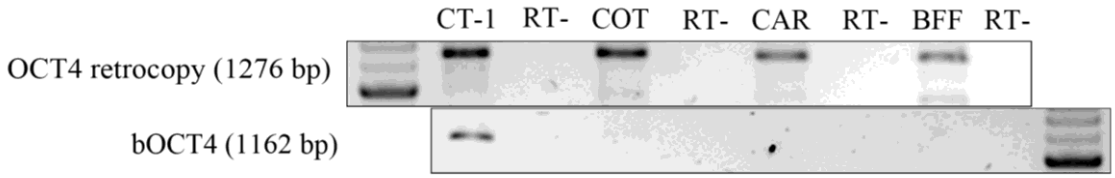



**Figure 3.2.** RT-PCR for bovine OCT4 open reading frame and transcribed OCT4 retrocopy. Primers are unique to 5' and 3' regions of the bovine OCT4 retrocopy located on chromosome 7. COT is cotyledon cDNA, CAR is caruncle cDNA and BFF is bovine fetal fibroblast cDNA. OCT4 is only expressed in CT-1 cells whereas the retrocopy was amplified from all cDNAs tested. No genomic DNA contamination was detected.

**OCT4 protein is localized to nuclei and cytoplasm in CT-1 cells.**

CT-1 cells express *OCT4* mRNA that is 88% conserved to human *OCT4A*, the variant associated with pluripotency maintenance in human ESC (Cauffman et al., 2006). The predicted bovine OCT4 amino acid sequence is 94% similar to the human OCT4A isoform and 89% similar between the first 134 amino acids that contain the N-terminal transaction domain specific to human OCT4A and not found in human OCT4B. As the OCT4B isoform has been found to be cytoplasmic even though it has the conserved nuclear localization signal, it was necessary to determine if the bovine OCT4 is indeed translated and imported to the nucleus similarly to human OCT4A. Western blot analysis using polyclonal and monoclonal antibodies raised against the human N-terminal 134 amino acid OCT4 epitope recognized a single band within whole CT-1 protein lysates that was the estimated size of bovine OCT4.

Further fractionation of CT-1 protein lysates into nuclear and cytoplasmic fractions (Fig. 3.3) as well as immunocytochemistry analysis (Fig. 3.4a) demonstrated that OCT4 is localized in nuclei of CT-1 cells. Further analysis of OCT4 localization revealed that OCT4 was enriched but not entirely restricted to the nucleus. Many cells within the CT-1 colony exhibited increased OCT4 perinuclear or cytoplasmic localization (Fig. 3.4b). OCT4 restriction to the nucleus is also less defined in binucleate cells. However, recombinant bovine OCT4 expressed in transfected HEK 293 cells is nuclear and similar to human OCT4 expressed in the human embryonic carcinoma NTERA cell line (Fig. 3.4c and 3.4d).

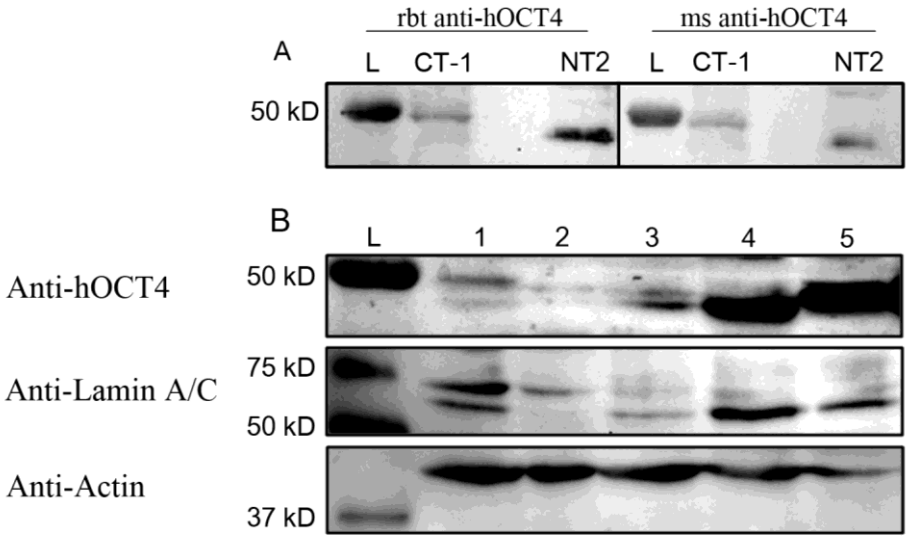



**Figure 3.3.** OCT4 protein is expressed in CT-1 cells and is enriched in the nuclear fraction. A) Polyclonal and monoclonal anti human OCT4 antibodies both detect bovine OCT4 in CT-1 whole cell lysate (75 µg) and NTERA (NT2) whole cell lysate (18 µg). B) CT-1 protein lysates were fractionated into nuclear (Lane 1) and cytoplasmic (Lane 2) protein lysates. CT-1 whole cell lysate (Lane 3), NTERA whole cell lysate (Lane 4) and transfected HEK 293 cells expressing HA tagged bovine OCT4 (Lane 5) were run as positive controls. 75 µg of lysate was run for each sample. Polyclonal anti human OCT4 detects enrichment of OCT4 in nuclear fraction. Immunoblots with anti Lamin A/C and Actin antibodies were performed as nuclear fraction enrichment and loading controls respectively.

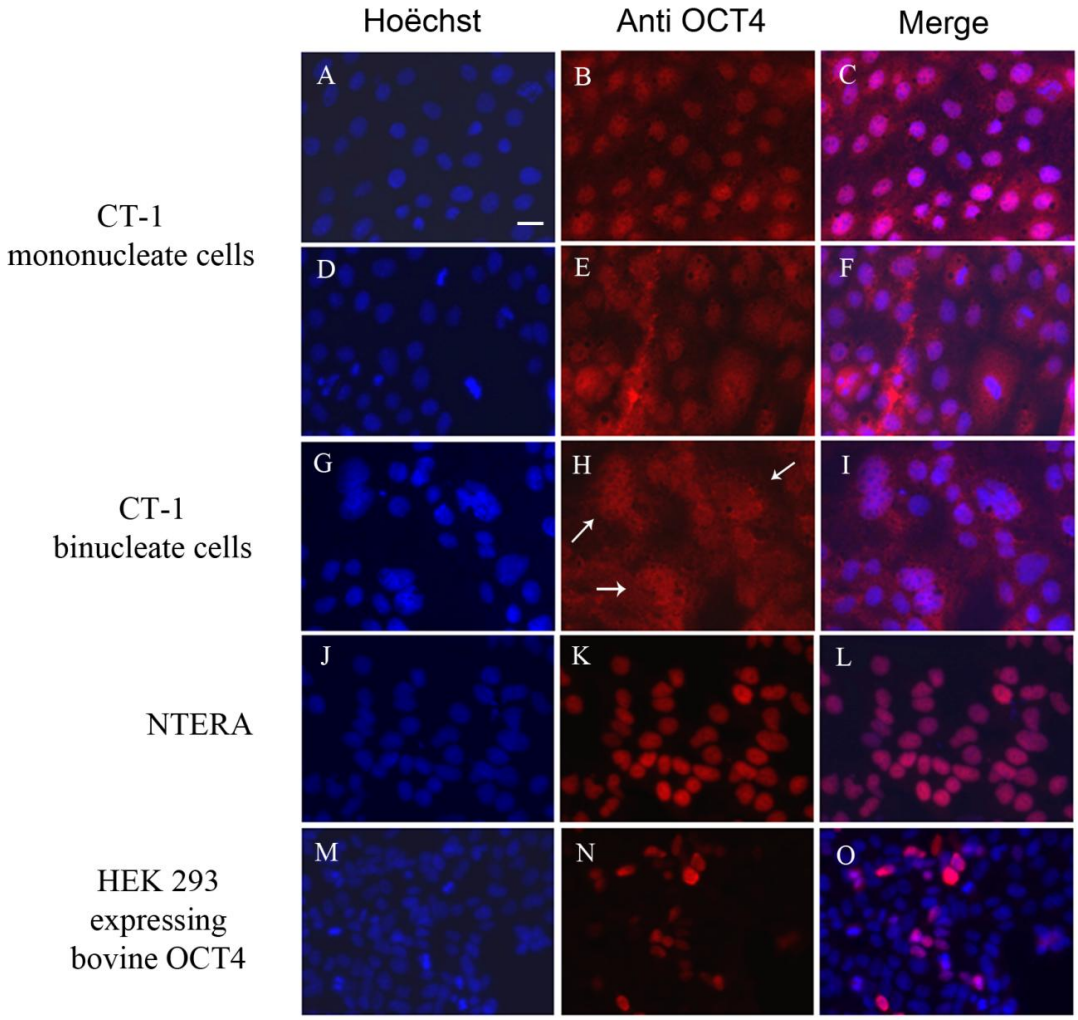



**Figure 3.4.** OCT4 protein typically either exhibits well defined nuclear localization in some mononucleate CT-1 cells (A-C) or a less defined nuclear/cytoplasmic distribution (D-F). Binucleate cells (white arrows) also exhibit ambiguous OCT4 localization (G-I). NTERA (J-L) and transfected HEK 293 cells expressing HA tagged OCT4 (M-O) display strict nuclear localization. Mononucleate cells undergoing mitosis exhibit cellular distribution due to nuclear membrane breakdown. Bar = 15 µm.

Since OCT4 and CDX2 are involved in a reciprocal inhibitory feedback loop that contributes to TE and ICM lineage specification and segregation during mouse development, localization of OCT4 protein with CDX2 was assessed by double label immunocytochemistry of CT-1 cells (Fig. 3.5). CDX2 immunofluorescence varied cell to cell, but all mononucleate and binucleate cells exhibited CDX2 and OCT4 co-localization.

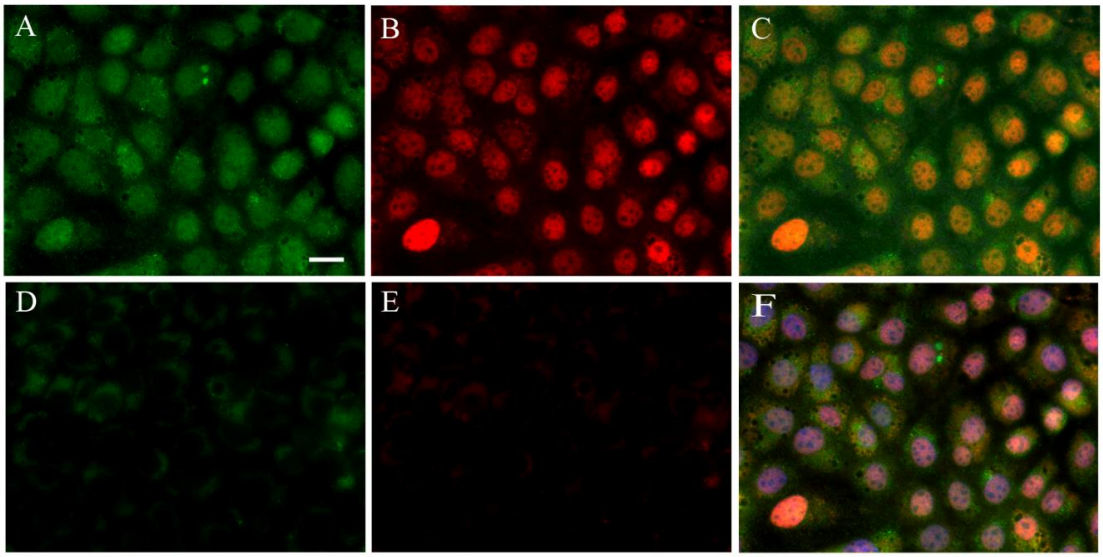



**Figure 3.5.** Many mononucleate cells of the CT-1 monolayer exhibit nuclear and perinuclear OCT4 localization (A) and nuclear CDX2 localization (B). CDX2 and OCT4 co-localize in CT-1 nuclei (C). Immunocytochemistry was also conducted on CT-1 cells with secondary antibodies only (D-E). Background immunofluorescence was observed to be higher than background fluorescence in single label immunocytochemistry with either conjugate (images not shown). OCT4 and CDX2 images were merged with Hoechst staining for nuclear reference (F). Bar = 10 µm.

**CDX2 regulates expression levels of other trophoblast lineage regulators.**

Manipulations of OCT4 or CDX2 levels in mouse ESC can disrupt the OCT4 and CDX2 equilibrium maintained by the reciprocal inhibitory feedback loop and permit induction of pro-differentiation gene networks (Niwa et al., 2005). To determine whether OCT4 and CDX2 function in a similar equilibrium in CT-1 cells, OCT4 and CDX2 levels were modulated and their effects on other trophoblast specific expression levels were assessed. Optimization of Lipitoid-based siRNA transfection in CT-1 cells and *OCT4* and *CDX*2 siRNA oligonucleotide validations were described previously (Schiffmacher and Keefer, 2010). Relative expression levels of *CDX2* were significantly downregulated by 62% from untransfected and nonspecific siRNA transfected CT-1 levels (Fig. 3.6, $p<0.05$). *IFNT* transcript levels were also significantly diminished from untransfected controls (33 ± 13% of untransfected levels), which agrees with a similar study showing that CDX2 directly induces *IFNT* expression (Sakurai et al., 2009). Downregulation of *CDX2* also significantly reduced expression of bHLH transcription factors *HAND1* and *MASH2*, as well as *SOX15* ($p<0.05$). Of the candidate genes tested, only the ETS family transcription factor *ETS2* was significantly upregulated following *CDX2* knockdown. While CDX2 is known to repress *OCT4* transcription and induce ETS family transcription factor *ELF5* expression in mice, a decrease in *CDX2* message did not have an effect on either *OCT4* or *ELF5* levels in CT-1 cells. GATA transcription factors *GATA2*, *GATA3*, and *GATA6*, which are all expressed during mouse and bovine trophoblast development, were not affected by a decrease in *CDX2*, nor were *ID2* or *FGFR2* transcript levels.

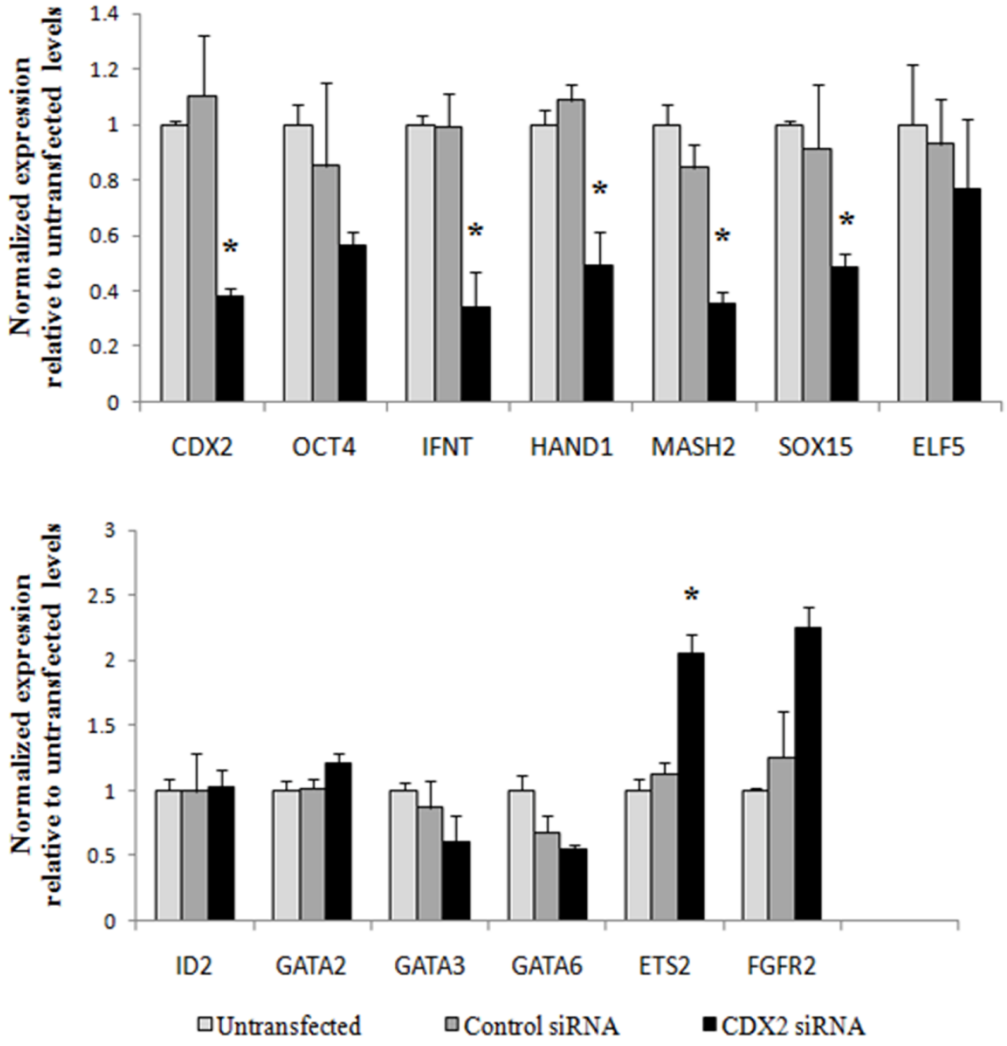



**Figure 3.6.** Expression of bovine trophoblast lineage genes in response to siRNA mediated downregulation of CDX2 for 48 hours. Candidate gene mRNA levels were determined by qRT-PCR and Ct values were normalized to ACTG1 Ct values. Data are represented as means and standard errors of normalized Ct values relative to ACTG1 normalized 'untransfected' treatment Ct mean (set at 1) for each gene. Means are generated from 3 independent experimental cDNA replicates. Asterisks (*) located above values denote a significant difference between that treatment value and all others values of the corresponding gene ($p<0.05$).

Overexpression of CDX2 affected expression of many of the same candidate genes affected by *CDX2* knockdown. Exogenous CDX2 expression was induced with 10 ng per ml doxycycline, which causes a significant increase in tetracycline responsive gene induction (Sup. Fig. 1, Sup. Fig 4, Jiang et al., 2001) With Lipitoid-based plasmid transfection efficiency at about 9% (Schiffmacher and Keefer, 2010), exogenous *CDX2* expression was increased five-fold (5.1 ± 2.5) relative to endogenous *CDX2* expression levels detected in untransfected CT-l cells or CT-1 cells transfected with empty pTHE vector (Fig. 3.7). This increase in overall *CDX2* expression did not exert positive feedback on endogenous *CDX*2 expression, since qRT-PCR performed with primers amplifying within the native *CDX*2 3' UTR did not produce any significant change from pTHE transfection control levels. *OCT4* transcript levels were significantly reduced, yet there was no reduction in OCT4 protein in response to CDX2 overexpression as determined by immunofluorescence (Supp. Fig. 4, G-I). Transcript levels of *IFNT* as well as transcription factors *ELF5*, *SOX15*, *ETS2*, and *MASH2* were also increased ($p<0.05$).

A 56% attenuation of *OCT4* levels was achieved following Lipitoid-based *OCT4* siRNA transfection (Fig. 3.8). Doxycycline induction of HA tagged OCT4 in transfected cells resulted in a 5-fold increase in *OCT4* transcript levels (Fig. 3.9). While similar degrees of siRNA mediated gene suppression and overexpression of *OCT4* were achieved as those with *CDX2*, there were no significant changes in any candidate genes tested. In addition, there was no observable loss of endogenous CDX2 protein in CT-1 cells overexpressing recombinant HA-tagged OCT4 (Sup. Fig 4, D-F). Untransfected neighboring cells exhibited comparable levels of CDx2 immunofluorescence to transfected cells.

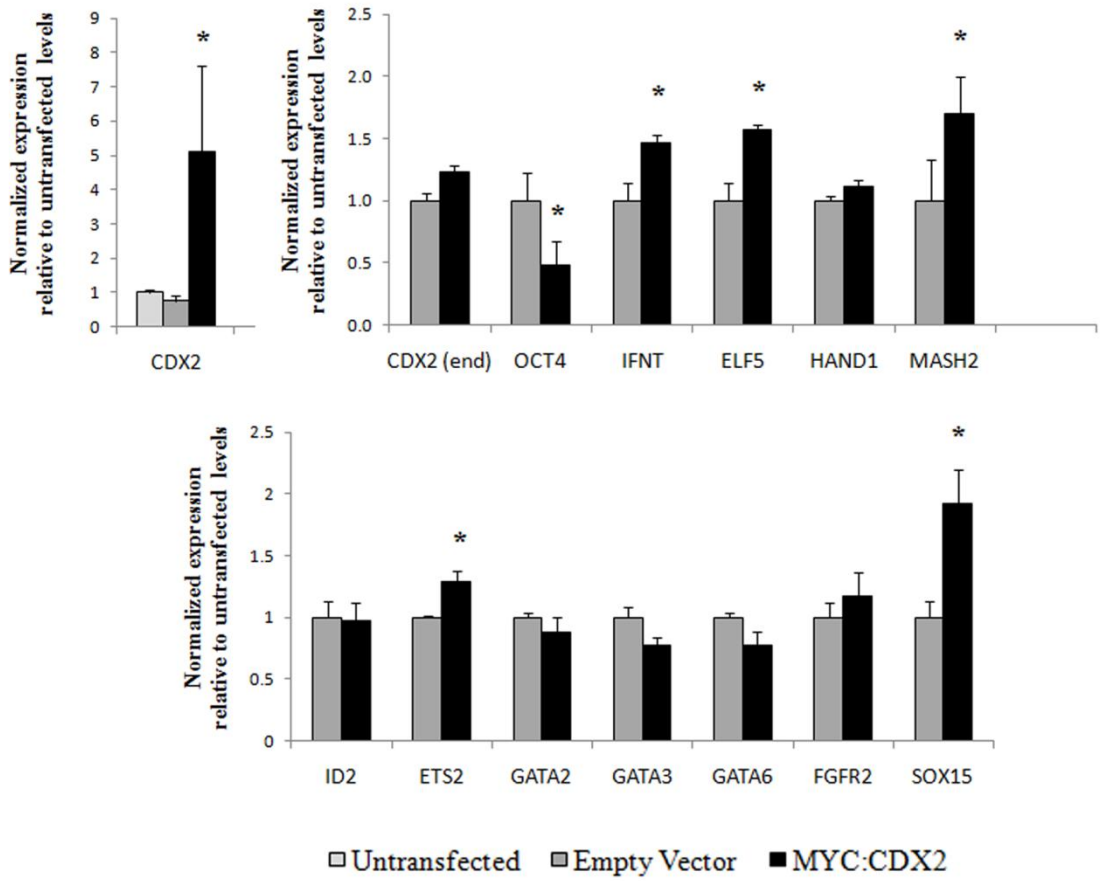



**Figure 3.7.** Expression of bovine trophoblast lineage genes in response to CDX2 overexpression. CT-1 cells were transfected with pTHE.MYC:CDX2 and incubated for 48 hrs in CT-1 medium with 10 ng per ml doxycycline before RNA extraction. Candidate gene mRNA levels were determined by qRT-PCR and Ct values were normalized to ACTG1. Data are represented as the treatment means and standard errors of normalized Ct values relative to the ACTG1 normalized, 'vector only' treatment Ct means (set at 1) for each gene. Means were generated from 3 independent experimental cDNA replicates. Asterisks (*) located above values denote a significant difference between that treatment value and all other values of the corresponding gene ($p < 0.05$).

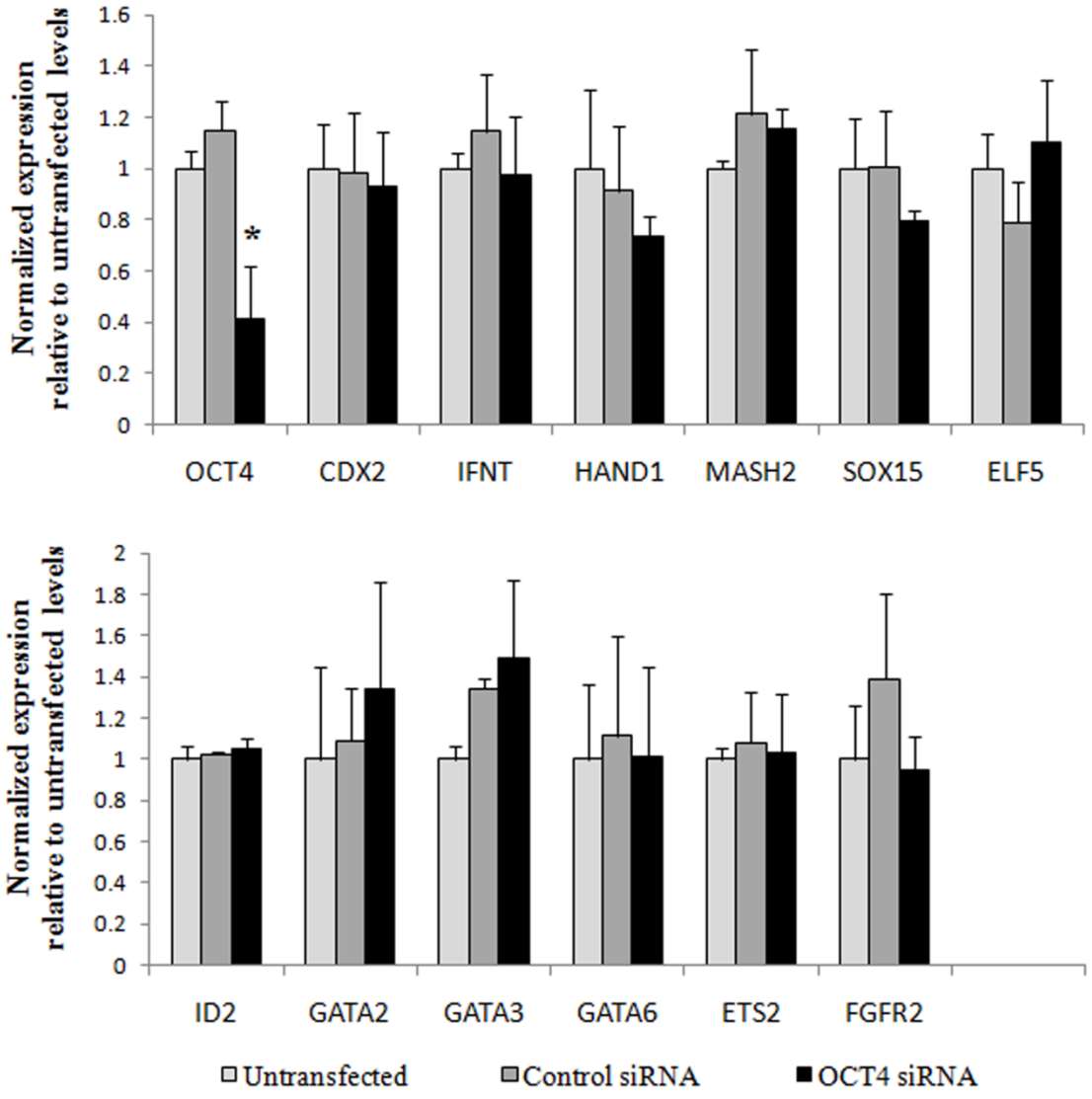



**Figure 3.8.** Expression of bovine trophoblast lineage genes in response to siRNA mediated downregulation of OCT4 for 48 hours. Candidate gene mRNA levels were determined by qRT-PCR and Ct values were normalized to ACTG1. Data are represented as the 'siRNA' treatment means and standard errors of normalized Ct values relative to the ACTG1 normalized, 'untransfected' treatment Ct mean (set at 1) for each gene. Means are generated from 3 independent experimental cDNA replicates. Asterisks (*) located above values denote a significant difference between that treatment value and all other values of the corresponding gene ($p < 0.05$).

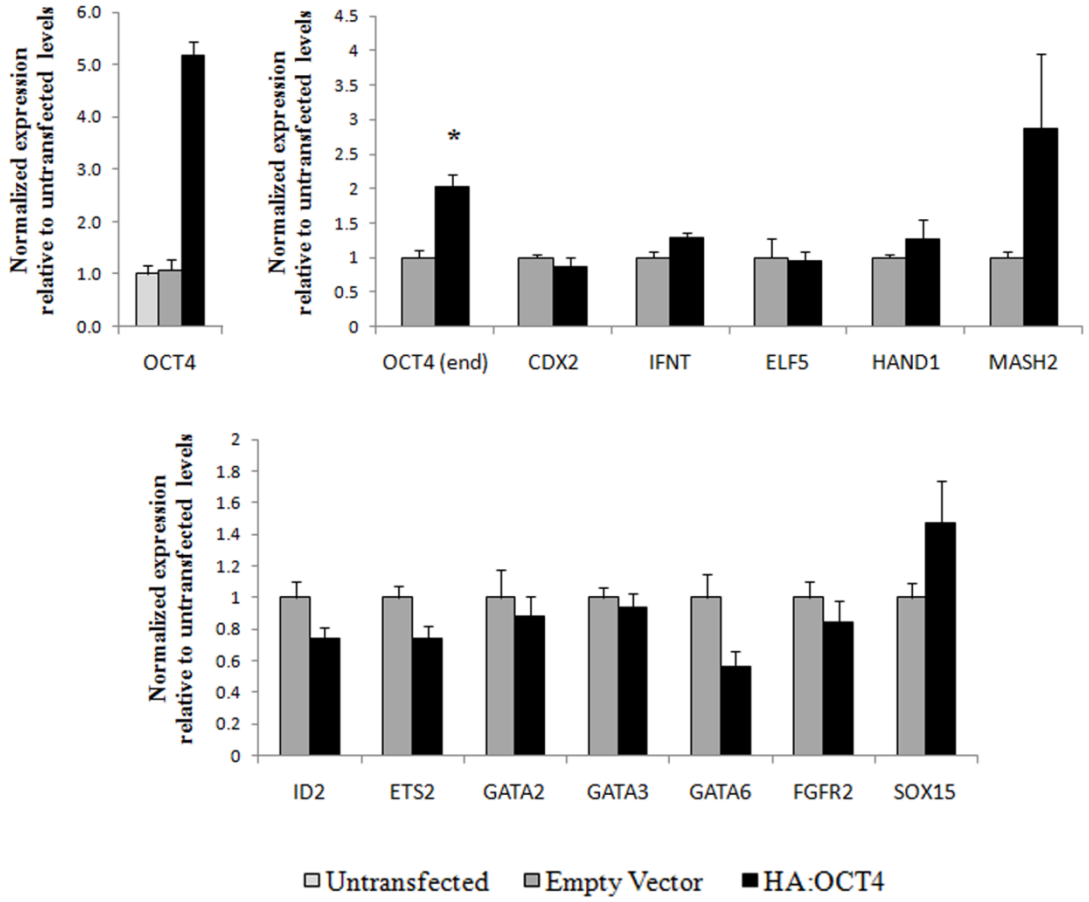



**Figure 3.9.** Expression of bovine trophoblast lineage genes in response to OCT4 overexpression. CT-1 cells were transfected with pTHE.HA:OCT4 and incubated for 48 hrs in CT-1 medium with 10 ng per ml doxycycline before RNA extraction. Candidate gene mRNA levels were determined by qRT-PCR and Ct values were normalized to ACTG1. Data are represented as the treatment means and standard errors of normalized Ct values relative to the ACTG1 normalized, vector-only treatment Ct mean (set at 1) for each gene. Means were generated from 3 independent experimental cDNA replicates. Asterisks (*) located above values denote a significant difference between that treatment value and all other values of the corresponding gene ($p < 0.05$).

## Discussion

Previous studies demonstrated that transcriptional networks in mouse ESC and TSC were sensitive to forced changes in OCT4 and CDX2 levels (Ivanova et al., 2006; Niwa et al., 2000; Niwa et al., 2005). For example, a 50% increase in OCT4 expression in mESC can shift differentiation commitment towards endodermal lineages while a 50% siRNA mediated reduction induces a switch to an active trophoblast program (Niwa et al., 2000). Similarly, CDX2 overexpression has a similar effect in facilitating trophoblast differentiation, due in part to an OCT4/CDX2 reciprocal inhibitory feedback loop at the core of the transcriptional hierarchy in place at the time of ICM and TE lineage commitment. After the bovine TE is formed, we hypothesized that a similar feedback loop may be central for pre-filamentous stage trophoblast elongation. We first investigated whether the CT-1 cell line could serve as an in vitro model for bovine TE. The expression profile of candidate genes in CT-1 cells in comparison to gene expression profiles of bovine embryos at different periods of elongation is summarized in Table 3.4. During elongation, mononucleate cells differentiate into binucleate cells that secrete placental lactogen (CSH1) and Groups I and II bovine pregnancy-associated glycoproteins (PAGs). Bovine PAG-1 family protein PAG9 is expressed in embryos as early as 17 dpc (Arnold et al., 2006) and is abundantly expressed in post-attachment placentas by 45 dpc (Green et al, 2000). Group II PAGs including PAG-2 are not restricted to binucleate cells but also expressed in mononucleate cells as early as the blastocyst stage and well into post-attachment development.

**Table 3.4.** Comparison of selected gene expression profiles between the CT-1 cell line and in vivo bovine embryo stages.

| Pluripotency Transcription Factors | CT-1 Cell Line | Spherical TE* | Ovoid TE* | Cotyledonary Tissue |
|---|---|---|---|---|
| **OCT4** | ✓ | ✓ | ✓ | NO |
| **NANOG** | NO | NO | ✓ | NO |
| **SOX2** | NO | ✓ | ✓ | NO |
| **TE Lineage Transcription Factors** | | | | |
| **CDX2** | ✓ | ✓ | ✓ | NO |
| **Errβ** | ✓ | ✓ | Unconfirmed | NO |
| **EOMES** | NO | ✓ | **NO** | Unconfirmed |
| **Trophoblast Related Factors** | | | | |
| **HAND1.** | ✓ | ✓ | ✓ | ✓ |
| **ETS2** | ✓ | ✓ | ✓ | ✓ |
| **GATA6** | ✓ | ✓ | ✓ | NO |
| **FGF2** | ✓ | ✓ | ✓ | NO |
| **FGFR2** | ✓ | ✓ | ✓ | NO |
| **ELF5** | ✓ | NO | ✓ | NO |
| **ID2** | ✓ | Unconfirmed | Unconfirmed | NO |
| **SOX15** | ✓ | Unconfirmed | Unconfirmed | NO |
| **MASH2** | ✓ | ✓ | ✓ | ✓ |
| **Lef1** | NO | Unconfirmed | Unconfirmed | ✓ |
| **Bovine Specific TE Markers** | | | | |
| **IFNτ** | ✓ | ✓ | ✓ | ✓ |
| **PAG9** | NO | NO | NO | ✓ |
| **PAG2** | ✓ | ✓ | ✓ | ✓ |
| **CSH1** | NO | NO | NO | ✓ |
| **SSLP1 (SOLD1)** | ✓ | Unconfirmed | ✓ | NO |

* Spherical and ovoid stage data compiled from Arnold et al., 2006, Degrelle et al., 2005, and Green et al., 2000.

Secreted protein of Ly-6 domain 1 (SOLD1), previously identified as the early elongation marker c12 (Degrelle et al., 2005), is expressed in the mononucleate cells of the ruminant cotyledon and is also expressed in BT-1 cells, a bovine trophectoderm cell line similar to CT-1 (Ushizawa et al, 2009). Expression of trophoblast markers exhibiting enriched expression profiles during the ovoid early elongation stage (*IFNT*, *PAG-2* and *SSLP1*) but not expression of filamentous stage markers (*CSH1* and *PAG-9*) suggest that CT-1 cells possess a similar transcriptional hierarchy directing mononucleate cell identity and differentiation as observed at the bovine ovoid stage (Table 3.4). Expression of *OCT4* and *ELF5* in CT-1 cells is also similar to the expression profile of these genes during the ovoid stage (Degrelle et al. 2005). Therefore, we investigated the interaction between CDX2 and OCT4 in directing the transcriptional circuitry.

This study expands upon the current list of trophoblast regulators that are regulated by CDX2. Overexpression and knockdown experiments of this study add *ELF5*, *MASH2*, *ETS2*, *SOX15*, and *HAND1* to the list of downstream effectors of CDX2 in the bovine early trophoblast. The mouse orthologs of HAND1, MASH2, and SOX15 are directly involved in terminal differentiation of trophoblast stem cells into giant cells (Yamada et al., 2006, Scott et al., 2000). Our results also confirm earlier findings that CDX2 is a key regulator of *IFNT* expression (Imakawa et al., 2006; Sakurai, et al., 2009) as *IFNT* was downregulated following knockdown of *CDX2*, and overexpression of CDX2 induced *IFNT* transcript levels. Both *MASH2* and *SOX15* levels changed in a similar fashion to *IFNT* levels, suggesting CDX2 is a potent activator of these genes as well. CDX2 induction of *MASH2* may be independent of ID2, a transcriptional repressor of *MASH2*, as CDX2 overexpression or knockdown did not elicit any changes in *ID2*

transcript levels that could potentially affect *MASH2* levels (Janatpour et al., 2000). In mouse trophoblast stem cells, MASH2 induces cell proliferation and inhibits HAND1 directed giant cell differentiation (Scott et al., 2000). In bovine trophoblast cells, this function of MASH2 may be conserved to promote mononucleate cell proliferation.

Induction of *ELF5* following CDX2 overexpression in CT-1 cells suggests that the CDX2-ELF5-EOMES circuit responsible for mouse trophoblast maintenance is conserved in bovine trophoblast. In conjunction with FGF4/FGFR2 signal transduction, CDX2, EOMES, and ELF5 form the core of the transcriptional network that maintains multipotency within the extra embryonic ectoderm. CDX2 and EOMES are required for early mouse TE maintenance (Strumpf et al., 2005; Russ et al., 2000) and are hypothesized to induce *Elf5* later on (Ng et al., 2008). *Elf5* can also be induced by FGF/FGFR signaling (Metzger et al., 2007). ELF5 in turn completes a positive feedback loop by directly activating *Cdx2* and *Eomes* gene expression. The importance of ELF5 as a key TSC regulator is highlighted by the fact the *Elf5* promoter, but not the promoters of *Cdx2* and *Eomes*, is preferentially methylated and silenced in the mouse ICM lineage (Ng et al., 2008). Further investigation is needed to determine whether the CDX2-ELF5-EOMES circuit is conserved in bovine trophoblast. EOMES expression is restricted to the germinal disk at the filamentous stage and analysis of *EOMES* by RT-PCR revealed expression in ovoid stage extra-embryonic tissues (Degrelle et al., 2005). However, this result did not determine if the trophoblast or endoderm expresses *EOMES*. *EOMES* expression was not detected in either CT-1 cells in our study (Table 4) or in BT-1 cells (Ushizawa et al., 2005).

In addition to ELF5, ETS2 is another ETS transcription factor that promotes trophoblast development. Transgenic mice expressing inactive ETS2 fail to develop a normal extra-embryonic ectoderm tissue lineage containing a functional trophoblast stem cell population, and loss of ETS2 in TSC compromises self renewal that cannot be rescued by CDX2 overexpression. (Wen et al., 2007). ETS2 may play a central, conserved role in trophoblast developmental potential as ETS2 induces *CDX2*. Our data demonstrate that *ETS2* is induced following CDX2 suppression and CDX2 over-expression. A small but significant induction in ETS2 expression suggests a possible positive feedback loop in place between CDX2 and ETS2. However, an average two-fold induction of ETS2 transcription following *CDX2* knockdown indicates that *ETS2* levels may be induced to stabilize the trophoblast gene network by sensing when *CDX2* levels need to be increased.

Our initial RT-PCR screen confirmed earlier findings of *GATA6* expression in CT-1 cells (Bai et al, 2009). GATA6 expression is not restricted to the trophoblast but is also expressed in primitive endoderm and extra embryonic mesoderm at the ovoid stage (Degrelle et al., 2005). Trophoblast specific functions of GATA6 are unknown, but GATA6 was recently demonstrated to regulate the balance between lung bronchioalveolar stem cell proliferation and differentiation (Zhang et al., 2007). In addition to GATA6, GATA2 and GATA3 are also expressed in CT-1 cells (Bai et al., 2009). Our data do not indicate any significant relationship between CDX2 or OCT4 and GATA proteins, although OCT4 overexpression indicated possible negative regulation of *GATA6* ($p<0.07$). GATA transcription factors function in a parallel trophoblast circuit that regulates downstream genes in common with CDX2 (Ralston et al., 2010). *Gata3* is

co-induced with *Cdx2* by TEAD4 and participates in TE lineage commitment by positively regulating *CDX2, IFNT* and other trophoblast specific genes (Bai et al., 2009; Home et al., 2009; Ralston et al, 2010). GATA3 also restricts GATA2-mediated trophoblast giant cell (TGC) differentiation by occupying GATA regulatory elements (Ray et al., 2009).

Our data from *CDX2* and *OCT4* overexpression and knockdown suggests that that the reciprocal inhibitory feedback loop between CDX2 and OCT4 is uncoupled in CT-1 cells or cannot be experimentally confirmed with siRNA knockdown of only about 60% or overexpression in only 9% of cells. Assessment of the effects of overexpression of one gene on the other by immunofluorescence also showed no evidence of an OCT4/CDX2 inhibitory feedback loop (Supp. Fig. 4). The fact that individual cells overexpressing recombinant OCT4 or CDX2 were no different in protein levels of the other gene from untransfected neighboring cells suggests that OCT4 and CDX2 are not potent repressors of each other in CT-1 cells. However, it should be noted that there are limitations in utilizing immunofluorescence to semi-quantify protein levels within cells. The nuclear localization of bovine OCT4 containing the N-terminal transactivation domain indicated that bovine OCT4 protein may function as a transcription factor regulating gene expression of the early bovine trophoblast (Cauffman et al., 2006). While many gene transcript levels were altered following forced changes in *CDX2* levels, there were no gene candidates within our panel that significantly responded to overexpression or suppression of *OCT4* except for an increase in endogenous *OCT4* by positive feedback following overexpression. *MASH2* and *SOX15* levels slightly increased following OCT4 overexpression. However, these findings were not statistically significant ($p < 0.06$,

$p<0.17$, respectively). It was hypothesized that *IFNT* would be upregulated following *OCT4* knockdown or downregulated upon OCT4 overexpression, as OCT4 binds to ETS2 in vitro and quenches ETS2-mediated *IFNT* induction (Ezashi et al., 2001). It is possible that OCT4 repression of *IFNT* is a temporal, protein level, and cell-specific mechanism. For example, morula and ICM cells expressing high levels of OCT4 may inhibit any expression of *IFNT* potentially activated by ETS2 (Ezashi et al., 2001). Within the bovine trophoblast derived cell lines, *IFNT* is induced by CDX2 and DLX3 as well (Ezashi et al., 2008; Sakurai et al., 2009), and these additional signals may override any potential repression by OCT4 in vivo.

OCT4 may interact with unidentified trophoblast specific proteins that may form heterodimers with OCT4 to regulate downstream gene expression similar to the OCT4-SOX2 heterodimer in the ICM. A possible candidate is SOX15, which is expressed in CT-1 cells and has been demonstrated to interact with OCT4 at adjacent octamer and sox binding sites in mouse ESC (Maruyama et al., 2005). The primary role of OCT4 may also be as an epigenetic component in a manner similar to the OCT4 association with the histone H3 Lys 9 (H3K9) methyltransferase ESET that silences trophoblast genes in the ICM (Yuan et al., 2009). To address this, co-immunoprecipitation of OCT4 complexes from CT-1 lysates could be performed. However, the poor efficiency of immunoprecipitating recombinant OCT4 with the polyclonal anti-human OCT4 antibody as shown in Fig. 3.3 suggests that a polyclonal anti-bovine OCT4 antibody may be more suitable for the task (Sup. Fig. 3).

In this present study, we demonstrate the utility of the CT-1 cell line as a genetic model for studying the transcriptional networks governing developmental potential in the

early bovine trophoblast. The CT-1 cell line appropriately displays a profile of proteins that are temporally expressed in the pre-attachment trophoblast. Our overexpression and siRNA mediated knockdown experiments indicate a conserved central role for CDX2 in governing the trophoblast specific transcription factor hierarchy. These studies confirm previous findings of CDX2 regulation of *IFNT* in CT-1 cells and *Oct4* in mice ESC, and provide new evidence for regulation of *MASH2*, *HAND1* and *SOX15* by CDX2. Our inability to detect significant changes in candidate gene expression by altering *OCT4* expression suggests that OCT4 may play a peripheral role in early ungulate trophoblast development, although it is possible the limitations in CT-1 transfection efficiency could not provide the resolution necessary to observe any effects. Future studies will determine the roles of CDX2 and other trophoblast transcription factors in this study in mononucleate cell proliferation and binucleate cell differentiation rates. For these studies, we predict that CT-1 cells stably transfected with our doxycycline inducible plasmids would prove an excellent model for such long term experiments.

# Chapter 4: Sense and antisense RNAs containing an *OCT4* retrocopy are expressed in CT-1 cells.


**Abstract**

Recent studies have demonstrated that sense and antisense RNA transcripts from gene retrocopies within mammalian genomes may act as endogenous siRNAs to regulate their parental protein-coding genes. *OCT4* (also known as *POU5F1*), a regulator of pluripotency, is known to have 6 retrocopies in the human genome. Our analysis of *OCT4* expression in the bovine trophectoderm CT-1 model cell line identified transcripts corresponding to an *OCT4* retrocopy located in the first intron of *ETF1* on bovine chromosome 7. The 3' truncated *OCT4* retrocopy contains *OCT4* exons 1-4 and 27 bp of exon 5 with 10 substitutions and no other mutations. It is fused at the 3' end with a 5' truncated bovine LINE (L1 BT). Any presumptive protein expressed from this retrocopy could not be detected by Western blot using an antibody binding to the OCT4 N-terminus. Strand specific RT-PCR revealed both strands of this locus are transcribed. While nested 5' and 3' RACE did not identify discrete ends for either transcript, nested PCRs from 5' and 3' RACE cDNAs indicate that *OCT4* retrocopy-containing RNAs transcribed from both strands may be over 5 Kb in length. Alternatively, multiple overlapping RNAs across the locus could be simultaneously transcribed. Our results suggest that double stranded RNAs could be created from this locus.

## Introduction

Transposable elements are a unique group of mobile DNA sequences that possess the ability to "cut and paste" themselves into and out of genomes. About 90% of transposable elements are Class I retroelements (REs) and are the most abundant in human DNA (Bannert and Kurth, 2004). REs spread throughout the genome by reverse transcribing expressed element RNA intermediates into cDNA and then integrating the cDNA into a new genomic location. REs are further subdivided into those which contain long terminal repeat (LTR) and those which do not contain LTRs. The most abundant types of non-LTR retroelements are long interspersed nuclear elements (LINES), short interspersed nuclear elements (SINES), and mRNA derived retrocopies of parental genes (Gogvadze and Buzdin, 2009; Vinckenbosch et al., 2006). Retrotransposition of mRNA requires enzymatic activity of LINE machinery transcribed and translated from two open reading frames (ORFs) located within the LINE sequence (Esnault et al., 2000). Retrotransposition of intact mRNA results in full length retrocopies that are intronless and possess a polyadenylated tail. However, many retrocopies lack the necessary regulatory elements of the original parental gene promoter that were not transcribed as part of the mRNA 5' untranslated region (5' UTR). Therefore, transcription is presumably not induced, and these nonfunctional sequences acquire mutations over time. Degenerate retrocopies were once all thought to be nonfunctional and collectively termed retropseudogenes or processed pseudogenes. However, transcriptome-wide analyses in multiple species indicate that as many as 20% of retrocopies or processed pseudogenes are transcribed (Gerstein et al., 2007), often by acquiring new regulatory elements at the

site of retrotransposition (Bai et al., 2008; Marques et al., 2005; Vinckenbosch et al., 2006).

Retrocopies that express known functional gene products are regarded as retrogenes, and are common in mammals (Marques et al., 2005; Pan et al., 2009). Genome-wide explorations of 6 mammalian species as well as chicken and pufferfish (*Tetraodon*) inferred that about 3% of protein-coding genes have a putative retrogene copy (Yu et al., 2007). Detailed characterization of putative retrogenes reveals that their creation is unique and may contribute to new gene formation. For example, the mouse *Pmse2* transcript that encodes a proteasome activator subunit was retrotransposed into a transcriptionally active LINE1 (L1) retroelement. L1 promoter driven expression of the intronless *Pmse2* ORF resulted in the formation of a constitutively active retrogene producing protein identical to the parental gene expressed protein (Zaiss et al., 1999). When a retrocopy is retrotransposed into a site downstream of another gene promoter, chimeric genes can be formed. During the evolution of the owl monkey, a new world primate, retrotransposition of full length *CYPA* cDNA into the *TRIM5* open reading frame created a new exon and TRIMCYP fusion protein. For the owl monkey, this new chimeric protein conveyed HIV-1 resistance. (Saya et al., 2004). Retrogenes not only contribute to new gene formation and speciation, but can also contribute to phenotypic variation, as evidenced by a strong association of FGF4 retrogene expression and chondrodysplaysia in short legged dog breeds (Parker et al., 2009). Retrocopies can also form new genes when fused in frame with LINE intermediates (Iwashita et al., 2001).

Transcription factors expressed in human embryonic stem cells (ESC) that regulate pluripotency were found to have an enriched number of daughter retrogenes and

processed pseudogenes over transcription factors that are not part of the ESC transcriptome. Pluripotency regulators OCT4, NANOG, and STELLA were found to have 6, 10, and 16 retrocopies respectively (Pain et al., 2005; Boothe et al., 2004). The higher incidence of retrotransposition of these genes over others may be due to their higher probability of germ line transmission as *OCT4*, *NANOG* and *STELLA* mRNA are all expressed in germ cells. The most recent human retrocopy, *NANOGP8*, was estimated to have retrotransposed 5.2 million years ago and contains the full ORF without insertions or deletions and only 9 substitutions within 913 bases (Boothe et al., 2004). It was demonstrated that *NANOGP8* is a retrogene in many human cancer cell lines as they express NANOGP8 protein but do not express NANOG protein from the original *NANOG* gene. *OCT4* retrocopies are also transcribed in many cancer cell lines and cancer tissues and are often co-expressed with *OCT4* (Suo et al., 2005). *OCT4* retrocopies are expressed in many adult tissue types and cell lines and were often incorrectly identified as *OCT4* transcripts, resulting in a call for more thorough sequence analysis and *OCT4* primer design (Liedtke et al., 2008).

Our own detailed analysis of *OCT4* expression in the bovine blastocyst derived trophectoderm CT-1 cell line (Talbot et al., 2000) found that an *OCT4* retrocopy is co-expressed (Fig. 2.2). Sequence data indicated that it was transcribed from a processed *OCT4* retrocopy located in the first intron of eukaryotic translation termination factor 1 (*ETF1*) and matches what was previously described as the bovine pseudogene *POU5F1rs1* (van Eijk et al., 1999). As it is becoming increasingly evident that transcribed processed pseudogenes can encode functional gene products (Gerstein et al.,

2007), the focus of this study was to characterize the bovine chromosome 7 *OCT4* retrocopy locus and gain insight into a potential function.

## Methods and Materials

### Reverse transcription and touchdown PCR

Cell culture, RNA extraction, RNA quantification and Superscript III based CT-1 first strand cDNA synthesis (Invitrogen, Carlsbad, CA) have been described previously (Schiffmacher and Keefer, 2010a). All primers listed in Table 4.1 were developed using IDT Real Time PCR software (Integrated DNA Technologies, http://www.idtdna.com/analyzer/ Applications/Oligo Analyzer). PCR reactions for products shown in Fig. 2 consisted of 4 µl 5X HF iPROOF buffer, 0.2 µl iPROOF polymerase (172-5330, Bio-Rad, Hercules, CA), 1.6 µl 2.5 mM dNTP, 11.2 µl nuclease free $H_2O$, 1 µl each primer, and 1 µl diluted CT-1 cDNA. PCR reactions using primer sets 73-74, 73-141, and 146-74 were subjected to touchdown PCR in the following manner. Samples were initially denatured for 30 sec at 98 °C. The next 10 cycles consisted of a denaturation step at 98 °C for 10 sec, annealing for 30 sec, and extension at 72 °C for 45 sec. First round of annealing was performed at 69 °C, and at each subsequent step the annealing temperature was decreased by 1 °C. The remaining 25 cycles were performed as above at a fixed annealing temperature of 59 °C, before finishing with a final extension period at 72 °C for 10 min. PCR reactions using primer sets 73-198, 195-74 and 115-74 were also subjected to touchdown PCR using the same protocol but with the extension times modified to 30 sec. to maintain extension at ~30 sec/ Kb. To account for potential genomic DNA contamination amplification during RT-PCR, 1µg of CT-1 RNA was also used as RNA template for cDNA synthesis, however

reverse transcriptase was omitted from these reactions (RT-). Therefore, RT- samples did not contain cDNA. All PCR reactions performed using RT- templates failed to amplify product. An additional reaction for each primer set was also performed without cDNA to account for any reagent DNA contamination. Single amplicon bands were purified from agarose gels using a MinElute PCR Purification Kit (28004, Qiagen, Valencia, CA) and sequenced at the University of Maryland Biotechnology Institute: Center for Biosystems Research (College park, MD).

**Table 4.1.** Primers used for RT-PCR and RACE.

| QPCR Primers | Sequence (5' - 3') | Primer Length (nt) | NW_001495368.2 Negative Strand Location | Position (-3,80X,XXX) | Negative Strand Orientation |
|---|---|---|---|---|---|
| 146 | TCAGGGCACATCTGGGACTTG | 21 | 5' to OCT4 retrocopy | 6044-6024 | Sense |
| 163 | TCCACCCTGGCTCCTGTCTCTGGATTGCC | 29 | LINE L1 | 5586-5614 | Sense |
| 73 | GGGTGTGAAGTGGGTTTGTG | 20 | 5' to OCT4 retrocopy | 4496-4477 | Sense |
| 115* | GACACCTCGCTTCTGACTTC | 20 | Within OCT4 retrocopy | 4375-4356 | Sense |
| 160** | ACTCGCAGGCGCCGCGAAGCTGGACAAGG | 29 | Within OCT4 retrocopy | 4032-4004 | Sense |
| 195* | AGGAGTCCCAGGACATCAA | 19 | Within OCT4 retrocopy | 3982-3964 | Sense |
| 157 | GTCCGAGTGTGGTTTTGCAACCGTGTACC | 29 | Spans OCT4 retrocopy/ L1 junction | 3563-3535 | Sense |
| 159 | AGAGGGGTTGGTAGCTGTTCAAAAGGGAGC | 30 | 3' to OCT4/ L1-BT sequence | 2551-2522 | Sense |
| 155 | GGCAATCCAGAGACAGGAGCCAGGGTGGA | 29 | LINE L1 | 5614-5586 | Antisense |
| 156 | CCGCCATGGGGAAGGAAGGCACCCCCAAA | 29 | Spans OCT4 retrcopy 5' end | 4376-4404 | Antisense |
| 198* | GCAGCTTACACATGTTCTTGAA | 22 | Within OCT4 retrocopy | 3782-3803 | Antisense |
| 158 | CCTGACTGGTGTGAAGTGGTACACGGTTGC | 30 | Spans OCT4 retrocopy/ L1 junction | 3518-3547 | Antisense |
| 74 | CCATTCATCTGCTGATGGAC | 20 | L1_BT 3' fused to OCT4 retrocopy | 3221-3240 | Antisense |
| 141 | ATGTGTTCCCCATCCTGAAC | 20 | L1_BT 3' fused to OCT4 retrocopy | 2787-2806 | Antisense |
| 161 | TGGAGAGAAGCCAGTCTCTTCAGCAAGTGG | 30 | LINE L1 | 551-580 | Antisense |

* This primer also matches OCT4 (NM_174580)
** 27/29 match with OCT4 (NM_174580)

**5’ and 3’ RACE ready cDNA synthesis and rapid amplification of cDNA ends (RACE)**

5’ and 3’ RACE ready cDNAs were generated using the SMARTer RACE cDNA Amplification Kit (634924, Clontech, Mountain View, CA) according to kit instructions. One µg of purified CT-1 RNA was used per reaction. All kit recommendations for testing RNA purity, reagent quality control, and cDNA synthesis efficiency were successfully performed. RACE cDNAs were diluted in 100 µl Tricine-EDTA buffer for use as PCR template. First round RACE was performed using one specific primer and SMARTer RACE Universal Primer Mix (UPM) and run using the SMARTer RACE Program 1 modified for use with iPROOF DNA polymerase (Bio-Rad). Samples were initially denatured for 2 min at 98 °C. The first 5 cycles consisted of one denaturation step at 98 °C for 10 sec, followed by a combined annealing/extension step at 72 °C for 2 min. The next 5 cycles consisted of one denaturation step at 98 °C for 10 sec, an annealing step at 70 °C for 20 sec, and an extension step at 72 °C for 2 min. The remaining 27 cycles were performed as in the previous 5 cycles but at a fixed annealing temperature of 68 °C, before finishing with a final extension period at 72 °C for 10 min. A 5 µl aliquot of each RACE reaction was separated by agarose gel electrophoresis. RACE samples typically exhibited multiple bands of varying intensity of ethidium bromide staining. RACE products were purified and diluted 100-fold to be used as template for a second round of nested RACE using one specific primer downstream of the primer used in first round RACE and the abridged Nested Universal Primer (NUP). Templates were also used for nested PCR using two specific primers and no NUP. Second round PCRs were performed according to SMARTer RACE Program 2 modified for use with iPROOF DNA polymerase (Bio-Rad). Samples were initially denatured for 2 min at 98 °C, followed by

35 cycles of the following three steps: denaturation at 98 °C for 10 sec, annealing at 68 °C for 30 sec, and extension at 72 °C for 45 sec. A final extension period was done at 72 °C for 10 min. PCR products were separated by agarose gel electrophoresis, and prominent DNA bands were excised and purified as previously described. PCR amplification of DNA using IPROOF DNA polymerase produces blunt ended amplicons. Therefore, purified PCR fragments were A-tailed. Briefly, 20 µl of purified PCR product was combined with 3 µl 10X PCR buffer, 3 µl 25 mM $MgCl_2$, 0.6 µl 10 mM dNTP, and 0.3 µl TAQ polymerase (10342-035, Invitrogen), and 2.1 µl nuclease-free $H_2O$. Reactions were incubated at 72 °C for 25 min, and purified again as previously described. A-tailed PCR fragments were cloned into pGEM-T Easy linearized plasmid according to the manufacturer's instructions (A1380, Promega, Madison, WI). Ligations were transformed into JM109 competent *E. coli* and screened using X-gal and IPTG selection. Individual colonies containing recombinant clones were selected and cultured before plasmid was purified using NucleoSpin Miniprep kit, (740588.10, Machery-Nagel, Bethlehem, PA). All clones were sequenced with T7 and SP6 primers.

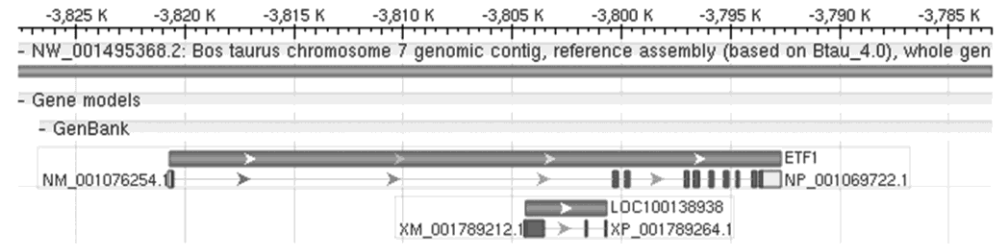
**Sequence Analysis**

All bovine genome sequence references are based on the *Bos taurus* Genome assembly version Btau_4.0. All sequence analyses of the *OCT4* retrocopy were based on the genomic loci of eukaryotic translation termination factor 1 (*ETF1*, Genbank Accession NM_001076254) intron 1, located on the chromosome 7 contiguous reference assembly (NW_001495368.2). For the *ETF1* gene and the intronic *OCT4* retrocopy (if *OCT4* transcriptional orientation is maintained), the sense strand is the negative strand. Retroelement analysis of the bovine *ETF1* intron was performed using the online software Repeatmasker, current version: open-3.2.8 (A.F.A. Smit, R. Hubley & P. Green, unpublished data). All sequence alignments were performed using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) or European Molecular Biology Laboratories-European Bioinformatics Institute's (EMBL-EBI) ClustalW2 program.

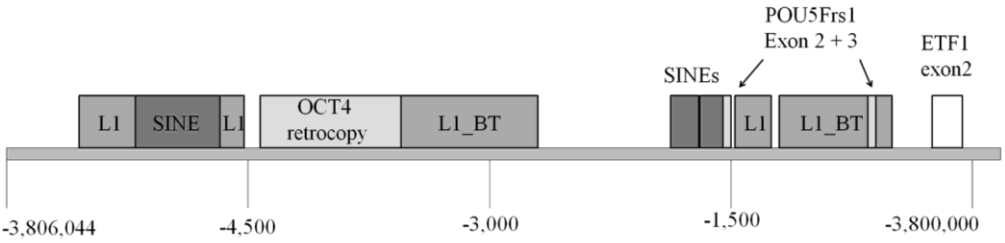
**Results**

Nucleotide polymorphisms of sequenced *OCT4* products amplified from bovine CT-1 cDNA indicated either allelic differences or a separate *OCT4*-like RNA transcript. BLAST analysis of the sequence containing the alternative nucleotide configuration resulted in a 100 % match with a predicted *OCT4*-like gene and mRNA (Genbank LOC100138938 and XM_001789212.1) that is located within the 20,198 bp first intron of *ETF1* on chromosome 7. The nucleotide differences also matched a putative *OCT4* pseudogene named *POU5F1rs1* previously detected in bovine genomic DNA (van Eijk et al., 1999). Further characterization of the computationally derived *OCT4*-like gene showed that the predicted first exon aligns with *OCT4* exons 1-4 and the first 27 bases of exon 5, indicating that this 861 bp sequence is an intronless 3' truncated OCT4 retrocopy. This fragment contains no insertions or deletions but has 10 substitutions, making it 98.9% identical to matching *OCT4* sequence (Genbank NM_174580.2). It is also the most conserved *OCT4* retrocopy obtained by bovine genome BLAST analysis (E = 0.0, Sup. Fig. 5). Predicted exons 2 and 3 of the LOC100138938 gene, however, were not well conserved with the remaining exon 5 sequence of *OCT4*, and were considered to be incorrectly predicted (Sup. Fig. 6). Sequence 3' to the *OCT4* retrocopy was BLASTed and matched a retroelement endonuclease-reverse transcriptase (ERT) coding sequence. Therefore, a detailed transposable element analysis was performed in this region using Repeatmasker software. Results showed that the *OCT4* retrocopy is fused at the truncated 3' end to a 5' truncated bovine L1 LINE (L1 BT), and is flanked by other LINE and SINE elements (Fig. 4.1).

**A.**




**B.**




**C.**

| Retroelement (Name and Class) | Position (-3,80X,XXX) | Length | % Subst. | % Deletions | % Insertions | Consensus Alignment C/ + | Begin | End | Left |
|---|---|---|---|---|---|---|---|---|---|
| L1MB8 LINE/L1 | 5694-5179 | 516 | 20.4 | 13 | 1.5 | C | 5491 | 6081 | 97 |
| ART2A SINE/RTE | 5175-4648 | 528 | 7.4 | 0.8 | 0.4 | C | 1 | 530 | 21 |
| BovB LINE/RTE | 4647-4516 | 132 | 3 | 6.8 | 0 | C | 2804 | 2944 | 358 |
| OCT4 retrocopy | 4400-3540 | 861 | 1.2 | 0 | 0 | + | 19 | 861 | 222 |
| L1_BT LINE/L1 | 3539-2705 | 835 | 4.3 | 0.6 | 0.1 | + | 7533 | 8371 | 19 |
| Bov-tA3 SINE/BovA | 1903-1751 | 153 | 13.1 | 0.7 | 0 | C | 13 | 166 | 47 |
| BOV-A2 SINE/BovA | 1734-1597 | 138 | 12.5 | 0 | 1.4 | + | 137 | 272 | 0 |
| L1-2_BT LINE/L1 | 1525-1263 | 263 | 23.7 | 8.3 | 4.7 | C | 1636 | 1878 | 137 |
| L1-2_BT LINE/L1 | 1236-519 | 718 | 19.9 | 11.6 | 4.3 | C | 583 | 1352 | 663 |
| ETF1 Exon 2 | 0326-0150 | 177 | | | | | | | |

**Figure 4.1.A**. NCBI Sequence viewer map of Bos Taurus chromosome 7 genomic reference assembly (based on Btau_4.0 whole genome shotgun sequence ref NW_001495368.2| Bt7_WGA801_4) that contains the *ETF1* gene (NM_001076254) and the *POU5F1* retrocopy (LOC100138938). Bases -3,827,652 : -3,782,998 (44,655 total) of the negative strand are shown. **B**. Magnification of *ETF1* intron 1 sequence flanking the *OCT4* retrocopy or "exon 1" of Genbank predicted *OCT4*-like gene (XM_001789212.1), in light gray. Exons 2 and 3 of the Genbank predicted *OCT4*-like gene are also in light gray. Exon 3 lies within a fragment of an L1-2_BT LINE. Other LINE fragments are in medium gray. SINE fragments are shaded dark gray. The *ETF1* exon 2 is added for reference (white). Representations are approximate in scale. **C**. Repeatmasker software analysis of retroelements in 1B. Elements either directly match (+) or compliment (C) the database consensus orientation. 'Begin' and 'End' indicate the position alignment of the match within the consensus, and 'Left' indicates the number of base pairs remaining in the consensus but missing from the end of the fragment.

The bovine specific L1 BT LINE is the most abundant LINE in the bovine genome and considered to still be active (Adelson et al., 2009). Flanking sequence immediately 5' to the retrocopy did not contain L1 BT sequence, suggesting that the retrocopy was not retrotransposed into a pre-existing L1 BT fragment. Alternatively, it is likely L1 enzymatic retrotransposition involved an RNA intermediate template switch between *OCT4* mRNA and L1 BT, resulting in a chimeric retrocopy that was singularly integrated into the locus. (Gogvadze and Buzdin, 2009).

This chimeric retrocopy contains the 5' region of *OCT4*, the 3' region of L1 BT including the poly (A) tail, and both parts are in the same transcriptional orientation. Based on these criteria, it was conceivable that the fused intronless ORFs could produce a chimeric fusion protein and reclassify the *OCT4* retrocopy as a retrogene. An amino acid (AA) prediction analysis and alignment was performed on the chimeric sequence (Sup. Fig. 7). Assuming the putative protein shares the conserved translational start site of bovine OCT4 and human OCT4A, it would contain an intact OCT4 N-terminal transactivation domain and a relatively intact Pit-Oct-Unc specific (POUs) DNA binding domain. The POU homeodomain would also be relatively conserved except for the loss of the C terminal end that is normally translated from exon 5. Continuation of the fused ERT domain does not occur, however, as a frame shift at the *OCT4*/L1 BT junction results in a nonsense mutation 16 AA from the template switch. It is still possible that an OCT4-like protein containing a short novel C-terminus may be expressed. However, this appears unlikely as immunoblotting of CT-1 whole cell lysates using OCT4 antibodies recognizing the first 134 AA detected only the native OCT4 protein and no bands for the predicted proteins 297 AA in length (Fig. 3.2).

Next, the transcribed RNA containing the *OCT4* retrocopy was characterized. Initial RT-PCR experiments using locus specific primers (Table 4.1) identified an oligo (dT) synthesized CT-1 cDNA containing the *OCT4* retrocopy and adjacent L1 BT sequence (Fig. 4.2, Sup. Fig. 6). In addition, a 2823 bp product was successfully amplified that contained additional L1 fragments and an ART2A SINE located 5' to the *OCT4* retrocopy. These PCR fragments also suggested that the predicted *OCT4*-like gene (LOC100138938) is incorrectly annotated in Genbank. To determine whether these transcripts were sense or antisense, cDNAs were reversed transcribed from CT-1 RNA using non-retroelement binding, strand specific primers that would prime reverse transcription across the *OCT4* retrocopy (Fig. 4.3). While our results indicated that both sense and antisense transcripts containing the *OCT4* retrocopy are expressed in CT-1 cells, further verification of transcript lengths would help distinguish these transcripts apart from each other. 5' and 3' RACE were performed to expand on transcript length and determine RNA ends. Nested RACE and additional nested PCR using locus specific primers resulted in amplification of distinct cDNA sequences. Sequence obtained from 5' RACE of the sense transcripts show the 5' end aligns over a one kilobase upstream of the *OCT4* retrocopy (Fig. 4.4). Attempts to extract the 3' end by 3' RACE were unsuccessful. Nested PCRs performed on first round RACE products also amplified fragments that spanned kilobases downstream. Assuming all PCR products were amplified from the same sense transcript, the resulting RNA would be over 5527 bp in length.

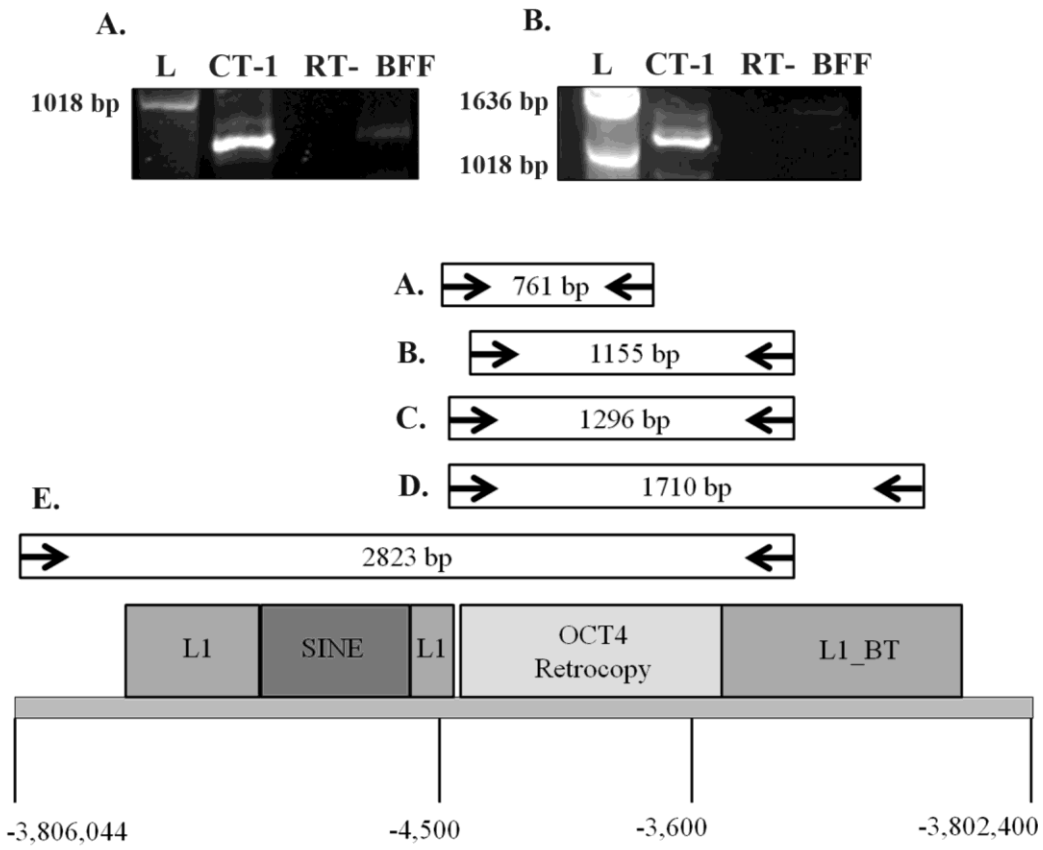



**Figure 4.2.** RT-PCR products containing the *OCT*4 retrocopy were successfully amplified from CT-1 cDNA using different primer sets. All products were sequenced and found to contain the *OCT4* retrocopy/ LI BT fusion junction unique to chromosome 7 (NW_001495368.2). Primer sets **A** (73/198) and **B** (115/74) amplified prominent bands from CT-1 cDNA but not from bovine fetal fibroblast (BFF) cDNA or negative control (RT-) cDNA. Primer sets **C** (73/74), **D** (73/141) and **E** (146/74) produced similar results (gels not shown). PCR fragment lengths and approximate primer locations are also displayed in the schematic representation.

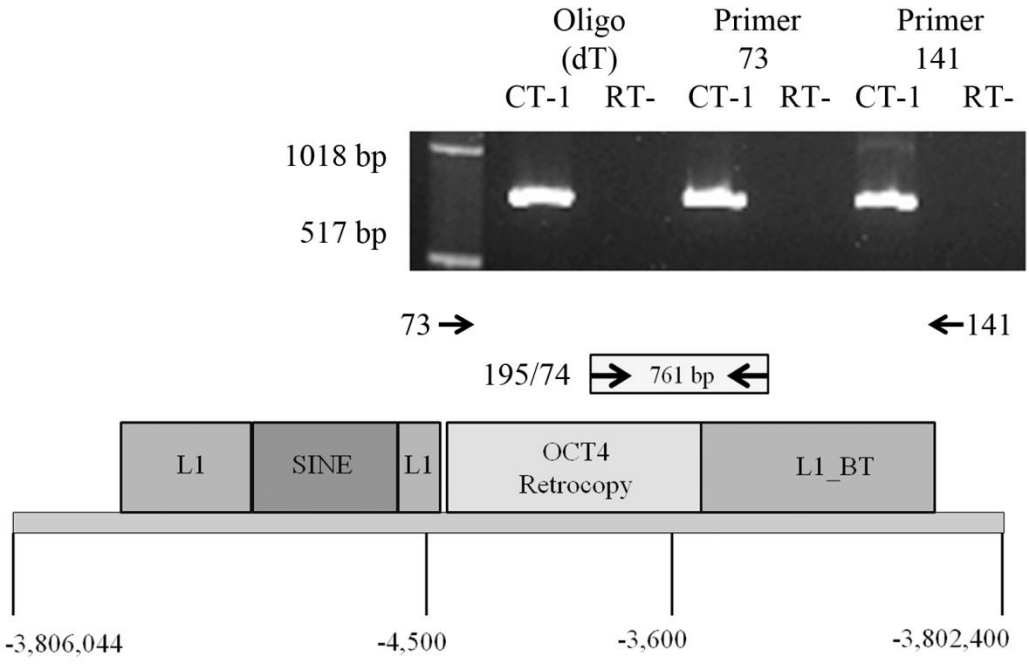



**Figure 4.3.** Transcripts containing the *OCT*4 retrocopy are transcribed from both the positive and negative DNA strands in CT-1 cells. Primers specific to regions flanking the *OCT*4 retrocopy/L1_BT fusion sequence were used as locus specific primers for reverse transcription of first strand cDNAs. Both primers do not bind to any retroelements. Primer 141 was used to synthesize cDNA from sense RNA and primer 73 was used to synthesize cDNA from antisense RNA. Primers 195 and 74 were then used to amplify the *OCT4* retrocopy/L1_BT sequence. No genomic DNA contamination was amplified from RT- negative controls for each cDNA. cDNA reverse transcribed using oligo(dT) primers served as a positive template control. All bands were sequenced. The negative DNA strand and approximate primer positions are represented in the schematic.

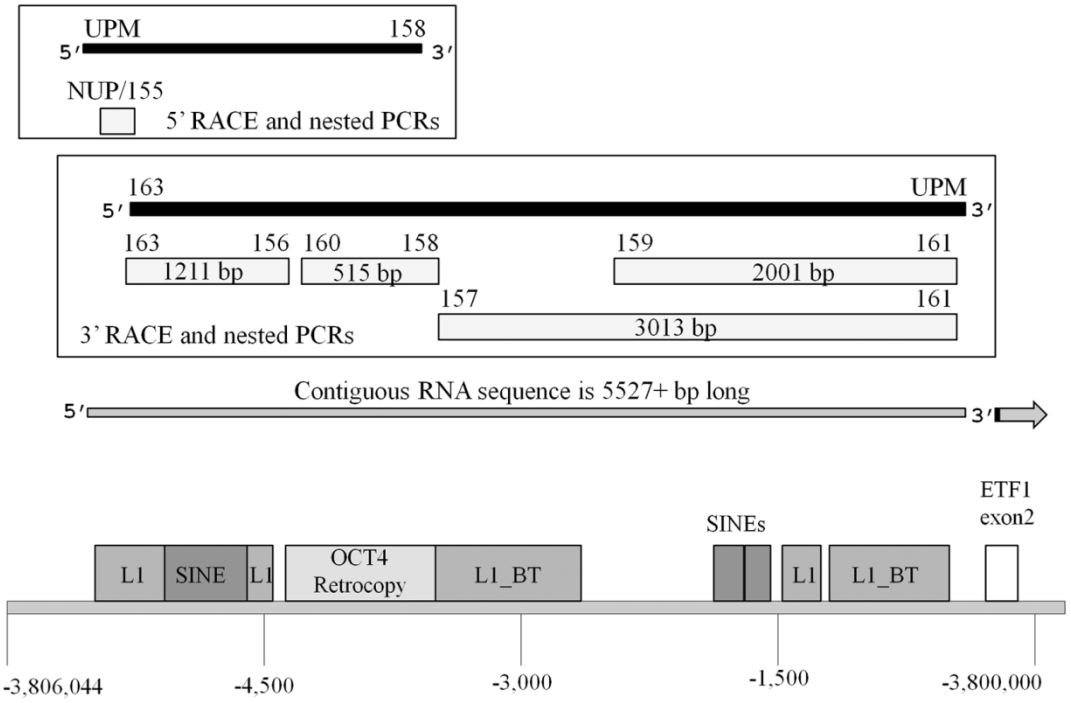



**Figure 4.4.** Schematic representation of RT-PCR products amplified from 5' and 3' RACE CT-1 cDNAs synthesized from sense RNA of *ETF1* intron 1 region. 3' RACE first strand synthesis was primed with modified oligo(dT) 3'-RACE CDS Primer A, and 5' RACE first strand synthesis was primed with modified oligo(dT)5'-RACE CDS Primer A and SMARTer II A oligonucleotide. These cDNAs were used in first round PCR reactions using Universal Primer Mix (UPM) and a locus specific primer. Single PCR products were never obtained but purified PCR reactions were used as template (black rectangle) for secondary nested PCR using either 2 locus specific primers or a locus specific primer and an abridged Nested Universal primer (NUP) to obtain sequence ends. Specific amplicons (light gray rectangles) are indicated by primers used, length, and approximate position within *OCT4* retrocopy region of *ETF1* intron 1. Overlapping PCR fragments were then used to construct a contiguous RNA sequence over 5527 bp long (dark gray rectangle). While a 5' end was successfully sequenced, the 3' end remained unidentified. No genomic DNA contamination was detected in any control RT- reactions. A minimum of two PCR fragments cloned into pGEM-T Easy were sequenced per PCR reaction.

Similar results were obtained in RACE of antisense transcripts (Fig. 4.5). Fragments amplified by 5' RACE overlapped with the second exon of *ETF1*. However, these sequences exhibited some type of RNA splicing as fragments aligned with genomic sequence lacked internal sequence. Similar results were obtained whenever PCR was performed using locus primers 160 and 161, which would presumably amplify a continuous fragment 3.5 Kb in length. The splicing phenomenon may be an in vitro artifact, as this region of the RNA is enriched in repetitive retroelements and may possess high levels of secondary structure. Interestingly, each splicing event involved recombining at short repeats (Sup. Fig. 8). Fragments corresponding to the 3' end of the antisense transcript were amplified. However it was found that the Universal Primer Mix RACE primer containing the d (T) stretch was hybridizing to a stretch of adenosine monophosphates within the sequence and not to a true poly (A) tail of mRNA. If all successfully amplified fragments were transcribed from the same cDNA, then the antisense RNA would be over 5239 bp in length.

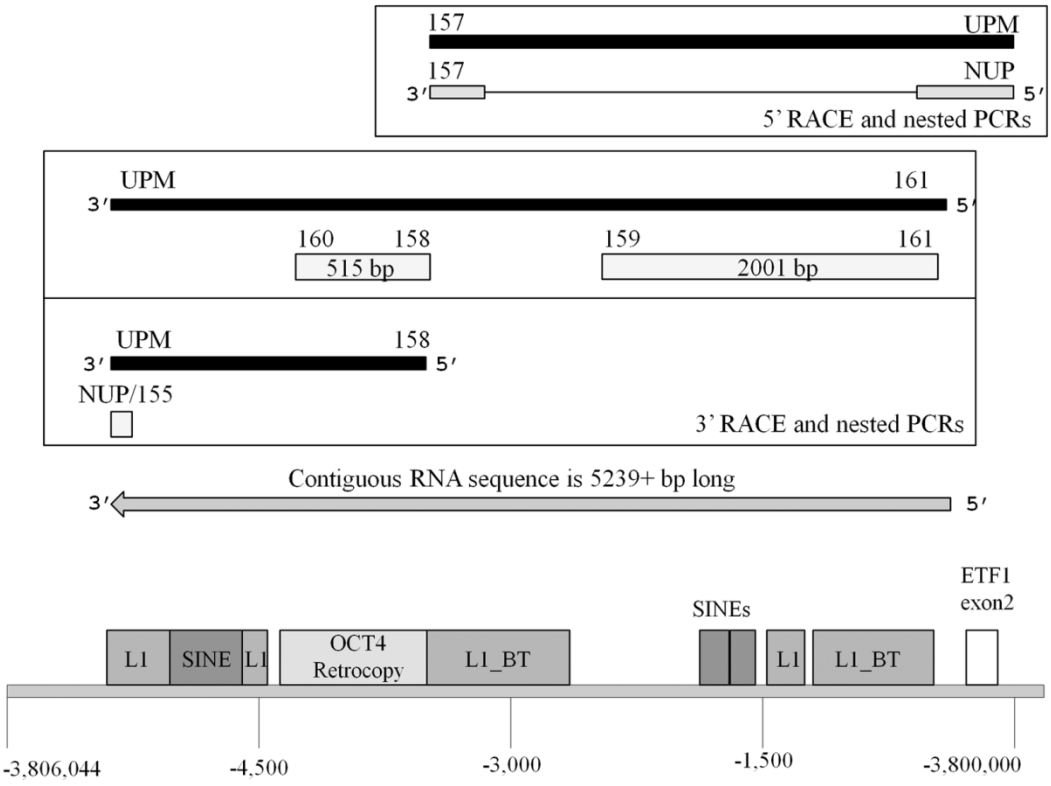



**Figure 4.5.** Schematic representation of RT-PCR products amplified from 5’ and 3’ RACE CT-1 cDNAs synthesized from antisense RNA of ETF1 intron 1 region. 5‘and 3’ RACE first strand synthesis were primed as described in Fig. 4. Single PCR products were never obtained but purified PCR reactions were used as template (black rectangle) for secondary nested PCR using either 2 locus specific primers or a locus specific primer and an abridged Nested Universal primer (NUP) to obtain sequence ends. Specific amplicons (light gray rectangles) are indicated by primers used, length, and approximate alignment with the OCT4 retrocopy region of ETF1 intron 1. A contiguous RNA sequence over 5239 bp long (dark gray rectangle) was constructed that is complementary to the 5527 bp sequence in Fig. 4. The 5’ end was successfully sequenced. However, a cDNA break/ligation splicing phenomenon was observed in the PCR fragment that omitted a few thousand base pairs. The 3’ RACE fragment was a result of the oligo (dT) primer hybridizing to a stretch of A’s within sequence rather than to a post transcriptionally modified poly A tail. No genomic DNA contamination was detected in any RT- reactions. A minimum of two PCR fragments cloned into pGEM-T Easy were sequenced per PCR reaction.

**Discussion**

In this study we demonstrate that long sense transcripts and natural antisense transcripts (NATs) are expressed from the bovine *ETF1* intron 1 that contain a conserved yet truncated *OCT4* and only lacking in a complete fifth exon. At the 3' end, the *OCT4* retrocopy is fused with a 5' truncated L1 BT suggesting that a chimeric retrocopy was formed during retrotransposition. Compared to other known bovine *OCT4* retrocopies, the chromosome 7 retrocopy is the least mutated with only 10 substitutions and no insertions or deletions (Sup. Fig. 5). RT-PCR performed using primers universal to all *OCT4* retrocopies produced only OCT4 and chromosome 7 OCT4 retrocopy sequences. Therefore, *OCT4* retrocopies on chromosomes 3, 4, and 21 are most likely processed pseudogenes or at least not expressed in CT-1 cells. While the low degree of mutation in the chromosome 7 *OCT4* retrocopy may suggest that the *OCT4*/ L1 BT retrotransposition into chromosome 7 is a relatively recent event, it also suggests that the conservation of this sequence may be critical for its gene product function.

Advances in transcriptome analysis have shown that noncoding RNA transcripts are often expressed and interspersed or even overlapped with other transcribed RNAs, and can be transcribed from intergenic and intragenic regions including introns (Mercer et al., 2009). Our results demonstrate that many kilobases of DNA flanking the *OCT4* retrocopy are transcribed from both strands. Our 5' and 3' RACE experiments did not consistently identify 5' and 3' RNA ends, which may be due to RNA secondary structure resulting from hybridization of cis retroelements that are part of the same long noncoding RNA. We were successful in consistently obtaining 5' and 3' ends of our control genes, *ETF1* and *OCT4*. It is also possible that multiple overlapping transcripts are expressed in

this region resulting in the creation of many cDNAs instead of one long noncoding RNA as represented by our contiguous sequences in Fig 4.4. and Fig. 4.5.

The mechanism for transcription of the sense and NAT RNAs requires further study. Transcription of NATs can be driven by acquiring regulatory elements from nearby retroelements (Conley et al., 2007). Transcription may also result from a "ripple effect" caused by rapid and strong induction of neighboring genes. Miki Ebisuya and colleagues found that rapid FGF mediated induction of "immediate early gene" *Erg1* caused upregulation of noncoding RNAs in neighboring intergenic regions as well as intragenic regions including intron 1 of *Etf1* (Ebisuya et al., 2008). There is shared synteny at this chromosomal locus between bovine and human genomes.

Possible functions of the sense and antisense *OCT4* retrocopy containing RNAs can be drawn from examples of other identified long noncoding RNAs and transcribed retrocopies (Mercer et al., 2009). Recently, it has been demonstrated that sense transcripts and NATs corresponding to gene exon regions are processed into smaller RNAs that act as endogenous siRNAs or endo-siRNAs (Werner et al., 2009). These siRNAs were once thought to induce silencing of transcribed pseudogenes and reduce transcriptional noise. Conversely, transcribed retrocopies have also been demonstrated to act as a source of NATs that hybridize with sense mRNAs and form double stranded RNAs (dsRNAs) that undergo DICER/ AGO2 mediated processing (Tam et al,. 2008; Watanabe et al., 2008). These endo-siRNAs are used for RNA interference of the parental gene transcripts. Furthermore, dsRNAs can also originate from transcription of both DNA strands at the same loci to produce endo-siRNAs. These findings suggest a possible mechanism for our *OCT4* retrocopy expression results.

OCT4 is a key regulator of pluripotency in early mouse embryonic development and is downregulated as cells commit to the trophectoderm lineage at the blastocyst stage (Nichols et al., 1998). Studies in mouse ESC show that OCT4 must be downregulated for differentiation of ESC into trophoblast cells to occur (Niwa et al., 2000). OCT4 is not rapidly downregulated in the trophectoderm in bovine blastocysts, but expressed into the early stages of trophoblast elongation (Degrelle at al, 2005; Roberts et al., 2004). OCT4 protein is also detected in the bovine trophectoderm derived CT-1 cell line (Schiffmacher and Keefer, 2010b). While the function of OCT4 in the early bovine trophoblast is unknown, OCT4 expression may be tightly regulated to ensure aberrant OCT4 levels do not interfere with the trophoblast transcriptional network. This study demonstrated that *OCT4* retrocopy-containing RNAs are transcribed from both DNA strands at the retrocopy locus within the first intron of *ETF1* in CT-1 cells. *OCT4* retrocopy-containing dsRNAs originating from these sense transcripts and NATs, or dsRNAs originating from NATs (acting in trans) and *OCT4* mRNA, could be further processed into endo-siRNAs to regulate *OCT4* expression. It is unknown whether these RNAs are expressed in vivo or if they are restricted to the bovine trophoblast. Expression analyses in bovine embryos similar to this study in CT-1 cells, as well as analyses of small RNA libraries derived from CT-1 cells and bovine embryos, could provide insight into this putative bovine specific, post-transcriptional mechanism of *OCT4* regulation.

# Chapter 5: Future directions

The results of this research demonstrate for the first time that the embryonic bovine trophectoderm derived CT-1 cell line can be used as a genetic model to study key regulators at the core of the trophectoderm transcriptional hierarchy. Our success in obtaining acceptable transfection efficiencies for both plasmid and oligonucleotide delivery in CT-1 cells allowed us to perform CDX2 and OCT4 overexpression and knockdown experiments to determine their roles in regulating other trophoblast specific regulators. Using qRT-PCR, we were able to identify potential target genes of CDX2. We detected changes in several gene levels in response to manipulation of CDX2 including *IFNT*, *ETS2*, *ELF*5, *HAND1*, *SOX15*, *OCT4*, and *MASH2* (Fig. 5.1). However, no significant changes in candidate gene expression in response to OCT4 overexpression or knockdown were detected, other than OCT4 autoregulation.

During our analysis of *OCT4* expression we unexpectedly identified a transcribed *OCT4* retrocopy expressed in CT-1 cells. Further experiments in the characterization of this transcribed retrocopy should include an expression profile analysis in bovine embryos. Also, Northern blot analysis would help elucidate discrete transcript sizes for both sense and antisense transcripts. Antisense transcript specific targeting by siRNA followed by qRT-PCR for *OCT4* levels would possible identify an *OCT4* endo-siRNA mechanism in place.

The overall results and achievements of this study have provided a rationale for an even more thorough analysis of the bovine trophoblast transcriptome. We have recently performed whole transcriptome-wide analyses of cDNAs of CT-1 cells, in vitro

bovine embryos and TE from in vitro derived embryos (8 dpc) and in vivo derived bovine embryos and TE from in vivo derived embryos (14 dpc) using Illumina digital gene expression (DGE): tag profiling. This unbiased approach will be used to identify and quantify genome-wide expression of transcripts, and will provide a more powerful and accurate assessment of the CT-1 transcriptome in comparison to in vivo and in vitro produced embryos. This study may also accurately identify the noncoding RNAs containing the *OCT4* retrocopy.

While experiments in this study were all based on transient transfections, stable CT-1 lines containing integrated doxycycline inducible *OCT4* or *CDX2* cassettes would prove an invaluable tool for future experiments. In the present study, single vector doxycycline inducible systems were constructed to overexpress hemagglutinin (HA) tagged OCT4 or MYC tagged CDX2 with the intent to create stably integrated CT-1 cell lines that can overexpress these recombinant proteins at different levels. The single vector doxycycline inducible system was initially tested and proved very sensitive to increasing doxycycline concentrations in transiently transfected HEK 293 cells (Sup. Fig. 2). The desire to tightly control recombinant protein levels was based on previous findings that lineage maintenance and differentiation is gene dosage dependent (Niwa et al., 2000). The recombinant proteins expressed from these vectors were tagged with the intent to use the tags for antibody pulldown in co-immunoprecipitations.

While our current transfection efficiency in CT-1 cells was acceptable for gene expression analysis by qRT-PCR, a higher efficiency or stably transfected cell lines would be necessary for adequate pull down of protein complexes to study binding of co-factors. Stable CT-1 cell lines expressing doxycycline inducible OCT4 or CDX2 would

also be needed for determining the effects of overexpression on CT-1 proliferation and differentiation. A disadvantage of the CT-1 cell line is their slow proliferation rate. Nevertheless, the creation of stably transfected CT-1 cell lines is feasible. With a 9% transfection rate and adequate amounts of Lipitoid, stable transfected CT-1 cells could be identified using a fluorescent reporter and clonally expanded by patching cells similar to methods used in embryo explant culture (Pant et al., 2008). Alternatively, our collagenase-dispase protocol could be used to dissociate transfected CT-1 colonies and seed new plates with aggregates as small as 10 cells. If enough CT-1 colonies are transfected, this passaging method could be used with antibiotic selection, although CT-1 cells are slow to respond to antibiotic selection. Then other approaches such as promoter reporter analyses of HAND1, MASH2, and SOX15 to determine CDX2 DNA binding and activation in their promoters as well as co-immunoprecipitations to identify potential protein interactions could be taken. While our results did not indicate a specific role for OCT4 in the bovine trophectoderm, identification of potential OCT4 cofactors, such as SOX15 which we have shown to be expressed in CT-1 cells, may provide clues to the function, if any, of OCT4 in bovine TE.

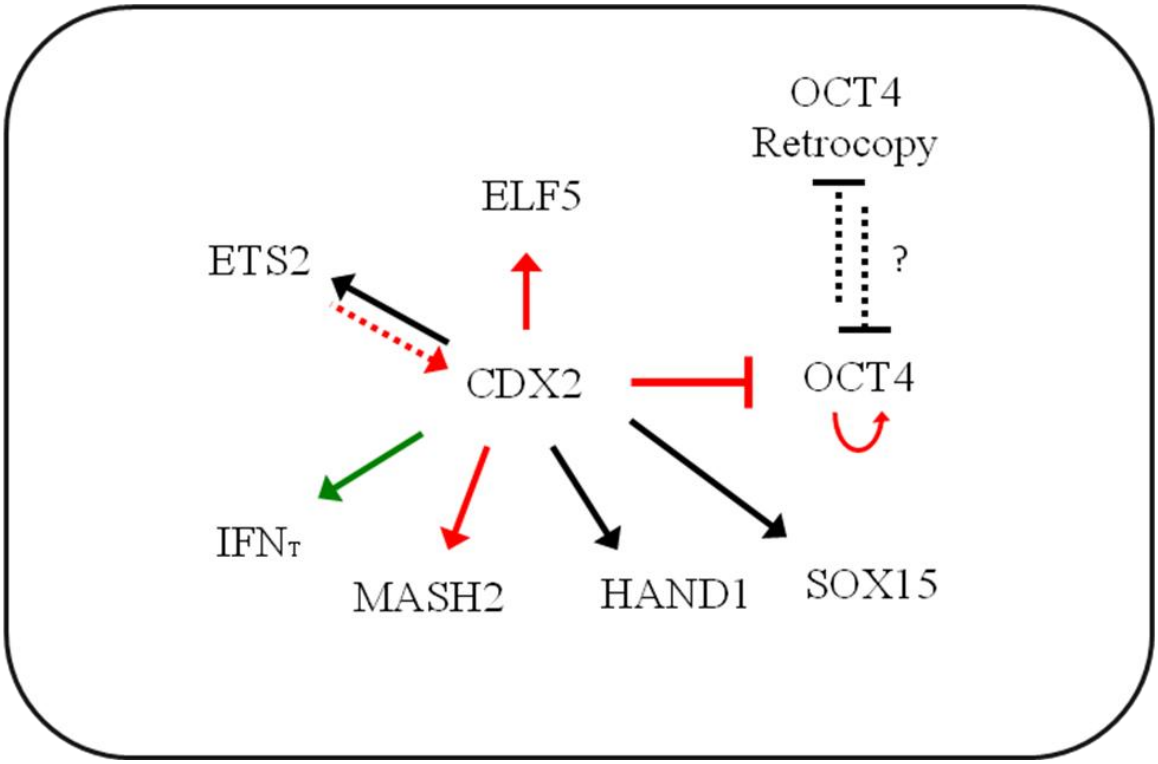
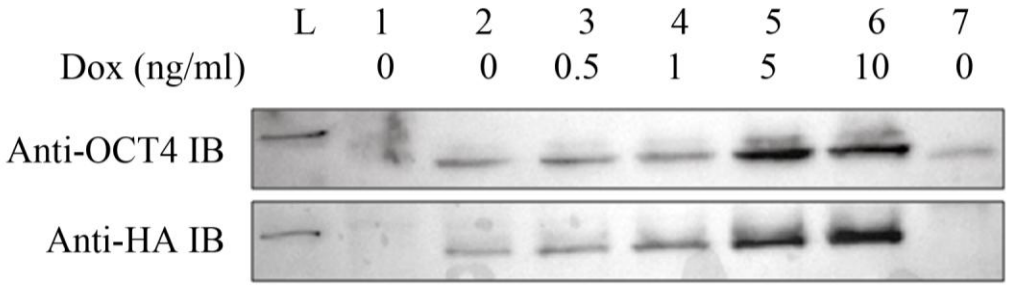
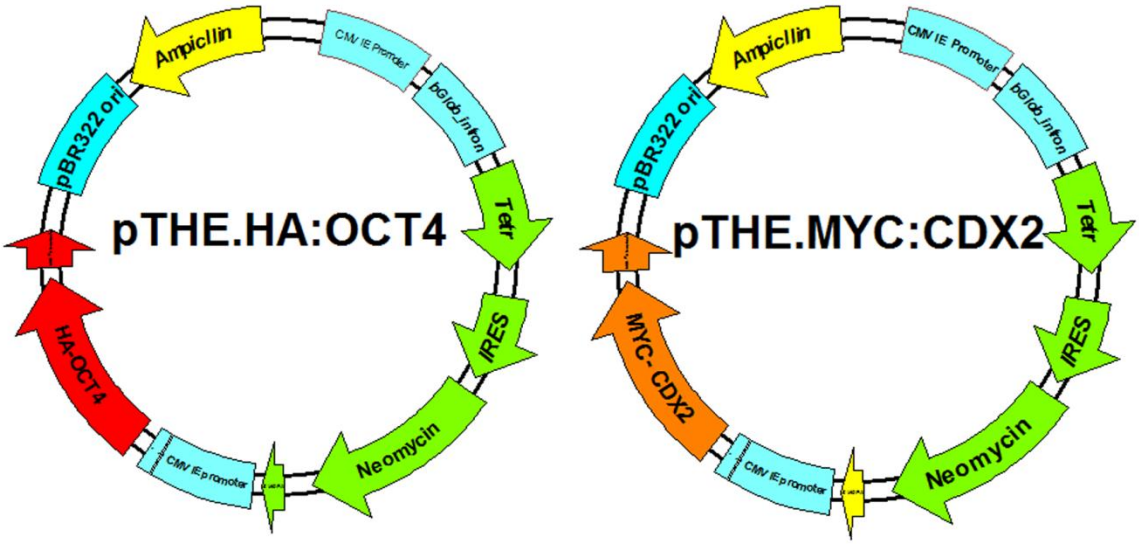



**Figure 5.1.** Model incorporating the results of this study to form a transcriptional regulatory network in CT-1 cells. Black arrows indicate results of this study that have not been previously elucidated in bovine or mouse models. Dashed arrows indicate speculative regulatory interactions that are based on our findings. Red arrows indicate regulatory mechanisms previously determined to be in conserved in mouse trophoblast stem cell maintenence. The green arrow indicates a similar finding from a parallel study (Sakurai et al., 2009).

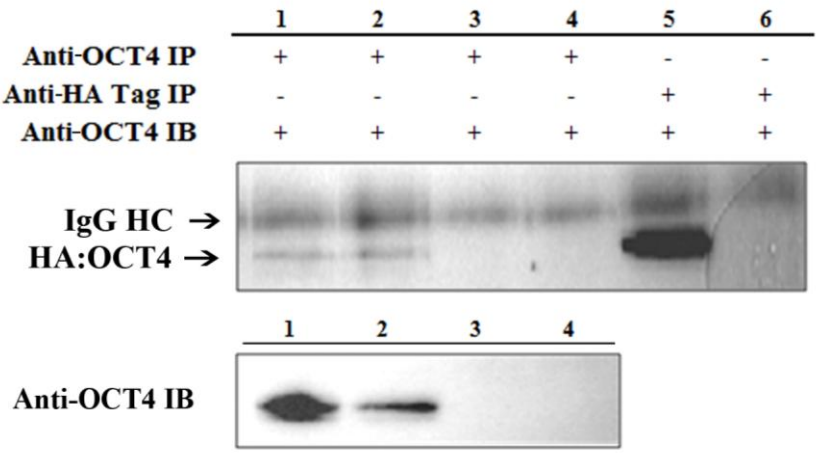
# Appendix



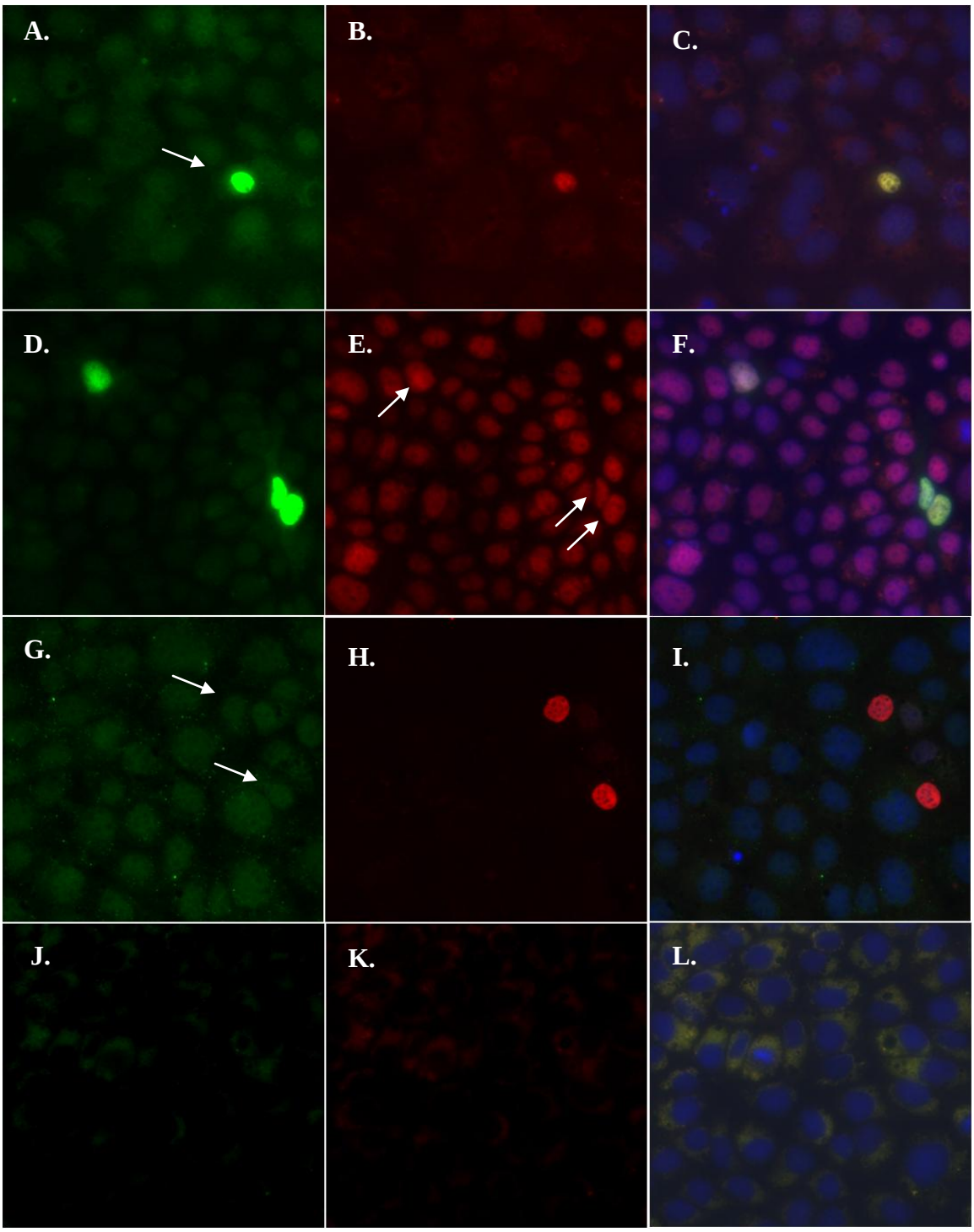

**Sup. Figure 1.** The doxycycline responsive promoter of pTHE.HA:OCT4 is sensitive at increasing doxycycline concentrations (ng per ml). HEK 293 cells were transiently transfected and treated for 48 hr with increasing concentrations of doxycycline in the medium (Lanes 2-6). 75 µg of whole cell lysate was loaded per well. Untransfected HEK 293 lysate serves as negative control (Lane 1) while NTERA lysate serves as an OCT4 positive control for OCT4 only (Lane 7). Both HA and OCT4 antibodies successfully detect HA tagged recombinant OCT4. L = ladder, DOX = doxycycline.




**Sup. Figure 2**. Plasmid schematics of pTHE.HA:OCT4 and pTHE.MYC:CDX2. Both plasmids utilize the pTHE plasmid (Jiang et al., 2001) to express doxycycline inducible, epitope tagged proteins. The pTHE plasmid (Addgene plasmid 12512) constitutively expresses a chimeric msin3 domain/ tetracycline repressor fusion protein that inhibits transcription at the tetracycline responsive promoter located on the same plasmid in the absence of tetracycline. HA epitope tagged OCT4 and MYC tagged CDX2 were cloned in the multiple cloning site located 3' to the tet responsive promoter. Many variations of these two plasmids have been constructed including plasmids containing C terminal HA epitope tagged OCT4. The neomycin ORF has also been substituted with visual selection marker ORFs including histone B2-GFP fusion protein, nuclear GFP, and nuclear dsRED2-1. However, IRES mediated translation of these coding sequences did not sufficiently express observable levels of protein fluorescence, even with high copy transfection rates in HEK 293 cells.



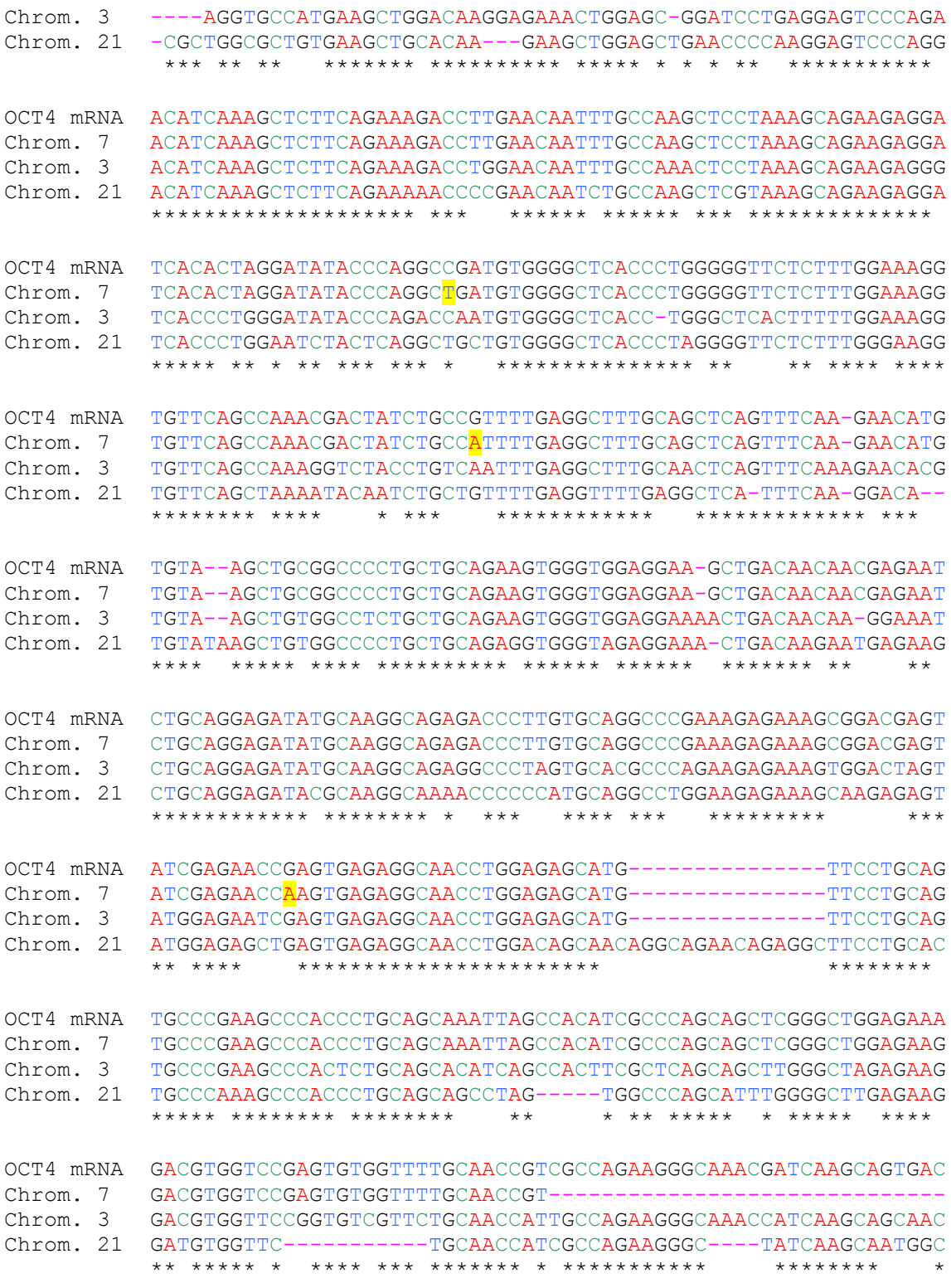

**Sup. Figure 3.** Comparison of recombinant bovine HA tagged OCT4 binding affinities between anti human OCT4 and anti HA antibodies. Efforts to pull down bovine OCT4 from CT-1 cells were unsuccessful even though a range of buffers varying in stringency were tried. To determine whether the issue is due to insufficient OCT4 protein or antibody specificity, immunoprecipitations were conducted on HEK 293 cells transiently transfected with doxycycline inducible vector expressing HA epitope tagged OCT4. HEK 293 cells were transfected with pTHE.HA:OCT4 and induced for 48 hours with 10 ng per ml (lane 2) and 100 ng per ml doxycycline (lanes 1 and 6). Untransfected HEK 293 cells (lane 4) and HEK 293 cells transfected with empty pTHE vector and induced with 100 ng per ml doxycycline (lane 3 and 6) served as negative controls. Immunoprecipitations (IP) were performed with 750 µg protein and 1 µg of anti-OCT4 antibody (lanes 1-4) or anti-HA antibody (lanes 5-6). Eluted samples were immunoblotted (IB) with anti-OCT4 antibody. An OCT4 immunoblot with 75 µg (10%) of original protein lysates served as a loading control.

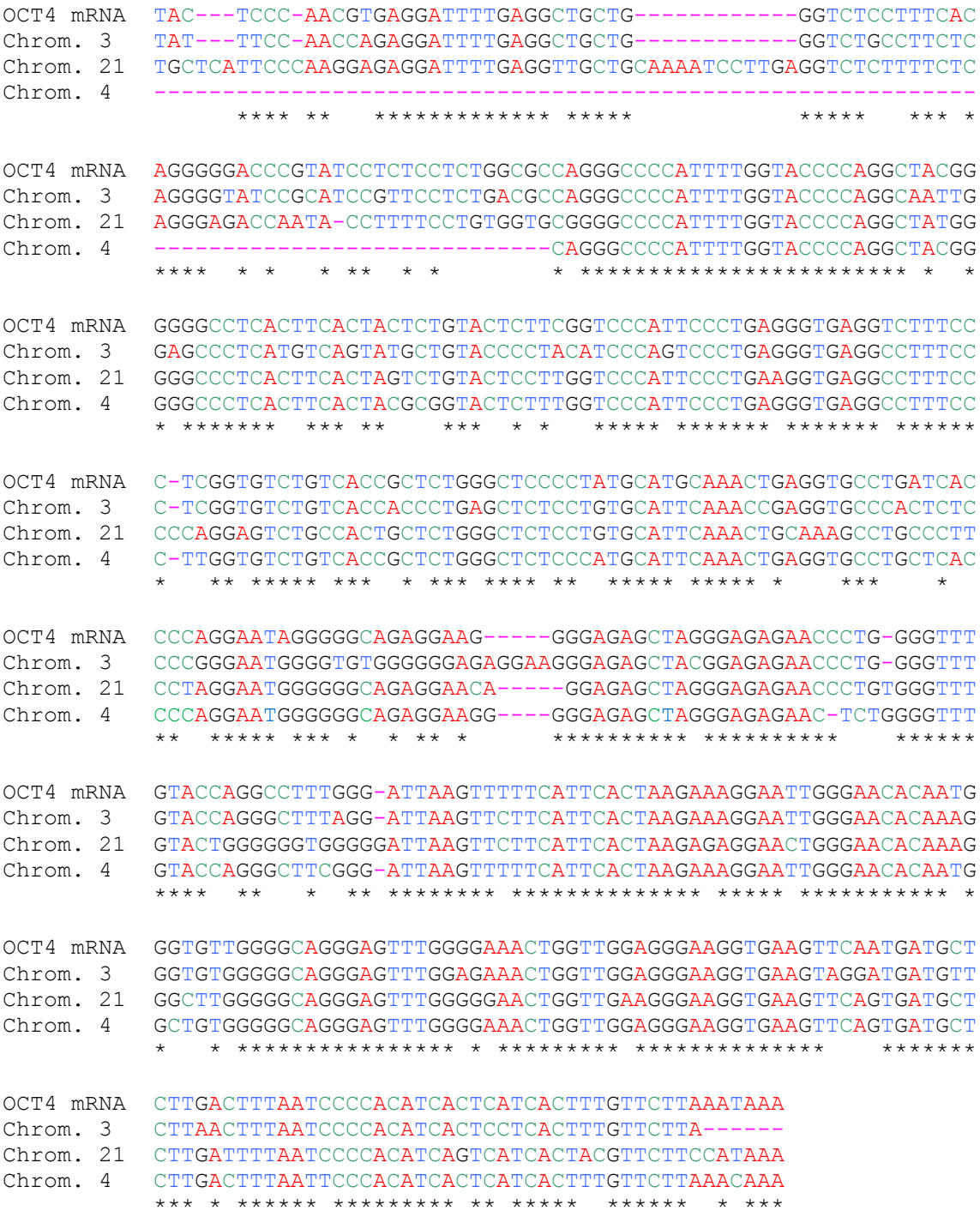

A.
B.
C.
D.
E.
F.
G.
H.
I.
J.
K.
L.

**Supp. Figure 4.** Double label immunocytochemistry to detect both doxycycline-induced proteins and endogenous proteins. Because our antibodies for the HA epitope and endogenous CDX2 were unsuitable together for double label immunocytochemistry, the polyclonal anti-human OCT4 antibody was utilized to detect recombinant HA:OCT4 protein for the following experiments. The anti-human OCT4 antibody detected both low levels of endogenous OCT4 and overexpressed recombinant OCT4 (A, white arrow). The overexpression of OCT4 was localized to the same cells that were immunolabeled with anti-HA antibody (B). CT-1 cells overexpressing HA:OCT4 (D) did not show any reduction of endogenous CDX2 protein (E, white arrows) when compared to levels in untransfected neighboring cells expressing CDX2. Similarly, CT-1 cells overexpressing MYC:CDX2 protein (H, white arrows), did not show any observable reduction in anti-human OCT4 labeling (G, white arrows) when compared to levels in untransfected neighboring cells. Images J and K are controls for nonspecific immunofluorescence of secondary antibodies in experiments performed without primary antibodies. Images C, F, I, and L are merged images of the 2 preceding images with Hoëchst DNA staining.

**Supp. Figure 5.** Alignment of bovine *OCT4* retrocopies with full length *OCT4* mRNA (NM 174580). The translational start site is boxed. Single nucleotide polymorphisms between the chromosome 7 retrocopy and *OCT4* mRNA are highlighted in yellow.

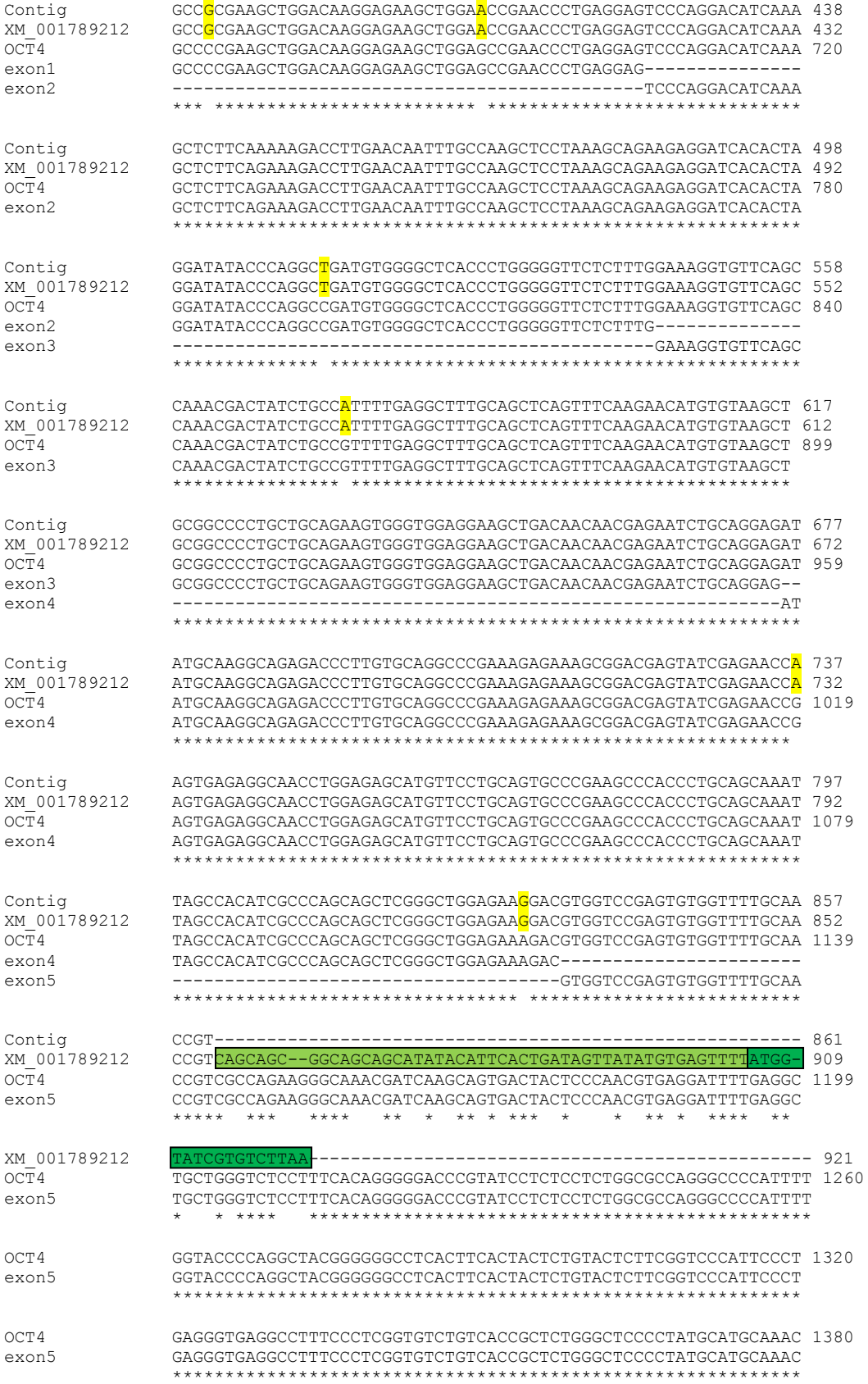
```
OCT4 mRNA   GGAGCTGGAAGTGAAGGCCCGCATGGGGGACCTGCACCGAGAGTCTAGGAGTCTGGGGG

OCT4 mRNA   TGGAGAGGGGCCTGGGTGGAGATCCCTGGCTTTCCCCTTCCAGACACCACCGCCACCAGC

OCT4 mRNA   AGGCAAACACCCTCCGCCTCAGTTTCTCCCACCCCCACGGTCCCTTCCCCCCACCCATCC

OCT4 mRNA   AGGGGGCGGGGCCAGAGGTCAAGGCTAGTGGGTGGGATTGGGGAGGGAGAGAGGTGTTGA

OCT4 mRNA   GCAGTCTCTAGGAGATCCCTCGTTTTCCTAGGCCCCCGGCTCGGGGTGCCTTCCTTCCCC
Chrom. 7    ----------------------------------------GGGGTGCCTTCCTTCCCC
Chrom. 3    ------------------------------------------------------------
Chrom. 21   --AGTCCCTTGAAGAGCTCTCATTTTCCTAGGCCCTGGGCTTGGGGCACATTCCTCCCCC
              **** ** * *** * *** *************  **** ****  * ***** ****

OCT4 mRNA   ATGGCGGGACACCTCGCTTCTGACTTCGCCTTCTCGCCCCCGCCGGGCGGTGG-AGGCGA
Chrom. 7    ATGGCGGGACACCTCGCTTCTGACTTCGCCTTCTCGCCCCCGCCGGGCGGTGG-AGGCGA
Chrom. 3    ----------------------------GCCTTCTCGCCCCTGCTGGGCAGTGG-AGGCCA
Chrom. 21   ACAGCGGGACACCTGGCTTCTGACTTTGCCTTCTCACCCCGGC-ACATGGTGGCAGGCGA
            *  *********** *********** ******** **** **      **** **** *

OCT4 mRNA   TGGGCCGGGAGGGCCAGAGCCGGGCTGGGTTGATCCTCGGACCTGGATGAGCTTCCAAGG
Chrom. 7    TGGGCCGGGAGGGCCAGAGCCGGGCTGGGTTGATCCTCGGACCTGGATGAGCTTCCAAGG
Chrom. 3    TCGGCAGGGAGGGCCACAGCCAGGCTGGGTTGATCCTCGGACCTGGATGAGCTTCCAAGG
Chrom. 21   CAGGCTGAGAAGGCTGAAGCCAGGCTAGG----------AACCTGGATGAGCTTCCAGGG
              *** * ** ***   **** **** ************ ***************** **

OCT4 mRNA   GCCTCCCGGTGGGTCGGGGATCGGGCCGGGGGTTGTGCCTGGCGCCGAGGTGTGGGGGCT
Chrom. 7    GCCTCCCGGTGGGTCGGGGATCGGGCCGGGGGTTGTGCCTGGAGCCGAGGTGTGGGGGCT
Chrom. 3    GCCTCCGGGTGGGTCAGGAATCTGGCTGGAGCTTGTGCCCAGCAC-GAGGTGTGGGGGCT
Chrom. 21   GCTTCCT--TGGGTGGGAGATGGGGCTGGGGGTTGGGCCTGGTGCTCAGGTGTTGGGGCT
            ** *** *******  *  **  *** ** * *** ***  *     ****** ******

OCT4 mRNA   TCCCCCGTGCCCCCCGCCCTATGACTTGTGTGGAGGGATGGCCTACTGTGCGCCGCAGGT
Chrom. 7    TCCCCCGTGCCCCCCGCCCTATGACTTGTGTGGAGGGATGGCCTACTGTGCGCCGCAGGT
Chrom. 3    TCCCCTGTGCCCCC-ACCCTATGACTTGTGCAGAGGGATGGCCTGCTGTGCACCGCAGGT
Chrom. 21   GACCCTGTGCTCTT-GCCCTATGACCTCTGTGGAGGAACCGCC---TGTGCCCCTCAGGT
              *** **** *    ********* * **  **** *  ****  ***** ** *****

OCT4 mRNA   TGGAGTGGGGCCGGTGCCCCCAGGCGGCCTGGAG-ACCCCTCAGCCCGAGGGCGAGGCGG
Chrom. 7    TGGAGTGGGGCCGGTGCCCCCAGGCGGCCTGGAG-ACCCCTCAGCCCGAGGGCCAGGCGG
Chrom. 3    TAGAGTGGGGCTGGTGCCCCAAGGTGTCCTGGAGGGTGCCTCAGCCCGAGGGCGAGGCGG
Chrom. 21   T-GAGTGGAGCGGGTGCCCCAAGGTAGCCCAGAG-ACCCCTCAGCCTGAGGGCGAGACAG
            * ****** ** ******** ***   **  ***    ******** ********* * *

OCT4 mRNA   GAGCCGGGGTGGAGAGCAACTCC------GAGGGGGCCTCCCCGGACCCCTGCGCCGCAC
Chrom. 7    GAGCCGGGGTGGAGAACAACTCC------GAGGGGGCCTCCCCGGACCCCTGCGCCGCAC
Chrom. 3    GAGCCAGGGTGGAAAGCAACTCC------GAGGGGGCCTCCCCTGGACC-----------
Chrom. 21   AAGCTGAGATGGAGAGCAACCCCAGCTCTGAGGGAGCCTCTCCGGAACCCTGCGCC--AC
             ***   * **** * **** **      ***** ***** * *** *********  **

OCT4 mRNA   CCGCAGGCGCCCCGAAGCTGGACAAGGAGAAGCTGGAGCCGAACCCTGAGGAGTCCCAGG
Chrom. 7    TCGCAGGCGCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCAGG
```


Chrom. 3 ----AGGTGCCATGAAGCTGGACAAGGAGAAACTGGAGC-GGATCCTGAGGAGTCCCAGA
Chrom. 21 -CGCTGGCGCTGTGAAGCTGCACAA---GAAGCTGGAGCTGAACCCCAAGGAGTCCCAGG
OCT4 mRNA ACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAGGA
Chrom. 7 ACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAGGA
Chrom. 3 ACATCAAAGCTCTTCAGAAAGACCTGGAACAATTTGCCAAACTCCTAAAGCAGAAGAGGG
Chrom. 21 ACATCAAAGCTCTTCAGAAAAACCCCGAACAATCTGCCAAGCTCGTAAAGCAGAAGAGGA
OCT4 mRNA TCACACTAGGATATACCCAGGCCGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAAGG
Chrom. 7 TCACACTAGGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAAGG
Chrom. 3 TCACCCTGGGATATACCCAGACCAATGTGGGGCTCACC-TGGGCTCACTTTTTGGAAAGG
Chrom. 21 TCACCCTGGAATCTACTCAGGCTGCTGTGGGGCTCACCCTAGGGGTTCTCTTTGGGAAGG
OCT4 mRNA TGTTCAGCCAAACGACTATCTGCCGTTTTGAGGCTTTGCAGCTCAGTTTCAA-GAACATG
Chrom. 7 TGTTCAGCCAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAA-GAACATG
Chrom. 3 TGTTCAGCCAAAGGTCTACCTGTCAATTTGAGGCTTTGCAACTCAGTTTCAAAGAACACG
Chrom. 21 TGTTCAGCTAAAATACAATCTGCTGTTTTGAGGTTTTGAGGCTCA-TTTCAA-GGACA--
OCT4 mRNA TGTA--AGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAA-GCTGACAACAACGAGAAT
Chrom. 7 TGTA--AGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAA-GCTGACAACAACGAGAAT
Chrom. 3 TGTA--AGCTGTGGCCTCTGCTGCAGAAGTGGGTGGAGGAAAACTGACAACAA-GGAAAT
Chrom. 21 TGTATAAGCTGTGGCCCCTGCTGCAGAGGTGGGTAGAGGAAA-CTGACAAGAATGAGAAG
OCT4 mRNA CTGCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGT
Chrom. 7 CTGCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGT
Chrom. 3 CTGCAGGAGATATGCAAGGCAGAGGCCCTAGTGCACGCCCAGAAGAGAAAGTGGACTAGT
Chrom. 21 CTGCAGGAGATACGCAAGGCAAAACCCCCCATGCAGGCCTGGAAGAGAAAGCAAGAGAGT
OCT4 mRNA ATCGAGAACCGAGTGAGAGGCAACCTGGAGAGCATG---------------TTCCTGCAG
Chrom. 7 ATCGAGAACCAAGTGAGAGGCAACCTGGAGAGCATG---------------TTCCTGCAG
Chrom. 3 ATGGAGAATCGAGTGAGAGGCAACCTGGAGAGCATG---------------TTCCTGCAG
Chrom. 21 ATGGAGAGCTGAGTGAGAGGCAACCTGGACAGCAACAGGCAGAACAGAGGCTTCCTGCAC
OCT4 mRNA TGCCCGAAGCCCACCCTGCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAA
Chrom. 7 TGCCCGAAGCCCACCCTGCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAG
Chrom. 3 TGCCCGAAGCCCACTCTGCAGCACATCAGCCACTTCGCTCAGCAGCTTGGGCTAGAGAAG
Chrom. 21 TGCCCAAAGCCCACCCTGCAGCAGCCTAG-----TGGCCCAGCATTTGGGGCTTGAGAAG
OCT4 mRNA GACGTGGTCCGAGTGTGGTTTTGCAACCGTCGCCAGAAGGGCAAACGATCAAGCAGTGAC
Chrom. 7 GACGTGGTCCGAGTGTGGTTTTGCAACCGT----------
Chrom. 3 GACGTGGTTCCGGTGTCGTTCTGCAACCATTGCCAGAAGGGCAAACCATCAAGCAGCAAC
Chrom. 21 GATGTGGTTC----------TGCAACCATCGCCAGAAGGGC----TATCAAGCAATGGC


OCT4 mRNA TAC---TCCC-AACGTGAGGATTTTGAGGCTGCTG----------GGTCTCCTTTCAC
Chrom. 3 TAT---TTCC-AACCAGAGGATTTTGAGGCTGCTG----------GGTCTGCCTTCTC
Chrom. 21 TGCTCATTCCCAAGGAGAGGATTTTGAGGTTGCTGCAAAATCCTTGAGGTCTCTTTTCTC
Chrom. 4 ----------
OCT4 mRNA AGGGGGACCCGTATCCTCTCCTCTGGCGCCAGGGCCCCATTTTGGTACCCCAGGCTACGG
Chrom. 3 AGGGGTATCCGCATCCGTTCCTCTGACGCCAGGGCCCCATTTTGGTACCCCAGGCAATTG
Chrom. 21 AGGGAGACCAATA-CCTTTTCCTGTGGTGCGGGGCCCCATTTTGGTACCCCAGGCTATGG
Chrom. 4 ----------CAGGGCCCCATTTTGGTACCCCAGGCTACGG
OCT4 mRNA GGGGCCTCACTTCACTACTCTGTACTCTTCGGTCCCATTCCCTGAGGGTGAGGTCTTTCC
Chrom. 3 GAGCCCTCATGTCAGTATGCTGTACCCCTACATCCCAGTCCCTGAGGGTGAGGCCTTTCC
Chrom. 21 GGGCCCTCACTTCACTAGTCTGTACTCCTTGGTCCCATTCCCTGAAGGTGAGGCCTTTCC
Chrom. 4 GGGCCCTCACTTCACTACGCGGTACTCTTTGGTCCCATTCCCTGAGGGTGAGGCCTTTCC
OCT4 mRNA C-TCGGTGTCTGTCACCGCTCTGGGCTCCCCTATGCATGCAAACTGAGGTGCCTGATCAC
Chrom. 3 C-TCGGTGTCTGTCACCACCCTGAGCTCTCCTGTGCATTCAAACCGAGGTGCCCACTCTC
Chrom. 21 CCCAGGAGTCTGCCACTGCTCTGGGCTCTCCTGTGCATTCAAACTGCAAAGCCTGCCCTT
Chrom. 4 C-TTGGTGTCTGTCACCGCTCTGGGCTCTCCCATGCATTCAAACTGAGGTGCCTGCTCAC
OCT4 mRNA CCCAGGAATAGGGGGCAGAGGAAG-----GGGAGAGCTAGGGAGAGAACCCTG-GGGTTT
Chrom. 3 CCCGGGAATGGGGTGTGGGGGGAGAGGAAGGGAGAGCTACGGAGAGAACCCTG-GGGTTT
Chrom. 21 CCTAGGAATGGGGGGCAGAGGAACA-----GGAGAGCTAGGGAGAGAACCCTGTGGGTTT
Chrom. 4 CCCAGGAATGGGGGGCAGAGGAAGG----GGGAGAGCTAGGGAGAGAAC-TCTGGGGTTT
OCT4 mRNA GTACCAGGCCTTTGGG-ATTAAGTTTTTCATTCACTAAGAAAGGAATTGGGAACACAATG
Chrom. 3 GTACCAGGGCTTTAGG-ATTAAGTTCTTCATTCACTAAGAAAGGAATTGGGAACACAAAG
Chrom. 21 GTACTGGGGGGTGGGGGATTAAGTTCTTCATTCACTAAGAGAGGAACTGGGAACACAAAG
Chrom. 4 GTACCAGGGCTTCGGG-ATTAAGTTTTTCATTCACTAAGAAAGGAATTGGGAACACAATG
OCT4 mRNA GGTGTTGGGGGCAGGGAGTTTGGGGAAACTGGTTGGAGGGAAGGTGAAGTTCAATGATGCT
Chrom. 3 GGTGTGGGGGCAGGGAGTTTGGAGAAACTGGTTGGAGGGAAGGTGAAGTAGGATGATGTT
Chrom. 21 GGCTTGGGGGCAGGGAGTTTGGGGGAACTGGTTGAAGGGAAGGTGAAGTTCAGTGATGCT
Chrom. 4 GCTGTGGGGGCAGGGAGTTTGGGGAAACTGGTTGGAGGGAAGGTGAAGTTCAGTGATGCT
OCT4 mRNA CTTGACTTTAATCCCCACATCACTCATCACTTTGTTCTTAAATAAA
Chrom. 3 CTTAACTTTAATCCCCACATCACTCCTCACTTTGTTCTTA------
Chrom. 21 CTTGATTTTAATCCCCACATCAGTCATCACTACGTTCTTCCATAAA
Chrom. 4 CTTGACTTTAATTCCCACATCACTCATCACTTTGTTCTTAAACAAA

**Supp. Figure 6**. CLUSTAL 2.0.10 multiple sequence alignment of contiguous *OCT4* retrocopy sequenced from RT-PCR (fig 2.), predicted *OCT4* retrogene ORF (XM_001789212.1) located on chromosome 7 (locus LOC100138938), and bovine *OCT4* mRNA (Genbank Accession NM_174580). OCT4 exons are also aligned for reference. The shared translational start sites for OCT4 and the predicted *OCT4* retrogene are labeled green. Ten single nucleotide polymorphisms between aligned sequences are highlighted in yellow. Exons 2 and 3 of the predicted *OCT4* retrogene are boxed and shaded green.

```
OCT4            GGAGCTGGAAGTGAAGGCCCGCATGGGGGACCTGCACCGAGAGTCTAGGAGTCTGGGGGC 60
exon1           GGAGCTGGAAGTGAAGGCCCGCATGGGGGACCTGCACCGAGAGTCTAGGAGTCTGGGGCC
                ************************************************************

OCT4            TGGAGAGGGGCCTGGGTGGAGATCCCTGGCTTTCCCCTTCCAGACACCACCGCCACCAGC 120
exon1           TGGAGAGGGGCCTGGGTGGAGATCCCTGGCTTTCCCCTTCCAGACACCACCGCCACCAGC
                ************************************************************

OCT4            AGGCAAACACCCTCCGCCTCAGTTTCTCCCACCCCCACGGTCCCTTCCCCCCACCCATCC 180
exon1           AGGCAAACACCCTCCGCCTCAGTTTCTCCCACCCCCACCGTCCCTTCCCCCCACCCATCC
                ************************************************************

OCT4            AGGGGGCGGGGCCAGAGGTCAAGGCTAGTGGGTGGGATTGGGGAGGGAGAGAGGTGTTGA 240
exon1           AGGGGGCGGGGCCAGAGGTCAAGGCTAGTGGGTGGGATTGGGGAGGGAGAGAGGTGTTGA
                ************************************************************

Contig          -----------------------------------------GGGGTGCCTTCCTTCCCC 18
XM_001789212      ---------------------------------------CCTTCCTTCCCC 12
OCT4            GCAGTCTCTAGGAGATCCCTCGTTTTCCTAGGCCCCCGGCTCGGGGTGCCTTCCTTCCCC 300
exon1           GCAGTCTCTAGGAGATCCCTCGTTTTCCTAGGCCCCCGGCTCGGGGTGCCTTCCTTCCCC
                ************************************************************

Contig          ATGGCGGGACACCTCGCTTCTGACTTNGCCTTCTCGCCCCCGACCGCCGGCGGAGACCAA 78
XM_001789212    ATGGCGGGACACCTCGCTTCTGACTTCGCCTTCTCGCCCCCGCCGGGCGGTGGAGGCGAT 72
OCT4            ATGGCGGGACACCTCGCTTCTGACTTCGCCTTCTCGCCCCCGCCGGGCGGTGGAGGCGAT 360
exon1           ATGGCGGGACACCTCGCTTCTGACTTCGCCTTCTCGCCCCCGCCGGGCGGTGGAGGCGAT
                ************************************************************

Contig          GGGCCGGGAGGGCCAGAGCCGGGCTGGGTTGATCCTCGGACCTGGATGAGCTACCAAGGG 138
XM_001789212    GGGCCGGGAGGGCCAGAGCCGGGCTGGGTTGATCCTCGGACCTGGATGAGCTTCCAAGGG 132
OCT4            GGGCCGGGAGGGCCAGAGCCGGGCTGGGTTGATCCTCGGACCTGGATGAGCTTCCAAGGG 420
exon1           GGGCCGGGAGGGCCAGAGCCGGGCTGGGTTGATCCTCGGACCTGGATGAGCTTCCAAGGG
                ************************************************************

Contig          CCTCCCGGTGGGTCGGGGATCGGGCCGGGGGTTGTGCCTGGAGCCGAGGTGTGGGGGCTT 198
XM_001789212    CCTCCCGGTGGGTCGGGGATCGGGCCGGGGGTTGTGCCTGGAGCCGAGGTGTGGGGGCTT 192
OCT4            CCTCCCGGTGGGTCGGGGATCGGGCCGGGGGTTGTGCCTGGCGCCGAGGTGTGGGGGCTT 480
exon1           CCTCCCGGTGGGTCGGGGATCGGGCCGGGGGTTGTGCCTGGCGCCGAGGTGTGGGGGCTT
                ****************************************** ******************

Contig          CCCCCGTGCCCCCCGCCCTATGACTTGTGTGGAGGGATGGCCTACTGTGCGCCGCAGGTT 258
XM_001789212    CCCCCGTGCCCCCCGCCCTATGACTTGTGTGGAGGGATGGCCTACTGTGCGCCGCAGGTT 252
OCT4            CCCCCGTGCCCCCCGCCCTATGACTTGTGTGGAGGGATGGCCTACTGTGCGCCGCAGGTT 540
exon1           CCCCCGTGCCCCCCGCCCTATGACTTGTGTGGAGGGATGGCCTACTGTGCGCCGCAGGTT
                ************************************************************

Contig          GGAGTGGGGCCGGTGCCCCCAGGCGGCCTGGAGACCCCTCAGCCCGAGGGCCAGGCGGGA 318
XM_001789212    GGAGTGGGGCCGGTGCCCCCAGGCGGCCTGGAGACCCCTCAGCCCGAGGGCCAGGCGGGA 312
OCT4            GGAGTGGGGCCGGTGCCCCCAGGCGGCCTGGAGACCCCTCAGCCCGAGGGCGAGGCGGGA 600
exon1           GGAGTGGGGCCGGTGCCCCCAGGCGGCCTGGAGACCCCTCAGCCCGAGGGCGAGGCGGGA
                ************************************************************

Contig          GCCGGGGTGGAGAACAACTCCGAGGGGGCCTCCCCGGACCCCTGCGCCGCACTCGCAGGC 378
XM_001789212    GCCGGGGTGGAGAACAACTCCGAGGGGGCCTCCCCGGACCCCTGCGCCGCACTCGCAGGC 372
OCT4            GCCGGGGTGGAGAGCAACTCCGAGGGGGCCTCCCCGGACCCCTGCGCCGCACCCGCAGGC 660
exon1           GCCGGGGTGGAGAGCAACTCCGAGGGGGCCTCCCCGGACCCCTGCGCCGCACCCGCAGGC
                ************* ************************************* *******
```


Contig GCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCAGGACATCAAA 438
XM_001789212 GCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCAGGACATCAAA 432
OCT4 GCCCCGAAGCTGGACAAGGAGAAGCTGGAGCCGAACCCTGAGGAGTCCCAGGACATCAAA 720
exon1 GCCCCGAAGCTGGACAAGGAGAAGCTGGAGCCGAACCCTGAGGAG----------
exon2 ----------TCCCAGGACATCAAA
Contig GCTCTTCAAAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAGGATCACACTA 498
XM_001789212 GCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAGGATCACACTA 492
OCT4 GCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAGGATCACACTA 780
exon2 GCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAGGATCACACTA
Contig GGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAAGGTGTTCAGC 558
XM_001789212 GGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAAGGTGTTCAGC 552
OCT4 GGATATACCCAGGCCGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAAGGTGTTCAGC 840
exon2 GGATATACCCAGGCCGATGTGGGGCTCACCCTGGGGGTTCTCTTTG----------
exon3 ----------GAAAGGTGTTCAGC
Contig CAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACATGTGTAAGCT 617
XM_001789212 CAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACATGTGTAAGCT 612
OCT4 CAAACGACTATCTGCCGTTTTGAGGCTTTGCAGCTCAGTTTCAAGAACATGTGTAAGCT 899
exon3 CAAACGACTATCTGCCGTTTTGAGGCTTTGCAGCTCAGTTTCAAGAACATGTGTAAGCT
Contig GCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCTGCAGGAGAT 677
XM_001789212 GCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCTGCAGGAGAT 672
OCT4 GCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCTGCAGGAGAT 959
exon3 GCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCTGCAGGAG--
exon4 ----------AT
Contig ATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTATCGAGAACCA 737
XM_001789212 ATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTATCGAGAACCA 732
OCT4 ATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTATCGAGAACCG 1019
exon4 ATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTATCGAGAACCG
Contig AGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCTGCAGCAAAT 797
XM_001789212 AGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCTGCAGCAAAT 792
OCT4 AGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCTGCAGCAAAT 1079
exon4 AGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCTGCAGCAAAT
Contig TAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTGGTTTTGCAA 857
XM_001789212 TAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTGGTTTTGCAA 852
OCT4 TAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAAGACGTGGTCCGAGTGTGGTTTTGCAA 1139
exon4 TAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAAGAC----------
exon5 ----------GTGGTCCGAGTGTGGTTTTGCAA
Contig CCGT---------- 861
XM_001789212 CCGTCAGCAGC--GGCAGCAGCATATACATTCACTGATAGTTATATGTGAGTTTTATGG- 909
OCT4 CCGTCGCCAGAAGGGCAAACGATCAAGCAGTGACTACTCCCAACGTGAGGATTTTGAGGC 1199
exon5 CCGTCGCCAGAAGGGCAAACGATCAAGCAGTGACTACTCCCAACGTGAGGATTTTGAGGC
XM_001789212 TATCGTGTCTTAA---------- 921
OCT4 TGCTGGGTCTCCTTTCACAGGGGGACCCGTATCCTCTCCTCTGGCGCCAGGGCCCCATTTT 1260
exon5 TGCTGGGTCTCCTTTCACAGGGGGACCCGTATCCTCTCCTCTGGCGCCAGGGCCCCATTTT
OCT4 GGTACCCCAGGCTACGGGGGGCCTCACTTCACTACTCTGTACTCTTCGGTCCCATTCCCT 1320
exon5 GGTACCCCAGGCTACGGGGGGCCTCACTTCACTACTCTGTACTCTTCGGTCCCATTCCCT
OCT4 GAGGGTGAGGCCTTTCCCTCGGTGTCTGTCACCGCTCTGGGCTCCCCTATGCATGCAAAC 1380
exon5 GAGGGTGAGGCCTTTCCCTCGGTGTCTGTCACCGCTCTGGGCTCCCCTATGCATGCAAAC

```
OCT4            TGAGGTGCCTGATCACCCCAGGAATGGGGGGCAGAGGAAGGGGAGAGCTAGGGAGAGAAC 1440
exon5           TGAGGTGCCTGATCACCCCAGGAATGGGGGGCAGAGGAAGGGGAGAGCTAGGGAGAGAAC
                ************************************************************

OCT4            CCTGGGGTTTGTACCAGGCCTTTGGGATTAAGTTTTTCATTCACTAAGAAAGGAATTGGG 1500
exon5           CCTGGGGTTTGTACCAGGCCTTTGGGATTAAGTTTTTCATTCACTAAGAAAGGAATTGGG
                ************************************************************

OCT4            AACACAATGGGTGTTGGGGCAGGGAGTTTGGGGAAACTGGTTGGAGGGAAGGTGAAGTTC 1560
exon5           AACACAATGGGTGTTGGGGCAGGGAGTTTGGGGAAACTGGTTGGAGGGAAGGTGAAGTTC 444
                ************************************************************

OCT4            AATGATGCTCTTGACTTTAATCCCCACATCACTCATCACTTTGTTCTTAAATAAA 1615
exon5           AATGATGCTCTTGACTTTAATCCCCACATCACTCATCACTTTGTTCTTAAATAAA 499
                ********************************************************
```

**Supp. Figure 7. A.** Sequence of putative OCT4 paralog protein translated from contiguous *OCT4* retrocopy mRNA sequence from RT-PCR (Fig 2.) if the OCT4 translational start site is conserved and the coding sequence contains no introns. The stop codon lies within the fused L1_BT sequence 3' to the *OCT4* retrocopy. **B.** Alignment of putative OCT4 paralog with bovine OCT4 and human OCT4A proteins. Dashes in sequence (-) and * indicate identical amino acid in the same position. Two dots (:) indicate a highly conserved amino acid substitution and a single dot (.) is a less conserved amino acid substitution. The first 133 amino acids are the N terminal transactivation domain. The lightly shaded box highlights 75 amino acids of the POUs DNA binding domain. Sequence in bold indicate the linker region. The dark gray shaded box is the POU homeodomain. Retrocopy amino acid sequence in bold italics is putative sequence translated from the 3' L1-BT LINE fragment fused 3' to the OCT4 retrocopy. **C.** Percent sequence similarity between the *OCT4* retrocopy lacking the L1 LINE contribution (bold italics), bovine OCT4 and human OCT4A proteins.

**A.**

MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWMSFQGPPGGSGIGPGVVPGAEVWG
LPPCPPPYDLCGGMAYCAPQVGVGPVPPGGLETPQPEGQAGAGVENNSEGASPDPCAAL
AGAAKLDKEKLEPNPEESQDIKALQKDLEQFAKLLKQKRITLGYTQADVGLTLGVLFGK
VFSQTTICHFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTS
IENQVRGNLESMFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRVPLHTSQDGCDPKV
YK*

**B.**

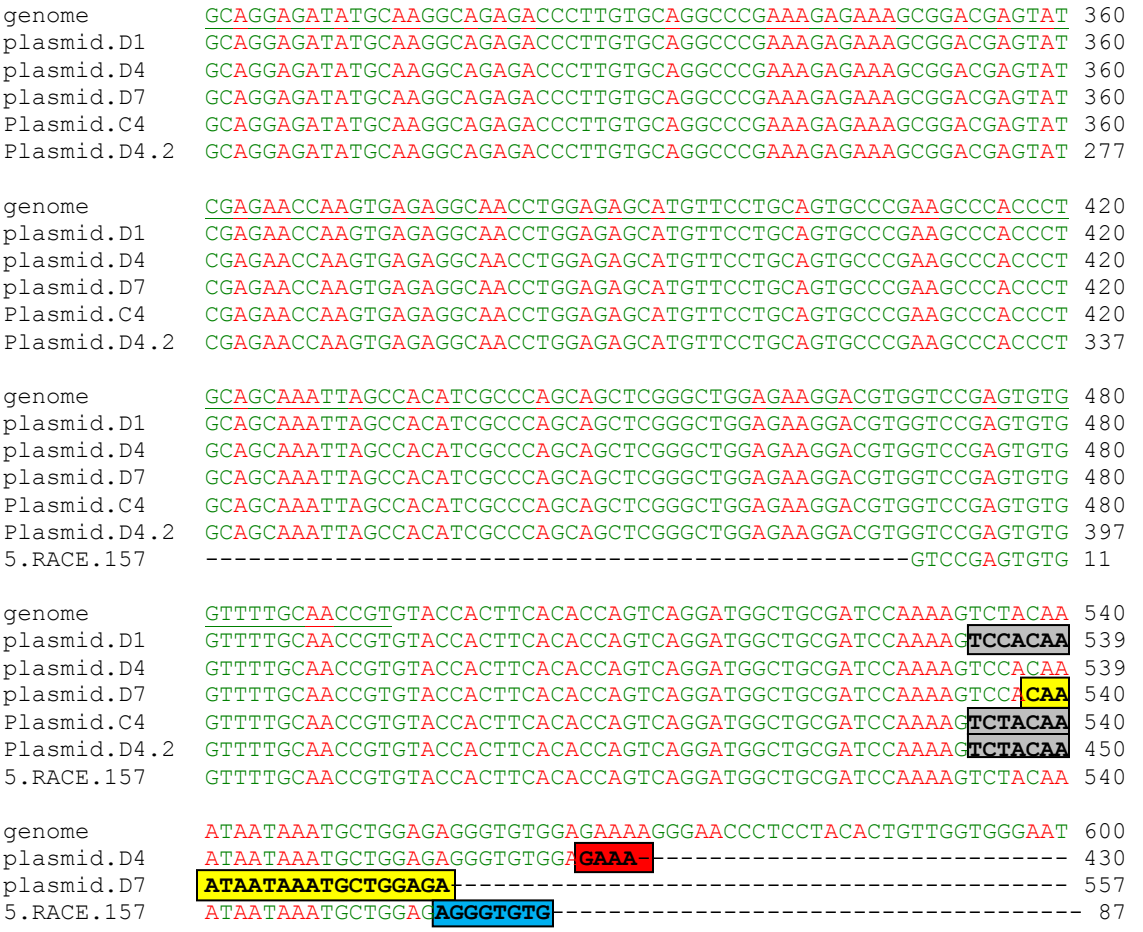
```
Bovine.OCT4     MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWMSFQGPPGGSGIGPGVVPGAEVWGL 60
Retrocopy       ------------------------------------------------------------ 60
Human.OCT4      ----------------------------------L--------P------G-S-----I 60
                **********************************:********.****** **:****:

Bovine.OCT4     PPCPPPYDLCGGMAYCAPQVGVGPVPPGGLETPQPEGEAGAGVESNSEGASPDPCAAPAG 120
Retrocopy       ------------------------------------Q------N------------L-- 120
Human.OCT4A     -------EF-------G------L--Q-----S-------V------D----E--TVTP- 120
                *******::*******.****** ** *****.****:**.***.**:****:**:. .*

Bovine.OCT4     APKLDKEKLEPNPEESQDIKALQKDLEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 180
Retrocopy       -A---------------------------------------------------------- 180
Human.OCT4A     -V--E-----Q-------------E----------------------------------- 180
                * **:***** ************:*********************************

Bovine.OCT4     QTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTSIENR 240
Retrocopy       -----H-----------------------------------------------------Q 240
Human.OCT4A     ------------------------------------------------------------ 240
                *****:**************************************************:

Bovine.OCT4     VRGNLESMFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYSQREDFEA 300
Retrocopy       ----------------------------------------VPLHTSQDGCDPKVYK
Human.OCT4A     ------NL-------------------------------------------A------- 300
                ******.:*********************************    . ...  .:  .***

Bovine.OCT4     AGSPFTGGPVSSPLAPGPHFGTPGYGGPHFTTLYSSVPFPEGEVFPSVSVTALGSPMHAN 360
Human.OCT4A     -----S-----F--------------S----A-----------A--P----T------S- 360
                *****:***** **************.****:***********.**.****:******:*
```

**c.**

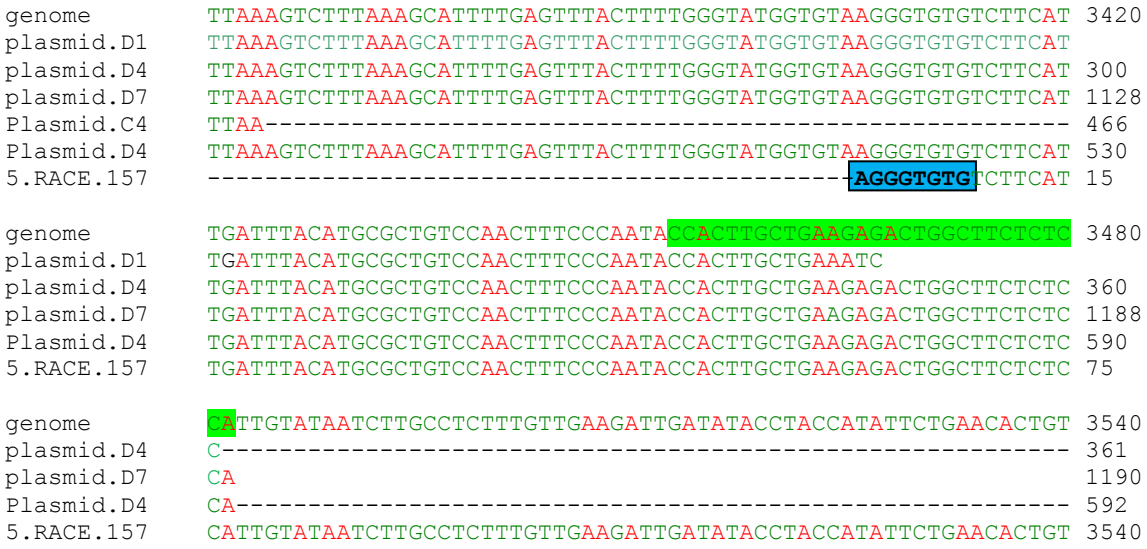
| | Name | Len(aa) | SeqB | Name | Len(aa) | Score |
|---|---|---|---|---|---|---|
| 1 | Human.OCT4A | 360 | 2 | Bovine.OCT4 | 360 | 90 |
| 1 | Human.OCT4A | 360 | 3 | Retrocopy | 281 | 90 |
| 2 | Bovine.OCT4 | 360 | 3 | Retrocopy | 281 | 97 |

**Supp. Figure 8.** Alignments of nested RT-PCR sequences (using primers 160/161) amplified from pooled 5' and 3' RACE cDNAs using primers 160 (highlighted in violet) and 161 (highlighted in green). The nested RT-PCR sequence (primers 157 and Universal nested primer) from 5' RACE cDNA (primers 157 and Universal primer mix) is also aligned. Clones were sequenced with SP6 and T7 sequencing primers. Underlined sequence in genome reference (-3800086 to -3804032 of NW_001495368.2) indicates chromosome 7 *OCT4* retrocopy. Clones were aligned to genome reference to display where sequence break/ligation events occurred. Sequences in colored boxes indicate the regions of discontinuity between clone sequences and the genome reference, and all lie within L1 LINE or SINE fragments. These short repeats may act as cohesive ends or redundant repeats for possible recombination. Sequences in colored boxes are present in clone sequences only once but are duplicated in the alignment to demonstrate conservation at two places within genome reference. Three of six clones exhibit the same discontinuity point/ repeated sequence. Continuous clone sequences match the original RT-PCR product sizes. Therefore, the deletions are not caused by transformed JM109 *E. coli.* The shortest deletion is about 1733 bp (clone D7) and about 2443 bp in the three identical clones (D1, C4 and D4.2).

```
CLUSTAL 2.0.11 multiple sequence alignment

genome          ACTCGCAGGCGCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCA 60
plasmid.D1      ACTCGCAGGCGCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCA 60
plasmid.D4      ACTCGCAGGCGCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCA
plasmid.D7      ACTCGCAGGCGCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCA 60
Plasmid.C4      ACTCGCAGGCGCCGCGAAGCTGGACAAGGAGAAGCTGGAACCGAACCCTGAGGAGTCCCA 60

genome          GGACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAG 120
plasmid.D1      GGACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAG 120
plasmid.D4      GGACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAG
plasmid.D7      GGACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAG 120
Plasmid.C4      GGACATCAAAGCTCTTCAGAAAGACCTTGAACAATTTGCCAAGCTCCTAAAGCAGAAGAG 120
Plasmid.D4.2    ----------------------NCCCTGGANCAATTTGCCCAGCTCCTAAAGCAGAAGAG 38

genome          GATCACACTAGGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAA 180
plasmid.D1      GATCACACTAGGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAA 180
plasmid.D4      GATCACACTAGGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTNGGAAA 41
plasmid.D7      GATCACACTAGGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTTGGAAA 180
Plasmid.C4      GATCACACTAGGATATACCCAGGCTGATGTGGGGCTCACCCTGGGGGTTCTTTTTGGAAA 180
Plasmid.D4.2    GATCNCNCTAGGATATACCCNGGCTGATGTGGGGCTCACCCTGGGGGTTCTCTTNGNAAA 98

genome          GGTGTTCAGCCAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACAT 240
plasmid.D1      GGTGTTCAGCCAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACAT 240
plasmid.D4      GGTGTTCAGCCAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACAT 240
plasmid.D7      GGTGTTCAGCCAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACAT 240
Plasmid.C4      GGTGTTCAGCCAAACGACTATCTGCCATTTTGAGGCTTTGCAGCTCAGTTTCAAGAACAT 240
Plasmid.D4.2    GG-GTTCAGCCNAACGACTATCTGCCATTTNGAGGCTTTGCAGCTCAGTTTCAAGAACAT

genome          GTGTAAGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCT 300
plasmid.D1      GTGTAAGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCT 300
plasmid.D4      GTGTAAGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCT 300
plasmid.D7      GTGTAAGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCT 300
Plasmid.C4      GTGTAAGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCT 300
Plasmid.D4.2    GTGTAAGCTGCGGCCCCTGCTGCAGAAGTGGGTGGAGGAAGCTGACAACAACGAGAATCT 217
```


genome GCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTAT 360
plasmid.D1 GCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTAT 360
plasmid.D4 GCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTAT 360
plasmid.D7 GCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTAT 360
Plasmid.C4 GCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTAT 360
Plasmid.D4.2 GCAGGAGATATGCAAGGCAGAGACCCTTGTGCAGGCCCGAAAGAGAAAGCGGACGAGTAT 277
genome CGAGAACCAAGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCT 420
plasmid.D1 CGAGAACCAAGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCT 420
plasmid.D4 CGAGAACCAAGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCT 420
plasmid.D7 CGAGAACCAAGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCT 420
Plasmid.C4 CGAGAACCAAGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCT 420
Plasmid.D4.2 CGAGAACCAAGTGAGAGGCAACCTGGAGAGCATGTTCCTGCAGTGCCCGAAGCCCACCCT 337
genome GCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTG 480
plasmid.D1 GCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTG 480
plasmid.D4 GCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTG 480
plasmid.D7 GCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTG 480
Plasmid.C4 GCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTG 480
Plasmid.D4.2 GCAGCAAATTAGCCACATCGCCCAGCAGCTCGGGCTGGAGAAGGACGTGGTCCGAGTGTG 397
5.RACE.157 ----------GTCCGAGTGTG 11
genome GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCTACAA 540
plasmid.D1 GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCCACAA 539
plasmid.D4 GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCCACAA 539
plasmid.D7 GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCCACAA 540
Plasmid.C4 GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCTACAA 540
Plasmid.D4.2 GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCTACAA 450
5.RACE.157 GTTTTGCAACCGTGTACCACTTCACACCAGTCAGGATGGCTGCGATCCAAAAGTCTACAA 540
genome ATAATAAATGCTGGAGAGGGTGTGGAGAAAAGGGAACCCTCCTACACTGTTGGTGGGAAT 600
plasmid.D4 ATAATAAATGCTGGAGAGGGTGTGGAGAAA---------- 430
plasmid.D7 ATAATAAATGCTGGAGA---------- 557
5.RACE.157 ATAATAAATGCTGGAGAGGGTGTG---------- 87
Break from genomic reference
genome GATTGTAAATCAATATATATACATTCTGGAGAAAGAAATGGCAACCCACTCCAGTATTCT 2340
plasmid.D7 ----------CAA-ATA-ATAAATGCTGGAGAAAGAAATGGCAACCCACTCCAGTATTCT 48
genome TGCCTGGGGAGTCCTATGGACGGAGGAGCCTGGTGGGCTGCTGTCTGTGGGGTCACACAG 2400
plasmid.D7 TGCCTGGGGAGTCCTATGGACGGAGGAGCCTGGTGGGCTGCTGTCTGTGGGGTCACACAG 108
genome AGTCATACATGACTGAAGCAAGTTAGCAGCAGCGGCAGCAGCATATACATTCACTGATAG 2460
plasmid.D7 AGTCATACATGACTGAAGCAAGTTAGCAGCAGCGGCAGCAGCATATACATTCACTGATAG 168
genome TTATATGTGTATATATACATTCTTTTTCATATTATTTTGCATTATGGTTTATTACTGGAT 2520
plasmid.D7 TTATATGTGTATATATACATTCTTTTTCATATTATTTTGCATTATGGTTTATTACTGGAT 228
genome ATTGAATATAGTTGCCTGTACTGTACAGTAGGACCTTGTTGTTTATCTGTTGATACTTTA 2580
plasmid.D7 ATTGAATATAGTTGCCTGTACTGTACAGTAGGACCTTGTTGTTTATCTGTTGATACTTTA 288
genome TATATAGTAGTATGTGTCTGTTAATCTCTGATTCCTAATTTATGCCTTCCCCCAACGCCT 2640
plasmid.D7 TATATAGTAGTATGTGTCTGTTAATCTCTGATTCCTAATTTATGCCTTCCCCCAACGCCT 348
genome TTCTCCTTTGGTAACCATAAGTTTTTTTTCCTGTCTGTCTGTTTTGTAAATAAATTTATG 2700
plasmid.D7 TTCTCCTTTGGTAACCATAAGTTTTTTTTCCTGTCTGTCTGTTTTGTAAATAAATTTATG 408
genome TCATATTTTAGATTCCACATTAAGTGATAGGATATGTCTTTTCTCTTCCTCAAATGGGAT 2760
plasmid.D7 TCATATTTTAGATTCCACATTAAGTGATAGGATATGTCTTTTCTCTTCCTCAAATGGGAT 468
genome TATTTCATTCTTTTTTATGGCTGTGCATGTGTGTGTATTTAGTTTTTTAAGCAGCCTTCA 2820
plasmid.D7 TATTTCATTCTTTTTTATGGCTGTGCATGTGTGTGTATTTAGTTTTTTAAGCAGCCTTCA 528

```
genome          TACTGTAACACAGTGGCTACACCAGTTTACCAACACCCTCCCACCTACAGTGTAGAAGGG 2880
plasmid.D7      TACTGTAACACAGTGGCTACACCAGTTTACCAACACCCTCCCACCTACAGTGTAGAAGGG 588

genome          TTCCCTTTTCTCCACAACCTCTCCAACATCTGTTATTTGTAGACTTTTTGATGATGACCA 2940
plasmid.D1      ----------TCCACAACCTCTCCAACATCTGTTATTTGTAGACTTTTTGATGATGACCA 50
plasmid.D7      TTCCCTTTTCTCCACAACCTCTCCAACATCTGTTATTTGTAGACTTTTTGATGATGACCA 648
Plasmid.C4      ----------TCCACAACCTCTCCAACATCTGTTATTTGTAGACTTTTTGATGATGACCA 43
Plasmid.D4.2    ----------TCCACAACCTCTCCAACATCTGTTATTTGTAGACTTTTTGATGATGACCA 50

genome          TCCTGGCTGATAGGAAATTTTTATGTGCATTTCTATGTGATAGGAAATTGTAATTTTTAT 3000
plasmid.D1      TCCTGGCTGATAGGAAATTTTTATGTGCATTTCTATGTGATAGGAAATTGTAATTTTTAT 110
plasmid.D4      ------------GAAATTTTTATGTGCATTTCTATGTGATAGGAAATTGTAATTTTTAT 43
plasmid.D7      TCCTGGCTGATAGGAAATTTTTATGTGCATTTCTATGTGATAGGAAATTGTAATTTTTAT 708
Plasmid.C4      TCCTGGCTGATAGGAAATTTTTATGTGCATTTCTATGTGATAGGAAATTGTAATTTTTAT 103
Plasmid.D4.2    TCCTGGCTGATAGGAAATTTTTATGTGCATTTCTATGTGATAGGAAATTGTAATTTTTAT 110

genome          GTGCATTTTTGTGATAATTAGCAGTGTTGAGCGTCCTTTGATGAAATGTCTTCTGCCCAT 3060
plasmid.D1      GTGCATTTTTGTGATAATTAGCAGTGTTGAGCGTCCTTTGATGAAATGTCTTCTGCCCAT 169
plasmid.D4      GTGCATTTTTGTGATAATTAGCAGTGTTGAGCGTCCTTTGATGAAATGTCTTCTGCCCAT 103
plasmid.D7      GTGCATTTTTGTGATAATTAGCAGTGTTGAGCGTCCTTTGATGAAATGTCTTCTGCCCAT 768
Plasmid.C4      GTGCATTTTTGTGATAATTAGCAGTGTTGAGCGTCCTTTGATGAAATGTCTTCTGCCCAT 163
Plasmid.D4.2    GTGCATTTTTGTGATAATTAGCAGTGTTGAGCGTCCTTTGATGAAATGTCTTCTGCCCAT 170

genome          TTTTTGATAGTGTTTGTTTTTTGTTCTTGAGTTAGTTGTATGAACTGTTTGTATGTTTTG 3120
plasmid.D1      TTTTTGATAGTGTTTGTTTTTTGTTCTTGAGTTAGTTGTATGAACTGTTTGTATGTTTTG 61
plasmid.D4      TTTTTGATAGTGTTTGTTTTTTGTTCTTGAGTTTGTTGTATGAACTGTTTGTATGTTTTG 16
plasmid.D7      TTTTTGATAGTGTTTGTTTTTTGTTCTTGAGTTTGTTGTATGAACTGTTTGTATGTTTTG 828
Plasmid.C4      TTTTTGATAGTGTTTGTTTTTTGTTCTTGAGTTTGTTGTATGAACTGTTTGTATGTTTTG 223
Plasmid.D4.2    TTTTTGATAGTGTTTGTTTTTTGTTCTTGAGTTAGTTGTATGAACTGTTTGTATGTTTTG 230

genome          GAAATTAACCACTGTTGGTCGCATCATTTGCAGGTATTTTCTCCCATTCAGTATGTTGTC 3180
plasmid.D1      GAAATTAACCACTGTTGGTCGCATCATTTGCAGGTATTTTCTCCCATTCAGTATGTTGTC 60
plasmid.D4      GAAATTAACCACTGTTGGTCGCATCATTTGCAGGTATTTTCTCCCATTCAGTATGTTGTC 60
plasmid.D7      GAAATTAACCACTGTTGGTCGCATCATTTGCAGGTATTTTCTCCCATTCAGTATGTTGTC 888
Plasmid.C4      GAAATTAACCACTGTTGGTCGCATCATTTGCAGGTATTTTCTCCCATTCAGTATGTTGTC 283
Plasmid.D4.2    GAAATTAACCACTGTTGGTCGCATCATTTGCAGGTATTTTCTCCCATTCAGTATGTTGTC 290

genome          TTTTTGTTTTGTTTATGGTTTCTTTGCTGTGTGAAAGTTTATAAATTTGATTAGGTCCCA 3240
plasmid.D1      TTTTGTTTTGGTTTATGGTTTCTTTGCTGTGTGAAAGTTTATAAATTTGATTAGGTCCCA 120
plasmid.D4      TTTTTGTTTTGTTTATGGTTTCTTTGCTGTGTGAAAGTTTATAAATTTGATTAGGTCCCA 120
plasmid.D7      TTTTTGTTTTGTTTATGGTTTCTTTGCTGTGTGAAAGTTTATAAATTTGATTAGGTCCCA 948
Plasmid.C4      TTTTTGTTTTGTTTATGGTTTCTTTGCTGTGTGAAAGTTTATAAATTTGATTAGGTCCCA 343
Plasmid.D4      TTTTTGTTTTGTTTATGGTTTCTTTGCTGTGTGAAAGTTTATAAATTTGATTAGGTCCCA 350

genome          TTTGTTTATTGTTGCTTTTATTTCTGTTGCTTTGGAAGACAGACCTAAGAAAACATTGGT 3300
plasmid.D1      TTTGTTTATTGTTGCTTTTATTTCTGTTGCTTTGGAAGACAGACCTAAGAAAACATTGGT 180
plasmid.D4      TTTGTTTATTGTTGCTTTTATTTCTGTTGCTTTGGAAGACAGACCTAAGAAAACATTGGT 180
plasmid.D7      TTTGTTTATTGTTGCTTTTATTTCTGTTGCTTTGGAAGACAGACCTAAGAAAACATTGGT 1008
Plasmid.C4      TTTGTTTATTGTTGCTTTTATTTCTGTTGCTTTGGAAGACAGACCTAAGAAAACATTGGT 403
Plasmid.D4      TTTGTTTATTGTTGCTTTTATTTCTGTTGCTTTGGAAGACAGACCTAAGAAAACATTGGT 410

genome          ACGAGTTATCTGAAAAGGTTTTACGTGTGTTCTCTTCTAGGAGTTTTATGGTATCGTGTC 3360
plasmid.D1      ACGAGTTATCTGAAAAGGTTTTACGTGTGTTCTCTTCTAGGAGTTTTATGGTATCGTGTC 189
plasmid.D4      ACGAGTTATCTGAAAAGGTTTTACGTGTGTTCTCTTCTAGGAGTTTTATGGTATCGTGTC 240
plasmid.D7      ACGAGTTATCTGAAAAGGTTTTACGTGTGTTCTCTTCTAGGAGTTTTATGGTATCGTGTC 1068
Plasmid.C4      ACGAGTTATCTGAAAAGGTTT-ACGTGTGTTCTCNTCTAGGAGTTTTATGGTATCGTGTC 462
Plasmid.D4      ACGAGTTATCTGAAAAGGTTTTACGTGTGTTCTCTTCTAGGAGTTTTATGGTATCGTGTC 470
```


genome TTAAAGTCTTTAAAGCATTTTGAGTTTACTTTTGGGTATGGTGTAAGGGTGTGTCTTCAT 3420
plasmid.D1 TTAAAGTCTTTAAAGCATTTTGAGTTTACTTTTGGGTATGGTGTAAGGGTGTGTCTTCAT
plasmid.D4 TTAAAGTCTTTAAAGCATTTTGAGTTTACTTTTGGGTATGGTGTAAGGGTGTGTCTTCAT 300
plasmid.D7 TTAAAGTCTTTAAAGCATTTTGAGTTTACTTTTGGGTATGGTGTAAGGGTGTGTCTTCAT 1128
Plasmid.C4 TTAA---------- 466
Plasmid.D4 TTAAAGTCTTTAAAGCATTTTGAGTTTACTTTTGGGTATGGTGTAAGGGTGTGTCTTCAT 530
5.RACE.157 ----------AGGGTGTGTCTTCAT 15
genome TGATTTACATGCGCTGTCCAACTTTCCCAATACCACTTGCTGAAGAGACTGGCTTCTCTC 3480
plasmid.D1 TGATTTACATGCGCTGTCCAACTTTCCCAATACCACTTGCTGAAATC
plasmid.D4 TGATTTACATGCGCTGTCCAACTTTCCCAATACCACTTGCTGAAGAGACTGGCTTCTCTC 360
plasmid.D7 TGATTTACATGCGCTGTCCAACTTTCCCAATACCACTTGCTGAAGAGACTGGCTTCTCTC 1188
Plasmid.D4 TGATTTACATGCGCTGTCCAACTTTCCCAATACCACTTGCTGAAGAGACTGGCTTCTCTC 590
5.RACE.157 TGATTTACATGCGCTGTCCAACTTTCCCAATACCACTTGCTGAAGAGACTGGCTTCTCTC 75
genome CATTGTATAATCTTGCCTCTTTGTTGAAGATTGATATACCTACCATATTCTGAACACTGT 3540
plasmid.D4 C---------- 361
plasmid.D7 CA 1190
Plasmid.D4 CA---------- 592
5.RACE.157 CATTGTATAATCTTGCCTCTTTGTTGAAGATTGATATACCTACCATATTCTGAACACTGT 3540

# Bibliography

Adelson D.L., Raison J.M., Edgar R.C. (2009) Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. Proc Natl Acad Sci U S A. 106(31):12855-60.

Ambrosetti D.C., Basilico C., Dailey L. (1997) Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol. 17(11):6321-9.

Ambrosetti D.C., Schöler H.R., Dailey L., Basilico C. (2000) Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J Biol Chem. 275(30):23387-97.

Armant D.R. (2005) Blastocysts don't go it alone. Extrinsic signals fine-tune the intrinsic developmental program of trophoblast cells. Dev Biol. 280(2):260-80.

Arnold D.R., Bordignon V., Lefebvre R., Murphy B.D., Smith L.C. (2006) Somatic cell nuclear transfer alters peri-implantation trophoblast differentiation in bovine embryos. Reproduction. 132(2):279-90.

Aston K.I., Li G.P., Hicks B.A., Sessions B.R., Davis A.P., Rickords L.F., Stevens J.R., White K.L. (2010) Abnormal levels of transcript abundance of developmentally important genes in various stages of preimplantation bovine somatic cell nuclear transfer embryos.Cloning Stem Cells. 12(1):23-32

Avilion AA, Nicolis S.K., Pevny L.H., Perez L., Vivian N., Lovell-Badge R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17(1):126-40.

Bai H., Sakurai T., Kim M.S., Muroi Y., Ideta A., Aoyagi Y., Nakajima H., Takahashi M., Nagaoka K., Imakawa K. (2009) Involvement of GATA transcription factors in the regulation of endogenous bovine interferon-tau gene transcription. Mol Reprod Dev. 76(12):1143-52.

Bai Y., Casola C., Betrán E. (2008) Quality of regulatory elements in Drosophila retrogenes. Genomics. 93(1):83-9.

Bannert N. and Kurth R. (2004) Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci U S A. 101 Suppl 2:14572-9

Barcroft L.C., Hay-Schmidt A., Caveney A., Gilfoyle E., Overstrom E.W., Hyttel P., Watson A.J. (1998) Trophectoderm differentiation in the bovine embryo: characterization of a polarized epithelium. J Reprod Fertil. 114(2):327-39.

Betteridge K.J., Flechon J.E. (1988) The anatomy and physiology of pre-attachment bovine embryos. Theriogenology 29:155-187.

Blomberg L., Hashizume K., Viebahn C. (2008) Blastocyst elongation, trophoblastic differentiation, and embryonic pattern formation. Reproduction. 135(2):181-95. Review.

Booth H.A., Holland P.W. (2004) Eleven daughters of NANOG. Genomics. (2):229-38.

Botquin V., Hess H., Fuhrmann G., Anastassiadis C., Gross M.K., Vriend G., Scholer H.R. (1998) New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. Genes Dev. 12(13):2073-90.

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., Young R.A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 122(6):947-56.

Catena R., Tiveron C., Ronchi A., Porta S., Ferri A., Tatangelo L., Cavallaro M., Favaro R., Ottolenghi S., Reinbold R., Scholer H., Nicolis S.K. (2004) Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. J Biol Chem. 279(40):41846-57.

Cauffman G., Liebaers I., Van Steirteghem A., Van de Velde H. (2006) POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. Stem Cells. 24(12):2685-91.

Cavaleri F., Scholer H.R. (2003) Nanog: a new recruit to the embryonic stem cell orchestra. Cell. 113(5):551-2.

Chai N., Patel Y., Jacobson K., McMahon J., McMahon A., Rappolee D.A. (1998) FGF is an essential regulator of the fifth cell division in preimplantation mouse embryos. Dev Biol. 198(1):105-15.

Chambers I., Colby D., Robertson M., Nichols J., Lee S., Tweedie S., Smith A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. 113(5):643-55.

Chazaud C., Yamanaka Y., Pawson T., Rossant J. (2006) Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell. 10(5):615-24.

Chen L., Yabuuchi A., Eminli S., Takeuchi A., Lu CW., Hochedlinger K., Daley G.Q. (2009) Cross-regulation of the Nanog and Cdx2 promoters. 19(9):1052-61.

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(14):6031-46.

Conley A.B., Miller W.J., Jordan I.K. (2008) Human cis natural antisense transcripts initiated by transposable elements. Trends Genet. 24(2):53-6.

Cooke F.N., Pennington K.A., Yang Q., Ealy A.D. (2009) Several fibroblast growth factors are expressed during pre-attachment bovine conceptus development and regulate interferon-tau expression from trophectoderm. Reproduction. 137(2):259-69.

Dalby B., Cates S., Harris A., Ohki E.C., Tilkins M.L., Price P.J., Ciccarone V.C. (2004) Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. Methods. 33(2):95-103.

Daniels R., Hall V.J, French A.J, Korfiatis N.A, Trounson A.O. (2001) Comparison of gene transcription in cloned bovine embryos produced by different nuclear transfer techniques. Mol Reprod Dev. 60(3):281-8.

de A Camargo L.S., Powell A.M., do Vale Filho V.R., Wall R.J. (2005) Comparison of gene expression in individual preimplantation bovine embryos produced by in vitro fertilisation or somatic cell nuclear transfer. Reprod Fertil Dev. 17(5):487-96.

Degrelle S.A., Campion E., Cabau C., Piumi F., Reinaud P., Richard C., Renard J.P., Hue I. (2005) Molecular evidence for a critical period in mural trophoblast development in bovine blastocysts. Dev Biol. 288(2):448-60.

Dietrich J.E., Hiiragi T. (2007) Stochastic patterning in the mouse pre-implantation embryo. Development. 134(23):4219-31.

Donnison M., Beaton A., Davey H.W., Broadhurst R., L'Huillier P., Pfeffer P.L. (2005) Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. Development. 132(10):2299-308.

Douglas G.C., VandeVoort C.A., Kumar P., Chang T.C., Golos T.G. (2009) Trophoblast stem cells: models for investigating trophectoderm differentiation and placental development. Endocr Rev. 30(3):228-40.

Dunlap K.A., Palmarini M., Varela M., Burghardt R.C., Hayashi K., Farmer J.L., Spencer T.E. (2006) Endogenous retroviruses regulate periimplantation placental growth and differentiation. Proc Natl Acad Sci U S A. 103(39):14390-5.

Ealy A.D., Yang Q.E. (2009) Control of interferon-tau expression during early pregnancy in ruminants. Am J Reprod Immunol. 61(2):95-106.

Ebisuya M., Yamamoto T., Nakajima M., Nishida E. (2008) Ripples from neighbouring transcription. Nat Cell Biol. 10(9):1106-13.

Esnault C., Maestre J., Heidmann T. (2000) Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 24(4):363-7.

Ezashi T., Telugu B.P., Alexenko A.P., Sachdev S., Sinha S., Roberts R.M. (2009) Derivation of induced pluripotent stem cells from pig somatic cells. Proc Natl Acad Sci U S A. 106(27):10993-8.

Ezashi T., Ghosh D., Roberts R.M. (2001) Repression of Ets-2-induced transactivation of the tau interferon promoter by Oct-4. Mol Cell Biol. 21(23):7883-91.

Ezashi T., Das P., Gupta R., Walker A., Roberts R.M. (2008) The role of homeobox protein distal-less 3 and its interaction with ETS2 in regulating bovine interferon-tau gene expression-synergistic transcriptional activation with ETS2. Biol Reprod. 79(1):115-24.

Farin P.W., Piedrahita J.A., Farin C.E. (2006) Errors in development of fetuses and placentas from in vitro-produced bovine embryos. Theriogenology. 65(1):178-91. Review.

Fujikura J., Yamato E., Yonemura S., Hosoda K., Masui S., Nakao K., Miyazaki Ji J., Niwa H. (2002) Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev. 16(7):784-9.

Garbayo J.M., Serrano B., Lopez-Gatius F. (2008) Identification of novel pregnancy-associated glycoproteins (PAG) expressed by the peri-implantation conceptus of domestic ruminants. Anim Reprod Sci. 103(1-2):120-34.

Gardner R.L. (2000) Flow of cells from polar to mural trophectoderm is polarized in the mouse blastocyst. Hum Reprod. 15(3):694-701.

Gerstein M.B., Bruce C., Rozowsky J.S., Zheng D., Du J., Korbel J.O., Emanuelsson O., Zhang Z.D., Weissman S., Snyder M. (2007) What is a gene, post-ENCODE? History and updated definition. Genome Res. 17(6):669-81.

Gogvadze E., Buzdin A. (2009) Retroelements and their impact on genome evolution and functioning. Cell Mol Life Sci. 66(23):3727-42.

Gray C.A., Taylor K.M., Ramsey W.S., Hill J.R., Bazer F.W., Bartol F.F., Spencer T.E. (2001) Endometrial glands are required for preimplantation conceptus elongation and survival. Biol Reprod. 64(6):1608-13.

Gray C.A, Burghardt R.C., Johnson G.A., Bazer F.W., Spencer T.E. (2002) Evidence that absence of endometrial gland secretions in uterine gland knockout ewes compromises conceptus survival and elongation. Reproduction. 124(2):289-300.

Green J.A., Xie S., Quan X., Bao B., Gan X., Mathialagan N., Beckers J.F., Roberts R.M. (2000) Pregnancy-associated bovine and ovine glycoproteins exhibit spatially and temporally distinct expression patterns during pregnancy. Biol Reprod. 62(6):1624-31.

Haffner-Krausz R., Gorivodsky M., Chen Y., Lonai P. (1999) Expression of Fgfr2 in the early mouse embryo indicates its involvement in preimplantation development. Mech Dev. 85(1-2):167-72.

Hall V.J., Ruddock N.T., French A.J. (2005) Expression profiling of genes crucial for placental and preimplantation development in bovine in vivo, in vitro, and nuclear transfer blastocysts. Mol Reprod Dev. (1):16-24.

Hamatani T., Daikoku T., Wang H., Matsumoto H., Carter M.G., Ko M.S., Dey S.K. (2004) Global gene expression analysis identifies molecular pathways distinguishing blastocyst dormancy and activation. Proc Natl Acad Sci U S A. 101(28):10326-31.

Hatano S.Y., Tada M., Kimura H., Yamaguchi S., Kono T., Nakano T., Suemori H., Nakatsuji N., Tada T. (2005) Pluripotential competence of cells associated with Nanog activity. Mech Dev. 122(1):67-79.

He S., Pant D., Schiffmacher A., Bischoff S., Melican D., Gavin W., Keefer C. (2006) Developmental expression of pluripotency determining factors in caprine embryos: Novel pattern of NANOG protein localization in the nucleolus. Mol Reprod. Dev. 73:1512-22.

He S., Pant D., Schiffmacher A., Meece A., Keefer C.L. (2008) Lymphoid enhancer factor 1-mediated Wnt signaling promotes the initiation of trophoblast lineage differentiation in mouse embryonic stem cells. Stem Cells. 26(4):842-9.

Hemberger M., Dean W., Reik W. (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol. 10(8):526-37.

Hiiragi T., Louvet-Vallee S., Solter D., Maro B. (2006) Embryology: does prepatterning occur in the mouse egg? Nature. 442(7099):E3-4.

Hill J.R., Burghardt R.C., Jones K., Long C.R., Looney C.R., Shin T., et al. (2000) Evidence for placental abnormality as the major cause of mortality in first-trimester somatic cell cloned bovine fetuses. Biol Reprod. 63:1787-94.

Home P., Ray S., Dutta D., Bronshteyn I., Larson M., Paul S. (2009) GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. J Biol Chem. 284(42):28729-37.

Huang C.Y., Uno T., Murphy J.E., Lee S., Hamer J.D., Escobedo J.A., Cohen F.E., Radhakrishnan R., Dwarki V., Zuckermann R.N. (1998) Lipitoids--novel cationic lipids for cellular delivery of plasmid DNA in vitro. Chem Biol. 5(6):345-54.

Imakawa K., Kim M.S., Matsuda-Minehata F., Ishida S., Iizuka M., Suzuki M., Chang K.T., Echternkamp S.E., Christenson R.K. (2006) Regulation of the ovine interferon-tau gene by a blastocyst-specific transcription factor, Cdx2. Mol Reprod Dev. 73(5):559-67.

Ivanova N., Dobrin R., Lu R., Kotenko I., Levorse J., DeCoste C., Schafer X., Lun Y., Lemischka I.R. (2006) Dissecting self-renewal in stem cells with RNA interference. Nature. 442(7102):533-8.

Iwashita S., Osada N., Itoh T., Sezaki M., Oshima K., Hashimoto E., Kitagawa-Arita Y., Takahashi I., Masui T., Hashimoto K., Makalowski W. (2003) A transposable element-mediated gene divergence that directly produces a novel type bovine Bcnt protein including the endonuclease domain of RTE-1. Mol Biol Evol. 20(9):1556-63.

Janatpour M.J., McMaster M.T., Genbacev O., Zhou Y., Dong J., Cross J.C., Israel M.A., Fisher S.J.(2000) Id-2 regulates critical aspects of human cytotrophoblast differentiation, invasion and migration. Development. 127(3):549-58.

Jang G., Jeon H.Y., Ko K.H., Park H.J., Kang S.K., Lee B.C., Hwang W.S. (2005) Developmental competence and gene expression in preimplantation bovine embryos derived from somatic cell nuclear transfer using different donor cells. Zygote.13(3):187-95.

Jiang W., Zhou L., Breyer B., Feng T., Cheng H., Haydon R., Ishikawa A., He T.C. (2001) Tetracycline-regulated gene expression mediated by a novel chimeric repressor that recruits histone deacetylases in mammalian cells. J Biol Chem. 276(48):45168-74.

Johnson M.H., McConnell J.M. (2004) Lineage allocation and cell polarity during mouse embryogenesis. Semin Cell Dev Biol. 15(5):583-97.

Johnson M.H., Ziomek C.A. (1981) The foundation of two distinct cell lineages within the mouse morula. Cell. 24(1):71-80.

Kan N.G., Stemmler M.P., Junghans D., Kanzler B., de Vries W.N., Dominis M., Kemler R. (2007) Gene replacement reveals a specific role for E-cadherin in the formation of a functional trophectoderm. Development. 134(1):31-41.

Keefer C.L., Pant D., Blomberg L., Talbot N.C. (2007) Challenges and prospects for the establishment of embryonic stem cell lines of domesticated ungulates. Anim Rerod Sci. Anim Reprod Sci. 98(1-2):147-68.

Keefer C.L. (2004) Production of bioproducts through the use of transgenic animal models, Anim Reprod Sci. 82-83:5-12.

Kirchhof N., Carnwath J.W., Lemme E., Anastassiadis K., Scholer H., Niemann H. (2000) Expression pattern of Oct-4 in preimplantation embryos of different species. Biol Reprod. 63(6):1698-705.

Kuijk E.W., Du Puy L., Van Tol H.T., Oei C.H., Haagsman H.P., Colenbrander B., Roelen B.A. (2008) Differences in early lineage segregation between mammals. Dev Dyn. 237(4):918-27.

Kuroda T., Tada M., Kubota H., Kimura H., Hatano S.Y., Suemori H., Nakatsuji N., Tada T. (2005) Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol Cell Biol. 25(6):2475-85.

Kurosaka S., Eckardt S., McLaughlin K.J. (2004) Pluripotent lineage definition in bovine embryos by Oct4 transcript localization. Biol Reprod. 71(5):1578-82.

Larue L., Ohsugi M., Hirchenhain J., Kemler R. (1994) E-cadherin null mutant embryos fail to form a trophectoderm epithelium. Proc Natl Acad Sci U S A. 91(17):8263-7.

Lee T.I., Rinaldi N.J., Robert F., Odom D.T., Bar-Joseph Z., Gerber G.K., Hannett N.M., Harbison C.T., Thompson C.M., Simon I., Zeitlinger J., Jennings E.G., Murray H.L., Gordon D.B., Ren B., Wyrick J.J., Tagne J.B., Volkert T.L., Fraenkel E., Gifford D.K., Young R.A. (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. Science. 298(5594):799-804.

Liedtke S., Stephan M., Kögler G. (2008) Oct4 expression revisited: potential pitfalls for data misinterpretation in stem cell research. Biol Chem. 389(7):845-50.

Liedtke S., Enczmann J., Waclawczyk S., Wernet P., Kögler G. (2007) Oct4 and its pseudogenes confuse stem cell research. Cell Stem Cell. 2007 Oct 11;1(4):364-6. No abstract available. Erratum in: Cell Stem Cell. 2008 Feb;2(2):190

Lim C.Y., Tam W.L., Zhang J., Ang H.S., Jia H., Lipovich L., Ng H.H., Wei C.L., Sung W.K., Robson P., Yang H., Lim B. (2008) Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. Cell Stem Cell. 3(5):543-54.

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.Y., Sung K.W., Lee C.W., Zhao X.D., Chiu K.P., Lipovich L., Kuznetsov V.A., Robson P., Stanton L.W., Wei C.L., Ruan Y., Lim B., Ng H.H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature Genetics. 38(4):431-40.

Marques A.C., Dupanloup I., Vinckenbosch N., Reymond A., Kaessmann H. (2005) Emergence of young human genes after a burst of retroposition in primates. PLoS Biol. 3(11):e357.

Martindill D.M., Risebro C.A., Smart N., Franco-Viseras Mdel M., Rosario C.O., Swallow C.J., Dennis J.W., Riley P.R.(2007) Nucleolar release of Hand1 acts as a molecular switch to determine cell fate. Nat Cell Biol. 9(10):1131-41.

Maruyama M., Ichisaka T., Nakagawa M., Yamanaka S. (2005) Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. J Biol Chem. 280(26):24371-9.

Meilhac S.M., Adams R.J., Morris S.A., Danckaert A., Le Garrec J.F., Zernicka-Goetz M. (2009) Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. Dev Biol. 331(2):210-21.

Mercer T.R., Dinger M.E., Mattick J.S. (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet. 10(3):155-9.

Metzger D.E., Xu Y., Shannon J.M. (2007) Elf5 is an epithelium-specific, fibroblast growth factor-sensitive transcription factor in the embryonic lung. Dev Dyn. 236(5):1175-92.

Michael D.D., Alvarez I.M., Ocon O.M., Powell A.M., Talbot N.C., Johnson S.E., Ealy A.D. (2006a) Fibroblast growth factor-2 is expressed by the bovine uterus and stimulates interferon-tau production in bovine trophectoderm. Endocrinology. 147(7):3571-9.

Michael D.D., Wagner S.K, Ocon O.M., Talbot N.C., Rooke J.A., Ealy A.D. (2006) Granulocyte-macrophage colony-stimulating-factor increases interferon-tau protein secretion in bovine trophectoderm cells. Am J Reprod Immunol. 56(1):63-7.

Mitalipov S.M., Kuo H.C., Hennebold J.D., Wolf D.P. (2003) Oct-4 expression in pluripotent cells of the rhesus monkey. Biol Reprod. 69(6):1785-92.

Mitsui K., Tokuzawa Y., Itoh H., Segawa K., Murakami M., Takahashi K., Maruyama M., Maeda M., Yamanaka S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell. 113(5):631-42.

Masui S., Nakatake Y., Toyooka Y., Shimosato D., Yagi R., Takahashi K., Okochi H., Okuda A., Matoba R., Sharov AA, Ko M.S., Niwa H. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol. 9(6):625-35.

Nakano H., Shimada A., Imai K., Takahashi T., Hashizume K. (2005) The cytoplasmic expression of E-cadherin and beta-catenin in bovine trophoblasts during binucleate cell differentiation. Placenta. 26(5):393-401.

Nakano H., Shimada A., Imai K., Takahashi T., Hashizume K. (2002) Association of Dolichos biflorus lectin binding with full differentiation of bovine trophoblast cells. Reproduction. 124(4):581-92.

Ng R.K., Dean W., Dawson C., Lucifero D., Madeja Z., Reik W., Hemberger M. (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nat Cell Biol. 10(11):1280-90.

Nichols J., Zevnik B., Anastassiadis K., Niwa H., Klewe-Nebenius D., Chambers I., Scholer H., Smith A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. 95(3):379-91.

Nishimoto M., Fukushima A., Okuda A., Muramatsu M. (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol Cell Biol. 9(8):5453-65.

Nishioka N., Yamamoto S., Kiyonari H., Sato H., Sawada A., Ota M., Nakao K., Sasaki H. (2008) Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech Dev. 125(3-4):270-83.

Nishioka N, Inoue K, Adachi K, Kiyonari H, Ota M, Ralston A, Yabuta N, Hirahara S, Stephenson RO, Ogonuki N, Makita R, Kurihara H, Morin-Kensicki EM, Nojima H, Rossant J, Nakao K, Niwa H, Sasaki H. (2009) The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Dev Cell. 16(3):398-410.

Niwa H., Toyooka Y., Shimosato D., Strumpf D., Takahashi K., Yagi R., Rossant J. (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell. 123(5):917-29.

Niwa H., Miyazaki J., Smith A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet. 24(4):372-6.

Okamoto K., Okazawa H., Okuda A., Sakai M., Muramatsu M., Hamada H. (1990) A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell. 60(3):461-72.

Okumura-Nakanishi S., Saito M., Niwa H., Ishikawa F. (2005) Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem. 280(7):5307-17.

Ohsugi M., Larue L., Schwarz H., Kemler R. (1997) Cell-junctional and cytoskeletal organization in mouse blastocysts lacking E-cadherin. Dev Biol. 185(2):261-71.

Pain D., Chirn G.W., Strassel C., Kemp D.M. (2005) Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. J Biol Chem. 280(8):6265-8.

Palmieri S.L., Peter W., Hess H., Scholer H.R. (1994) Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Dev Biol. 166(1):259-67.

Pan G., Li J., Zhou Y., Zheng H., Pei D. (2006) A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. FASEB J. 20(10):1730-2.

Pant D., Keefer C.L. (2009) Expression of pluripotency-related genes during bovine inner cell mass explant culture. Cloning Stem Cells. 11(3):355-65.

Parker H.G., VonHoldt B.M., Quignon P., Margulies E.H., Shao S., Mosher D.S., Spady T.C., Elkahloun A., Cargill M., Jones P.G., Maslen C.L., Acland G.M., Sutter N.B., Kuroki K., Bustamante C.D., Wayne R.K., Ostrander E.A. (2009) An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science. 325(5943):995-8

Pesce M. and Scholer H.R. (2001) Oct-4:Gatekeeper in the beginnings of mammalian development. Stem Cells. 19(4):271-8.

Plusa B., Hadjantonakis A.K., Gray D., Piotrowska-Nitsche K., Jedrusik A., Papaioannou V.E., Glover D.M., Zernicka-Goetz M. (2005a) The first cleavage of the mouse zygote predicts the blastocyst axis. Nature. 434(7031):391-5.

Plusa B, Frankenberg S, Chalmers A, Hadjantonakis AK, Moore CA, Papalopulu N, Papaioannou VE, Glover DM, Zernicka-Goetz M. (2005b) Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. J Cell Sci. 118:505-15.

Powell A.M., Talbot N.C., Wells K.D., Kerr D.E., Pursel V.G., Wall R.J. (2004) Cell donor influences success of producing cattle by somatic cell nuclear transfer. Biol Reprod. 71(1):210-6.

Ralston A., Rossant J. (2005) Genetic regulation of stem cell origins in the mouse embryo. Clin Genet. 68(2):106-12.

Ralston A., Rossant J. (2008) Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. Dev Biol. 313(2):614-29.

Ralston A., Cox B.J., Nishioka N., Sasaki H., Chea E., Rugg-Gunn P., Guo G., Robson P., Draper J.S., Rossant J. (2010) Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development. 137(3):395-403.

Ray S., Dutta D., Rumi M.A., Kent L.N., Soares M.J., Paul S. (2009) Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. J Biol Chem. 284(8):4978-88.

Remenyi A., Lins K., Nissen L.J., Reinbold R., Scholer H.R., 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(16):2048-59.

Roberts R.M., Ezashi T., Das P. (2004) Trophoblast gene expression: transcription factors in the specification of early trophoblast. Reprod Biol Endocrinol. 2:47.

Robinson R.S., Fray M.D., Wathes D.C., Lamming G.E., Mann G.E. (2006) In vivo expression of interferon tau mRNA by the embryonic trophoblast and uterine concentrations of interferon tau protein during early pregnancy in the cow. Mol Reprod Dev. 73(4):470-4.

Rodda D.J., Chew J.L., Lim L.H., Loh Y.H., Wang B., Ng H.H., Robson P. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem. 280(26):24731-7.

Rodríguez-Alvarez L., Sharbati J., Sharbati S., Cox J.F., Einspanier R., Castro F.O. (2010) Differential gene expression in bovine elongated (Day 17) embryos produced by somatic cell nucleus transfer and in vitro fertilization. Theriogenology. Epub ahead of print.

Rosner M.H., Vigano M.A., Ozato K., Timmons P.M., Poirier F., Rigby P.W., Staudt L.M. (1990) A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature. 345(6277):686-92.

Russ A.P., Wattler S., Colledge W.H., Aparicio S.A., Carlton M.B., Pearce J.J., Barton S.C., Surani M.A., Ryan K., Nehls M.C., Wilson V., Evans M.J. (2002) Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature. 404(6773):95-9.

Sakurai T., Sakamoto A., Muroi Y., Bai H., Nagaoka K., Tamura K., Takahashi T., Hashizume K., Sakatani M., Takahashi M., Godkin J.D., Imakawa K. (2009) Induction of endogenous interferon tau gene transcription by CDX2 and high acetylation in bovine nontrophoblast cells. Biol Reprod. 80(6):1223-31.

Sasaki H. (2010) Mechanisms of trophectoderm fate specification in preimplantation mouse development. Dev Growth Differ. Epub ahead of print.

Sayah D.M., Sokolskaja E., Berthoux L., Luban J. (2004) Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature. 430(6999):569-73.

Schiffmacher A., Keefer, C.L. (2008). Characterization of the bovine embryo-derived CT-1 cell line for studying regulation of transcription factors expressed with the bovine trophectoderm. Reprod. Fert. Dev. 20(1):136-137, abst. 113.

Schiffmacher A., Keefer C. (2010a) Optimization of nucleic acid transfection methods for bovine trophectoderm CT-1 cells. Placenta. In preparation.

Schiffmacher A., Keefer C. (2010b) The roles of OCT4 and CDX2 in directing early bovine trophoblast development. Biol Reprod. In preparation.

Scholer H.R., Ruppert S., Suzuki N., Chowdhury K., Gruss P. (1990) New type of POU domain in germ line-specific protein Oct-4. Nature. 344(6265):435-9.

Scholer H.R., Balling R., Hatzopoulos A.K., Suzuki N., Gruss P. (1989) Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. EMBO J. 8(9):2551-7.

Scott I.C., Anson-Cartwright L., Riley P., Reda D., Cross J.C. (2000) The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. Mol Cell Biol. 20(2):530-41.

Sheth B., Nowak R.L., Anderson R., Kwong W.Y., Papenbrock T., Fleming T.P. (2008) Tight junction protein ZO-2 expression and relative function of ZO-1 and ZO-2 during mouse blastocyst formation. Exp Cell Res. 314(18):3356-68.

Shimada A, Nakano H, Takahashi T, Imai K, Hashizume K. (2001) Isolation and characterization of a bovine blastocyst-derived trophoblastic cell line, BT-1: development of a culture system in the absence of feeder cell. Placenta. 22(7):652-62.

Sritanaudomchai H., Sparman M., Tachibana M., Clepper L., Woodward J., Gokhale S., Wolf D., Hennebold J., Hurlbut W., Grompe M., Mitalipov S. (2009) CDX2 in the formation of the trophectoderm lineage in primate embryos. Dev Biol. 335(1):179-87.

Strumpf D., Mao C.A., Yamanaka Y., Ralston A., Chawengsaksophak K, Beck F, Rossant J. (2005) Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development. 132(9):2093-102.

Suo G., Han J., Wang X., Zhang J., Zhao Y., Zhao Y., Dai J. (2005) Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun. 337(4):1047-51.

Suwińska A., Czołowska R., Ozdzeński W., Tarkowski A.K. (2008) Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. Dev Biol. 322(1):133-44.

Suzuki A., Raya A., Kawakami Y., Morita M., Matsui T., Nakashima K., Gage F.H., Rodriguez-Esteban C., Belmonte J.C. (2006a) Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. Nat Clin Pract Cardiovasc Med. Suppl 1:S114-22.

Suzuki A., Raya A., Kawakami Y., Morita M., Matsui T., Nakashima K., Gage F.H., Rodriguez-Esteban C., Izpisua Belmonte J.C. (2006b) Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. Proc Natl Acad Sci U S A. 103(27):10294-9.

Szyda A., Paprocka M., Krawczenko A., Lenart K., Heimrath J., Grabarczyk P., Mackiewicz A., Duś D. (2007) Optimization of a retroviral vector for transduction of human CD34 positive cells.
Acta Biochim Pol. 53(4):815-23.

Talbot N.C., Powell A.M., Camp M., Ealy A.D. (2007) Establishment of a bovine blastocyst-derived cell line collection for the comparative analysis of embryos created in vivo and by in vitro fertilization, somatic cell nuclear transfer, or parthenogenetic activation. In Vitro Cell Dev Biol Anim. 43(2):59-71.

Talbot N.C., Caperna T.J., Edwards J.L., Garrett W., Wells K.D., Ealy A.D. (2000) Bovine blastocyst-derived trophectoderm and endoderm cell cultures: interferon tau and transferrin expression as respective in vitro markers. Biol Reprod. 62(2):235-47.

Tam O.H., Aravin AA, Stein P., Girard A., Murchison E.P., Cheloufi S., Hodges E., Anger M., Sachidanandam R., Schultz R.M., Hannon G.J. (2008) Pseudogene-derived

small interfering RNAs regulate gene expression in mouse oocytes. Nature. 453(7194):534-8.

Tanaka S., Kunath T., Hadjantonakis A.K., Nagy A., Rossant J. (1998) Promotion of trophoblast stem cell proliferation by FGF4. Science. 282(5396):2072-5.

Tarkowski A.K., Wroblewska J. (1967) Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. J Embryol Exp Morphol. 18(1):155-80.

Tokuzawa Y., Kaiho E., Maruyama M., Takahashi K., Mitsui K., Maeda M., Niwa H., Yamanaka S. (2003) Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. Mol Cell Biol. 23(8):2699-708.

Tsai R.Y., McKay R.D. (2002) A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. Genes Dev. 16(23):2991-3003.

Ushizawa K, Takahashi T, Kaneyama K, Tokunaga T, Tsunoda Y, Hashizume K. (2005) Gene expression profiles of bovine trophoblastic cell line (BT-1) analyzed by a custom cDNA microarray. J Reprod Dev. 51(2):211-20.

Ushizawa K., Takahashi T., Hosoe M., Kizaki K., Hashizume K. (2009) Characterization and expression analysis of SOLD1, a novel member of the retrotransposon-derived Ly-6 superfamily, in bovine placental villi. PLoS One. 4(6):e5814.

Utku Y., Dehan E., Ouerfelli O., Piano F., Zuckermann R.N., Pagano M., Kirshenbaum K. (2006) A peptidomimetic siRNA transfection reagent for highly effective gene silencing. Mol Biosyst. 2(6-7):312-7.

van Eijk M.J., van Rooijen M.A., Modina S., Scesi L., Folkers G., van Tol H.T., Bevers M.M., Fisher S.R., Lewin H.A., Rakacolli D., Galli C., de Vaureix C., Trounson A.O., Mummery C.L., Gandolfi F. (1999) Molecular cloning, genetic mapping, and developmental expression of bovine POU5F1. Biol Reprod. 60(5):1093-103.

Vejlsted M., Avery B., Schmidt M., Greve T., Alexopoulos N., Maddox-Hyttel P. (2005) Ultrastructural and immunohistochemical characterization of the bovine epiblast. Biol Reprod. 72(3):678-86.

Viebahn C. (1999) The anterior margin of the mammalian gastrula: comparative and phylogenetic aspects of its role in axis formation and head induction. Curr Top Dev Biol. 46:63-103.

Vinckenbosch N., Dupanloup I., Kaessmann H. (2006) Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci U S A. 103(9):3220-5.

Vinot S., Le T., Ohno S., Pawson T., Maro B., Louvet-Vallee S. (2005) Asymmetric distribution of PAR proteins in the mouse embryo begins at the 8-cell stage during compaction. Dev Biol. 282(2):307-19.

Walker A.M., Kimura K., Roberts R.M. (2009) Expression of bovine interferon-tau variants according to sex and age of conceptuses. Theriogenology. 72(1):44-53

Wall R.J., Powell A.M., Paape M.J., Kerr D.E., Bannerman D.D., Pursel V.G., Wells K.D., Talbot N, Hawk H.W. (2005) Genetically enhanced cows resist intramammary Staphylococcus aureus infection. Nat Biotechnol. (4):445-51.

Wang H., Ding T., Brown N., Yamamoto Y., Prince L.S., Reese J., Paria B.C. (2008) Zonula occludens-1 (ZO-1) is involved in morula to blastocyst transformation in the mouse. Dev Biol. 318(1):112-25.

Wang L., Schultz G.A. (1996) Expression of Oct-4 during differentiation of murine F9 cells. Biochem Cell Biol. 74(4):579-84.

Watanabe T., Totoki Y., Toyoda A., Kaneda M., Kuramochi-Miyagawa S., Obata Y., Chiba H., Kohara Y., Kono T., Nakano T., Surani M.A., Sakaki Y., Sasaki H. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. 453(7194):539-43.

Wells K.D., Powell A.M. (2000) Blastomeres from somatic cell nuclear transfer embryos are not allocated randomly in chimeric blastocysts. Cloning. 2(1):9-22.

Wen F., Tynan J.A., Cecena G., Williams R., Múnera J., Mavrothalassitis G., Oshima R.G. (2007) Ets2 is required for trophoblast stem cell self-renewal. Dev Biol. 312(1):284-99.

Werner A., Carlile M., Swan D. (2009) What do natural antisense transcripts regulate? RNA Biol. 6(1):43-8.

Wilmut I., Schnieke A.E., McWhir J., Kind A.J., Campbell K.H. (1997) Viable offspring derived from fetal and adult mammalian cells, Nature. 85(6619):810-3.

Wrenzycki C., Herrmann D., Lucas-Hahn A., Lemme E., Korsawe K., Niemann H. (2004) Gene expression patterns in in vitro-produced and somatic nuclear transfer-derived preimplantation bovine embryos: relationship to the large offspring syndrome? Anim Reprod Sci. 82-83:593-603.

Wu Q., Chen X., Zhang J., Loh Y.H., Low T.Y., Zhang W., Zhang W., Sze S.K., Lim B., Ng H.H. (2006) Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem. 281(34):24090-4.

Xie S., Green J., Bixby J.B., Szafranska B., DeMartini J.C., Hecht S., Roberts R.M. (1997) The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes. Proc Natl Acad Sci U S A. 94(24):12809-16.

Xie, S., Low, B.G., Nagel, R.J., Kramer, K.K., Anthony, R.V., Zoli, A.P., Beckers, J.F., Roberts, R.M. (1991) Identification of the major pregnancy-specific antigens of cattle and sheep as inactive members of the aspartic proteinase family. Proc. Natl. Acad. Sci. U.S.A. 88, 10247–10251.

Yagi R., Kohn M.J., Karavanova I., Kaneko K.J., Vullhorst D., DePamphilis M.L., Buonanno A. (2007) Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development. 134(21):3827-36.

Yamada K., Kanda H., Tanaka S., Takamatsu N., Shiba T., Ito M. (2006) Sox15 enhances trophoblast giant cell differentiation induced by Hand1 in mouse placenta. Differentiation. 74(5):212-21.

Yang J., Chai L., Fowles T.C., Alipio Z., Xu D., Fink L.M., Ward D.C., Ma Y. (2008) Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. Proc Natl Acad Sci U S A. 105(50):19756-61.

Yu Z., Morais D., Ivanga M., Harrison P.M. (2007) Analysis of the role of retrotransposition in gene evolution in vertebrates. BMC Bioinformatics. 8:308.

Yuan P., Han J., Guo G., Orlov Y.L., Huss M., Loh Y.H., Yaw L.P., Robson P., Lim B., Ng H.H. (2009) Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev. 23(21):2507-20.

Zaiss D.M., Kloetzel P.M. (1999) A second gene encoding the mouse proteasome activator PA28beta subunit is part of a LINE1 element and is driven by a LINE1 promoter. J Mol Biol. 287(5):829-35.

Zernicka-Goetz M. (2006) The first cell-fate decisions in the mouse embryo: destiny is a matter of both chance and choice. Curr Opin Genet Dev. 16(4):406-412.

Zhang J., Tam W.L., Tong G.Q., Wu Q., Chan H.Y., Soh B.S., Lou Y., Yang J., Ma Y., Chai L., Ng H.H., Lufkin T., Robson P., Lim B. (2006) Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1.Nat Cell Biol. 8(10):1114-23.

Zhang J., Wang X., Li M., Han J., Chen B., Wang B., Dai J. (2006) NANOGP8 is a retrogene expressed in cancers. FEBS J. 273(8):1723-30.

Zhang M., Guller S., Huang Y. (2007) Method to enhance transfection efficiency of cell lines and placental fibroblasts. Placenta. 28(8-9):779-82.

Zhang Y., Goss A.M., Cohen E.D., Kadzik R., Lepore J.J., Muthukumaraswamy K., Yang J., DeMayo F.J., Whitsett J.A., Parmacek M.S., Morrisey E.E. (2008) A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. Nat Genetics. 40(7):862-70