

EXPRESSION AND ROLES OF BLASTOCYST LINEAGE-DETERMINING GENES  
DURING SOMATIC CELL REPROGRAMMING

By

Alexandra Moauro

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Physiology - Doctor of Philosophy

2022

## ABSTRACT

### EXPRESSION AND ROLES OF BLASTOCYST LINEAGE-DETERMINING GENES DURING SOMATIC CELL REPROGRAMMING

By

Alexandra Moauro

In order to properly use stem cells, it is important that we first understand how these cells are established and maintained. One of the most widely used stem cells are induced pluripotent stem cells (iPSCs) which provide great therapeutic promise and a novel source of ethical stem cells for research models. iPSCs are created by overexpression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*) in a somatic cell. As studies have sought to improve reprogramming efficiency and develop the most embryonically identical stem cells, our lab has uncovered that *OSKM* is not a specific cocktail for pluripotency formation. Instead *OSKM* induces additional cell fates including the formation of a multipotent stem cell termed induced extraembryonic endoderm stem (iXEN) cells. This raises the question as to how two distinct stem cell types arise in parallel. Interestingly, in embryo development we observe the same pluripotent and multipotent extraembryonic endoderm lineages form in parallel. Using our knowledge of normal embryo development, I set out to identify what blastocyst lineage markers can help us identify early iPSC and iXEN colonies as they start to form and mature. Of these markers, we observed that endogenous OCT4 is expressed in both iXEN and iPSC colonies. Based on the expression pattern of the key embryonic transcription factor, OCT4, we further focused on how this transcription factor may have a dual role in establishing iPSC and iXEN fates. Lastly, we altered the reprogramming cocktail using additional embryonic transcription factors to determine how these factors affect the propensity for pluripotency or extraembryonic endoderm fate.

## ACKNOWLEDGEMENTS

### Family

I would like to thank my close and extended family for their unconditional love and support throughout the years. To leave my main support system and move halfway across the country has been a long and taxing journey. I cherish your understanding, advice, and conversations throughout this time. I would like to say a special thank you to my parents, Mary and Paul Moauro, who always believed that I could achieve anything and for giving me every chance they could to accomplish my goals no matter how lofty. I would also like to thank my partner, Greg Forman, for his endless care, patience, and for never letting me doubt myself. I would even like to thank Onyx for the endless distractions and for being (mostly) a good boy.

### Friends

I would like to thank my friends, old and new, for their support and the sanity that they have given me. From every conversation to each outing, it has been an immense help in keeping me grounded and refreshed during this process. You have also taught me a lot about this world and have provided me with unique perspectives that I am forever grateful to have seen.

### Mentors

To every mentor from elementary school to graduate school and all my extracurriculars, I would like to say thank you for seeing my potential even when I did not. As a female first

generation college student, I am forever grateful for the support that you have provided me. I would have never thought at the beginning of college that I could have or would have gone on to pursue a terminal degree. I appreciate each of you for the knowledge and resources that you have provided me to get me to where I am.

### Ralston Lab

I would like to give a special thank you to the Ralston Lab for providing me the resources that I needed to complete my training and giving me the space I needed to grow as a researcher. I appreciate every piece of advice and personal connection that I made along the way from new to past lab members. Moving forward I am excited to see where everyone's success takes them.

## TABLE OF CONTENTS

|   |     |
|---|-----|
| LIST OF TABLES .....  | ix  |
| LIST OF FIGURES .....   | x   |
| KEY TO ABBREVIATIONS .....  | xii |
| CHAPTER 1. An overview of blastocyst derived and induced stem cells .....   | 1   |
| Abstract .....  | 2   |
| Section 1.1. Stem cells and their origin .....  | 3   |
| Stem cells .....  | 3   |
| Early embryonic development and the creation of embryo-derived stem cells .....   | 5   |
| Cell reprogramming and the creation of induced stem cells .....   | 9   |
| Section 1.2. Overview of the mechanism behind OSKM reprogramming ..   | 11  |
| Review of transcription factors OCT4, SOX2, KLF4 and c-MYC ....   | 11  |
| What is known about the mechanism of OSKM somatic cell reprogramming .....  | 14  |
| Section 1.3. The discovery and promise of iXEN .....  | 19  |
| Discovery of iXEN .....   | 19  |
| Significance of embryonic and induced XEN .....   | 21  |
| iXEN and iPSCs as a key to understanding OSKM somatic cell reprogramming .....  | 23  |
| APPENDIX .....  | 25  |
| CHAPTER 2. Utilization of molecular and cellular techniques to distinguishing between extraembryonic endoderm and pluripotent stem cells during OSKM somatic cell reprogramming ..... | 32  |
| Abstract .....  | 33  |
| Introduction .....  | 34  |
| Materials .....   | 35  |
| Media Preparation .....   | 35  |
| Preparing Mouse Embryonic Fibroblasts (MEFs) .....  | 36  |
| Preparing Replication-incompetent Retroviruses for Overexpression of OSKM for Reprogramming .....   | 37  |
| Viral Titer of Retrovirus .....   | 37  |
| OSKM Viral Reprogramming .....  | 38  |
| Picking and Passaging iXEN and iPSC Colonies .....  | 38  |
| Confocal Imaging of Reprogramming Cells .....   | 39  |
| Cell Sorting of Reprogramming Cells .....   | 39  |
| RNA-seq of Passaged Cell Lines .....  | 40  |
| scRNA-seq of Reprogramming Cells .....  | 41  |
| Methods .....   | 41  |
| Preparing Mouse Embryonic Fibroblasts (MEFs) .....  | 41  |

|  |     |
|--|-----|
| Preparing Replication-incompetent Retroviruses for<br>Overexpression of OSKM for Reprogramming .....   | 43  |
| Viral Titer of Retrovirus .....  | 44  |
| OSKM Viral Reprogramming .....   | 45  |
| Picking and Passaging iXEN and iPSC Colonies .....   | 46  |
| Fluorescently Activated Cell Sorting (FACS) of Reprogramming<br>Cells .....  | 46  |
| Confocal Imaging of Reprogramming Cells .....  | 48  |
| RNA-seq of Passaged Cell Lines .....   | 49  |
| scRNA-seq of Reprogramming Cells .....   | 52  |
| Notes .....  | 55  |
| Methods Section .....  | 55  |
| Final Thoughts on Analyzing iXEN and iPSC .....  | 58  |
| APPENDIX .....   | 60  |
|  |     |
| CHAPTER 3. A Closer examination of fluorescent reporters NANOG, OCT4,<br>GATA6 and GATA4 during somatic cell reprogramming reveals unexpected<br>expression in multiple colony types ..... | 69  |
| Abstract .....   | 70  |
| Introduction .....   | 71  |
| Materials & Methods .....  | 73  |
| Creation of <i>Nanog</i> <sup>mCherry</sup> Mouse .....  | 73  |
| Immunofluorescence and Confocal Microscopy .....   | 74  |
| Mouse Strains .....  | 74  |
| Mouse Embryonic Fibroblast (MEF) Preparation .....   | 75  |
| Reprogramming .....  | 75  |
| Colony Counting and Lab Images .....   | 76  |
| RNA isolation and qPCR .....   | 76  |
| Results .....  | 76  |
| Nanog-2A-mCherry fluorescent reporter mouse line .....   | 76  |
| NANOG-mCherry and OCT4-eGFP expression alone cannot<br>reliably identify putative iPSC colonies .....  | 77  |
| GATA6-H2B-Venus is an unreliable and non-specific reporter for<br>putative iXEN colony identification .....  | 81  |
| GATA4-H2B-eGFP is minimally expressed during reprogramming<br>but does not express in putative iPSC colonies .....   | 82  |
| Discussion .....   | 84  |
| Acknowledgments .....  | 88  |
| APPENDIX .....   | 89  |
|  |     |
| CHAPTER 4. OCT4 is expressed in cells fated for extraembryonic endoderm<br>formation during somatic cell reprogramming .....   | 104 |
| Abstract .....   | 105 |
| Introduction .....   | 106 |
| Materials & Methods .....  | 107 |
| Mouse Strains .....  | 107 |
| Mouse Embryonic Fibroblast (MEF) Preparation .....   | 107 |

|   |     |
|---|-----|
| Reprogramming .....   | 107 |
| Colony Counting and Lab Images .....  | 108 |
| Immunofluorescence and Confocal Microscopy .....  | 109 |
| qPCR and RNA-seq .....  | 109 |
| scRNA-seq .....   | 111 |
| Single Cell Sorting .....   | 114 |
| Visceral Endoderm Formation .....   | 114 |
| Results .....   | 115 |
| SKM induces the formation of iXEN cells and is sufficient to<br>activate endogenous OCT4 .....                                    | 115 |
| A population of cells expressing <i>Oct4</i> display an embryonic<br>primitive endoderm gene signature during reprogramming ..... | 118 |
| Endogenous <i>Oct4</i> is expressed within iXEN cell colonies during<br>reprogramming .....                                       | 121 |
| Endogenous <i>Oct4</i> is temporarily expressed within iXEN cells .....   | 123 |
| Discussion .....  | 126 |
| Acknowledgements .....  | 129 |
| APPENDIX .....  | 130 |
|   |     |
| CHAPTER 5. Evaluating the ability of exogenous <i>Sall4</i> to replace <i>Oct4</i> in<br>somatic cell reprogramming .....         | 147 |
| Abstract .....  | 148 |
| Introduction .....  | 149 |
| Materials & Methods .....   | 151 |
| Mouse Strains .....   | 151 |
| Mouse Embryonic Fibroblast (MEF) Preparation .....  | 151 |
| Reprogramming .....   | 152 |
| Colony Counting and Lab Images .....  | 153 |
| RNA Isolation and qPCR .....  | 153 |
| Results .....   | 153 |
| Exogenous <i>Sall4</i> is not sufficient to replace <i>Oct4</i> in the formation<br>of iPSC or iXEN cells .....                   | 153 |
| Exogenous <i>Nanog</i> plus <i>Sall4</i> are not sufficient to replace <i>Oct4</i> in<br>somatic cell reprogramming .....         | 156 |
| <i>Nanog</i> + <i>Sall4</i> , in conjunction with <i>SKM</i> , induces the expression of<br>extraembryonic endoderm markers ..... | 157 |
| Discussion .....  | 158 |
| Acknowledgements .....  | 160 |
| APPENDIX .....  | 161 |
|   |     |
| CHAPTER 6. Where to explore next in the complicated landscape of<br>reprogramming?.....   | 168 |
| Abstract .....  | 169 |
| Introduction .....  | 170 |
| Continued analysis of cell fate decisions using single-cell and longitudinal<br>studies .....                                     | 171 |

|  |     |
|--|-----|
| Using the second cell fate decision of embryo development to guide reprogramming studies ..... | 173 |
| Reanalyzing reprogramming studies to include analysis of iXEN and iTSCs .....                  | 174 |
| Evaluating the potential of iXEN cells .....   | 176 |
| Conclusions .....  | 177 |
| APPENDIX .....   | 179 |
| CURRICULUM VITAE .....   | 185 |
| Education and Professional History .....   | 186 |
| Education .....  | 186 |
| Professional and Academic Positions .....  | 186 |
| Honors and Awards .....  | 187 |
| Research .....   | 187 |
| Projects .....   | 187 |
| Future Goals and Research Interests .....  | 188 |
| Scholarship .....  | 188 |
| Papers .....   | 188 |
| Posters .....  | 189 |
| Presentations .....  | 190 |
| Grants .....   | 190 |
| Service .....  | 191 |
| Offices Held in Professional Organizations .....   | 191 |
| Volunteering .....   | 191 |
| Mentorship .....   | 191 |
| Outreach .....   | 192 |
| REFERENCES .....   | 193 |

## LIST OF TABLES

|   |     |
|---|-----|
| Table 2.1. qRT-PCR primers for detecting endogenous and viral transcripts. .... | 61  |
| Table 2.2. Antibodies for fluorescent imaging. ....                             | 62  |
| Table 3.1. qPCR primers for detecting endogenous transcripts. ....              | 102 |
| Table 3.2. Genotyping primers. ....   | 103 |
| Table 4.1. Top GO term pathways upregulated in OCT4-eGFP iXEN and XEN...        | 143 |
| Table 4.2. qPCR primers for detecting endogenous transcripts. ....              | 146 |
| Table 5.1. qPCR primers for detecting endogenous transcripts. ....              | 166 |
| Table 5.2. Genotyping primers. ....   | 167 |

## LIST OF FIGURES

|   |     |
|---|-----|
| Figure 1.1. Stem cell potency. ....   | 26  |
| Figure 1.2. Embryo development and the formation of embryonic stem cells. ....  | 28  |
| Figure 1.3. Somatic cell reprogramming. ....  | 30  |
| Figure 1.4. <i>Fbx15</i> gene expression in embryonic derived stem cell lines. ....   | 31  |
| Figure 2.1. iXEN and iPSC formation in OSKM reprogramming. ....   | 63  |
| Figure 2.2. Passaged iXEN and iPSC morphology. ....   | 65  |
| Figure 2.3. Confocal images of colonies on day 14 of reprogramming. ....  | 67  |
| Figure 2.4. Examples of cells on a hemacytometer before scRNA-seq library preparation. ....                                 | 68  |
| Figure 3.1. NANOG-mCherry expression during reprogramming shows specificity in cell lines but not colonies. ....            | 90  |
| Figure 3.2. Reprogramming with NANOG-mCherry & OCT4-eGFP reporters show the same fluorescence pattern and specificity. .... | 92  |
| Figure 3.3. GATA6-H2B-Venus expression during reprogramming shows specificity in cell lines but not colonies. ....          | 93  |
| Figure 3.4. GATA4-H2B-eGFP expression during reprogramming shows specificity in cell lines but not colonies. ....           | 95  |
| Figure 3.5. Fluorescent reporter summary. ....  | 97  |
| Figure 3.6. <i>Nanog</i> <sup>mCherry/+</sup> reporter creation and testing. ....   | 99  |
| Figure 3.7. <i>Nanog</i> <sup>mCherry/+</sup> embryos. ....   | 101 |
| Figure 4.1. OSKM and SKM reprogramming produces iXEN colonies that can be expanded to create stable cell lines. ....        | 131 |
| Figure 4.2. SKM and OSKM derived iXEN produce visceral endoderm and are transcriptionally indistinguishable. ....           | 133 |
| Figure 4.3. <i>Oct4</i> is expressed alongside early embryo development genes during OSKM reprogramming. ....               | 135 |

|  |     |
|--|-----|
| Figure 4.4. OCT4-eGFP is expressed in reprogramming somatic cells fated for different identities. ....   | 137 |
| Figure 4.5. Single cell sorting reveals that OCT4-eGFP expressing reprogramming cells are fated for non-pluripotent cell types. ....                           | 139 |
| Figure 4.6. OCT4-eGFP single cell sorted iXEN lines express the same extraembryonic endoderm markers and differentiation potential as embryo-derived XEN. .... | 141 |
| Figure 4.7. Expression of NANOG-mCherry and iPSC Markers in SKM and OSKM Reprogramming. ....   | 144 |
| Figure 4.8. Single cell sorting gating strategy, cells counted and cell growth outcomes.....   | 145 |
| Figure 5.1. <i>Sall4</i> +SKM somatic cell reprogramming failed to produce iPSC or iXEN colonies. ....   | 162 |
| Figure 5.2. <i>Nanog</i> + <i>Sall4</i> +SKM somatic cell reprogramming failed to produce iPSC or iXEN colonies. ....  | 164 |
| Figure 5.3. <i>Nanog</i> + <i>Sall4</i> +SKM somatic cell reprogramming leads to the expression of extraembryonic endoderm markers. ....                       | 165 |
| Figure 6.1. Embryo development and the formation of embryonic stem cells. ....   | 180 |
| Figure 6.2. Second cell fate decision in mouse embryo development. ....  | 182 |
| Figure 6.3. Human and mouse early embryo development. ....   | 184 |

## KEY TO ABBREVIATIONS

|           |  |
|-----------|--|
| AST-seq   | <b>A</b> llele <b>s</b> pecific <b>t</b> argeted <b>s</b> equencing                            |
| ATAC-seq  | <b>A</b> ssay for <b>t</b> ransposase- <b>a</b> ccessible <b>c</b> hromatin <b>s</b> equencing |
| BMP4      | <b>B</b> one <b>m</b> orphogenetic <b>p</b> rotein <b>4</b>                                    |
| ChIP-seq  | <b>C</b> hromatin <b>i</b> mmunoprecipitation <b>s</b> equencing                               |
| EPI       | <b>E</b> piblast   |
| ESC/ ES   | <b>E</b> mryonic <b>s</b> tem cells  |
| Exp.      | <b>E</b> xpression   |
| FACS      | <b>F</b> luorescently- <b>a</b> ctivated <b>c</b> ell <b>s</b> orting                          |
| FBS       | <b>F</b> etal <b>b</b> ovine <b>s</b> erum   |
| FGF4      | <b>F</b> ibroblast <b>g</b> rowth <b>f</b> actor <b>4</b>                                      |
| Fig.      | <b>F</b> igure   |
| ICM       | <b>I</b> nnner <b>c</b> ell <b>m</b> ass   |
| iPS/ iPSC | <b>I</b> nduced <b>p</b> luripotent <b>s</b> tem cells   |
| iTSC      | <b>I</b> nduced <b>t</b> rophoblast <b>s</b> tem <b>c</b> ells                                 |
| iXEN      | <b>I</b> nduced <b>e</b> xtraembryonic <b>e</b> ndoderm  |
| KM        | <i><b>K</b>lf4, <b>c</b>-<b>M</b>yc</i>  |
| LIF       | <b>L</b> eukemia <b>i</b> nhibitory <b>f</b> actor   |
| MACS      | <b>M</b> agnetic- <b>a</b> ctivated <b>c</b> ell <b>s</b> orting                               |
| MEF       | <b>M</b> ouse <b>e</b> mryonic <b>f</b> ibroblast  |
| MMLV      | <b>M</b> oloney <b>m</b> urine <b>l</b> eukemia <b>v</b> irus                                  |
| Neg.      | <b>N</b> egative   |

|               |   |
|---------------|---|
| OS            | <b>Oct4, Sox2</b>   |
| OSK           | <b>Oct4, Sox2, Klf4</b>   |
| OSKM          | <b>Oct4, Sox2, Klf4, c-Myc</b>  |
| PE**          | <b>Primitive endoderm</b>   |
| Pos.          | <b>Positive</b>   |
| PrESCs        | <b>Primitive endoderm stem cells</b>  |
| PS            | <b>Primitive streak</b>   |
| pXEN          | <b>Primitive extraembryonic endoderm stem cells</b>   |
| qPCR/ qRT-PCR | <b>Quantitative polymerase chain reaction/ quantitative reverse transcription polymerase chain reaction</b> |
| RNA-seq       | <b>RNA sequencing</b>   |
| scRNA-seq     | <b>Single cell RNA sequencing</b>   |
| SKM           | <b>Sox2, Klf4, c-Myc</b>  |
| TE            | <b>Trophectoderm</b>  |
| TS/TSC        | <b>Trophoblast stem Cells</b>   |
| VE            | <b>Visceral endoderm</b>  |
| XEN           | <b>Extraembryonic endoderm stem cells</b>   |

\*\* Note: Depending on the context PE can also mean extraembryonic endoderm

## CHAPTER 1.

An overview of blastocyst derived and induced stem cells

Moauro A. and Ralston A.

A. Moauro wrote the chapter and assembled the figures.

A. Ralston edited the chapter.

## **Abstract**

Stem cells are unique cells that are defined by their ability to self-renew and differentiate. Cells with a greater potency, or differentiation ability, can produce more cell types. Because of this, totipotent or pluripotent cells have been highly sought after for research models and stem cell therapies. The use of embryonic stem cells produces ethical and immunological concerns, while the use of adult stem cells is limited based on a restricted differentiation ability. These restraints made it challenging for stem cell research to progress. However, this changed in 2006 with the discovery that overexpression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*) in somatic cells could induced pluripotent stem cells (iPSCs). iPSCs changed the way in which pluripotent stem cells were collected and avoids the concerns of embryonic stem cell use. The discovery of iPSCs significantly propelled stem cell research forward thus launching many studies into how *OSKM* can function to create a pluripotent cell fate. Through this exploration, gene targets of *OSKM* and how *OSKM* binds to chromatin has been uncovered. This mechanistic exploration of *OSKM* also revealed that *OSKM* is not a precise cocktail for the induction of pluripotency, but instead produces a second multipotent extraembryonic stem cells type called induced extraembryonic endoderm stem (iXEN) cells. The discovery of iXEN has provided a new induced stem cell with unknow therapeutic promise and has raised the need for better understanding of the processes that govern final cell fate decisions during *OSKM* reprogramming. This chapter provides a broad overview of stem cells including their function and formation with a special emphasis on iPSC and iXEN formation during *OSKM* reprogramming.

## **Section 1.1. Stem cells and their origin**

### ***Stem cells***

Stem cells are a unique type of cell that have the capability to make more copies of itself through self-renewal and can differentiate to form another cell type. Stem cells are naturally found in embryos, fetuses, and adults. There are many different types of stem cells. A way to define a stem cell is through its differentiation potential. A stem cell that can give rise to any type of cell, including embryonic and extraembryonic lineages, is called a totipotent cell. There are also pluripotent stem cells, which can give rise to all three germ layers. These layers include the endoderm, mesoderm and ectoderm which can give rise to important structures such as the gut, heart, and skin respectively. After pluripotent follows multipotent, oligopotent and unipotent stem cells which can give rise to several, few or one cell type, respectively (Kolios & Moodley, 2012; S. Yamanaka, 2020; Zakrzewski et al., 2019) (Fig 1.1 A).

Given that stem cells can self-renew and differentiate, they serve as an excellent model for research. These cells can be studied to determine pathways involved in the maintenance of identity and self-renewal. They can also be taken out of their stem cell state and forced to differentiate, elucidating pathways that are important in establishing a new cell identity (Kolios & Moodley, 2012; Rowe & Daley, 2019). Many have also used stem cells to establish *in vitro* models of disease or create rare populations of cells in large quantities. These models can then be used in high throughput genetic and drug screens (Chien, 2008; Rowe & Daley, 2019). As researchers strive for better models that mimic disease and human physiology, they have developed more complex *ex vivo*

models. *In vivo*, cells are often not isolated and frequently coexist with other cells, extracellular matrices, and pathogens. This has pushed scientists to develop novel 3D tissue culture models. One of the most notable accomplishments is the establishment of 3D multicellular aggregates derived from stem cells and extracellular matrix to create mini organs termed organoids (Rowe & Daley, 2019).

Not only do stem cells serve as important research models, but they are used in medical treatments. To date, there are only a handful of FDA approved stem cell therapies. Approved therapies include the use of cord blood and hematopoietic stem cells (HSC) when used to treat appropriate disease ([www.fda.gov](http://www.fda.gov)). Although there are limited approved stem cell therapies, this continues to be an active area of research. Currently, there are over 600 active clinical trials through the NIH that explore the use of different types of stem cells to treat disease ([clinicaltrials.gov](http://clinicaltrials.gov)). Part of the reason stem cell therapies are highly sought after is that aging leads to irreversible damage and impaired organ function, resulting in disease. The hope is that stem cells can be used to replace these damaged pools of cells to restore function. Examples of degenerative disease that regenerative medicine is aiming to treat include Parkinson's disease, macular degeneration, heart failure, spinal cord injuries, cartilage defects and type I diabetes (S. Yamanaka, 2020).

To develop stem cell therapies, one must consider the type of stem cell to use and where the cell can be harvested. There are three main types of stem cell sources: embryo-derived stem cells, induced stem cells, and adult stem cells (Kolios & Moodley, 2012).

Embryo-derived stem cells are created from cultured embryos. Commonly used embryo-derived stem cells are pluripotent embryonic stem cells (ESCs). Another common way to collect stem cells is to force a differentiated cell into an undifferentiated stem cell state through cell reprogramming (Kolios & Moodley, 2012; S. Yamanaka, 2020). Frequently used induced stem cells are induced pluripotent stem cell (iPSC). Lastly, adult stem cells with limited differentiation potential can be collected from different organs (Kolios & Moodley, 2012). Examples of adult-derived stem cells include HSCs and mesenchymal stem cells (MSC). For both research and clinical trials, pluripotent stem cells are often used due to the unlimited expansion and myriad of differentiation potential (S. Yamanaka, 2020). To date, there are just under 200 clinical trials registered through the NIH that have used pluripotent stem cells (clinicaltrials.gov). Given that embryonic and induced stem cells are commonly used and highly sought after in research and clinical trials, it is important to understand their origins in order to understand their potential and promise.

### ***Early embryonic development and the creation of embryo-derived stem cells***

Understanding preimplantation embryo development tells us not only where embryo-derived stem cells originate but how different stem cells are formed, maintained, and differentiated *in vivo*. In research, embryonic development is often evaluated through a mouse model due to the developmental similarities to humans and the ability to genetically modify the mouse embryo. The mouse embryo is first formed at embryonic day 0 (E0), when sperm and egg come together to form a single cell called a zygote. From E0-2.5, the zygote then goes on to make several cleavage divisions, forming a ball of totipotent cells called a morula. By E2.75, or the 8-cell stage, the morula undergoes

compaction. During compaction, cells start to form an apical-basal polarity in which the apical surface faces the outside of the morula and the basal surface faces other cells of the morula. As the cells divide to form 16-cells (E2.75-3.0), they do so along the apical-basal domain. Cells containing the apical domain become outside cells and cells containing the basal domain become inside cells (Fleming, 1987; Johnson & McConnell, 2004; Lokken & Ralston, 2016; Plusa et al., 2005). This is the first point in development in which cells make their first cell fate decision and decide to become an inside cell or an outside cell (Fig. 1.2 A). The first cell fate decision is important as it provides insight into the loss of totipotency.

During the 16-cell stage, outside cells are also known as trophoblast cells. These cells play an important part in forming the placenta (Hemberger et al., 2020; Lokken & Ralston, 2016). Given trophoblast cell's restricted differentiation potential, it is now a multipotent cell type. As the embryo continues to develop, the trophoblast will go on to expand and differentiate to form the many cell types of the placenta allowing for critical maternal-fetal nutrient exchange and providing hormones to support the growing embryo (Hemberger et al., 2020). Commonly used markers of trophoblast include CDX2, TEAD4, GATA3, ELF5, YAP1 and EOMES (Hemberger et al., 2020; Lokken & Ralston, 2016).

During the 16-cell stage, inside cells become what is called the inner cell mass. These cells are no longer totipotent as they cannot give rise to trophoblast, but they have an expanded pluripotent potential. The trophoblast and inner cell mass continue to grow

and undergo cavitation at E3.0 to form a blastocyst. A blastocyst is a hollow sphere of trophectoderm cells in which a cluster of inner cell mass cells resides within one side of the trophectoderm sphere. By E3.75, the inner cell mass must make the second cell fate decision. During this decision, cells of the inner cell mass must decide to contribute to either the epiblast or primitive endoderm in a salt and pepper-like pattern. This is accomplished through RTK-ERK and PIK3 signaling. Cells of the epiblast express and secrete FGF4 to transiently and sporadically activate their own FGFR1/ERK signaling pathway. Cells of the primitive endoderm express FGFR1 receptors to receive a constant FGF4 signal to activate ERK while in parallel activating RTK and PI3K signaling through FGFR2 and PDGFRFA receptors (Azami et al., 2019; Bessonard et al., 2019; Chazaud et al., 2006; G. Guo et al., 2010; Kang et al., 2013, 2017; Molotkov et al., 2017; Simon et al., 2021; Y. Yamanaka et al., 2010). Once the epiblast and primitive endoderm have formed, they will continue to divide. Primitive endoderm cells then go on to migrate and separate out from the epiblast. The epiblast cells will continue to form a small clump of cells attached to the trophectoderm while the primitive endoderm forms a single cell layer lining the bottom of the epiblast (Lokken & Ralston, 2016) (Fig 1.2 A). The second cell fate decision provides important insight into the formation and loss of pluripotency.

The epiblast cells are pluripotent cells that can give rise to all cells of the fetus. In embryo development, the epiblast will continue to grow and eventually differentiate to form the three germ layers (endoderm, mesoderm, and ectoderm). Commonly used markers of the epiblast include OCT4, SOX2 and NANOG (Bassalart et al., 2018; Lokken & Ralston, 2016).

The primitive endoderm plays an important role in forming extraembryonic structures and provides important signals to the developing embryo (Belaoussoff et al., 1998; Stern & Downs, 2012; Stuckey et al., 2011; Thomas & Beddington, 1996). After the inner cell mass differentiates and forms the primitive endoderm, cells committed to this new lineage have lost the ability to contribute to all three germ layers and instead become multipotent (Nowotschin & Hadjantonakis, 2020). In development, the primitive endoderm will continue to grow and differentiate to form the visceral endoderm and parietal endoderm. Commonly used markers for primitive endoderm cells include GATA6, GATA4, SOX7, SOX17, PDGFR $\alpha$  and DAB2 (Bassalart et al., 2018; Lokken & Ralston, 2016).

During early embryo development, embryos can be grown *ex vivo* in media containing specific inhibitors, small molecules, and growth factors to select for different stem cell populations. Trophectoderm cells can be expanded to create trophoctoderm stem (TS) cells (Tanaka et al., 1998). Epiblast cells can be expanded to create embryonic stem cells (ESCs) (Evans & Kaufman, 1981; Martin, 1981). Primitive endoderm cells can be expanded to create extraembryonic endoderm (XEN) stem cells (Figure 1.2 A and B) (Kunath et al., 2005; Niakan et al., 2013). Each embryo-derived stem cell line maintains the appropriate gene expression and differentiation potential relative to their embryonic counterparts. Embryonic studies have provided insight into how to maintain and collect ESC, TS, and XEN cells. However, the use of ESC, TS and XEN cells has allowed us to better investigate pathways and transcription factors that are important in proper development.

### ***Cell reprogramming and the creation of induced stem cells***

Although stem cells can be collected from embryos and have been valuable tools as research models, there is a therapeutic limit to their use. One limitation is due to ethical concerns surrounding their source. As discussed, embryonic stem cells are derived from the embryo and current means of creating stem cell lines results in the sacrifice of an embryo. This creates ethical concerns when wanting to develop human embryonic stem cells. The second issue is that since stem cells are created from an embryo, each stem cell line has a unique major histocompatibility complex that may not match the recipient. This mismatch could result in immune rejection when administered as a therapy to a patient. These reasons created a need for induced stem cells in which a fully differentiated cell could be taken from a patient and converted into a stem cell.

However, this idea was not thought to be possible throughout the 19<sup>th</sup> and 20<sup>th</sup> century as it was previously postulated by August Weismann that cells only involved in inheritance maintain the entire genetic code and that non-germline somatic cells must discard unnecessary genes (Blaine et al., 2020). Additionally, Conrad Waddington hypothesized that differentiating cells are like balls rolling down a hill and that it would take an impossible amount of energy for a differentiated cell to roll back up the hill and undifferentiate (Waddington, 1942).

This changed in 1962 when the nucleus of a somatic cell was transferred into an enucleated egg which resulted in the formation of an embryo (Gurdon et al., 1958) (Fig 1.3. A). This pivotal discovery of somatic cell nuclear transfer showed that fully

differentiated cells maintain the necessary genetic components to turn on a totipotent state. This discovery was followed by cell fusion studies in which a differentiated cell was fused to an embryonic stem cell to create a pluripotent cell fate in the differentiated cell (Cowan et al., 2005; Tada et al., 2001). Cell fusion studies demonstrated that cellular factors could change a cell's identity. The identity of these factors was not uncovered until the establishment of transdifferentiation. Using knowledge from embryo development, researchers were able to identify critical transcription factors that guide differentiation. By overexpressing a key transcription factor, a cell could take on a new identity. Examples include the overexpression of MYOD, GATA1 and CBEP to create myoblasts, common lymphoid progenitors and macrophages respectively (Davis et al., 1987; Kulesa et al., 1995; Xie et al., 2004). This demonstrated that overexpressing select transcription factors leads to new cell identities.

These fundamental studies paved the way for the Shinya Yamanaka Lab to show that overexpression of key transcription factors can be used to create an induced pluripotent stem cell (Fig 1.3. A). Using an elegant study, the Yamanaka Lab in 2006 developed a list of 24 candidate transcription factors that each play an important role in ESC maintenance and identity. They overexpressed all 24 transcription factors using a retroviral delivery system to induce pluripotent gene expression. To only select for cells that could activate a pluripotent state, they created a fibroblast cell line that had a neomycin resistance gene inserted into a dispensable pluripotency gene, *Fbx15*. Any time a pluripotency state was achieved, the cell would metabolize neomycin and proliferate uninhibited in the high surrounding neomycin concentration. Once the lab was able to

achieve a pluripotent state with all 24 genes, they decreased the number of transcription factors necessary for iPSC formation, using an n-1 approach in which they removed one transcription factor at a time. Using this approach, they were able to determine that *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*) in combination were sufficient to establish a stable pluripotent network in differentiated cells after 2-3 weeks (Takahashi & Yamanaka, 2006).

To this day, *OSKM* reprogramming remains a widely used technique. Unfortunately, the efficiency of this process is low with only 1% of cells successfully reprogramming to become iPSCs. The remaining 99% of the cells are thought to stay as fibroblasts or exist in an alternate non-pluripotent state (S. Yamanaka, 2009). Due to this low efficiency, the field continues to study the mechanism of *OSKM* and develop new cocktails with improved efficiencies. Additionally, there is therapeutic concern over the use of reprogramming cells with DNA integrating retrovirus and the use of oncogenes as part of the reprogramming cocktail. This has led to the development of reprogramming through small molecule signaling and development on non-integrating viral delivery systems (Fusaki et al., 2009; Hou et al., 2013) (Fig 1.3. A).

## **Section 1.2. Overview of the mechanism behind *OSKM* reprogramming**

### ***Review of transcription factors OCT4, SOX2, KLF4 and c-MYC***

*Oct4*, *Sox2*, *Klf4* and *c-Myc* are all genes that encode for different transcription factors. These genes were first identified for the construction of a pluripotency cocktail based on their role in forming and maintaining ESCs (Takahashi & Yamanaka, 2006). Octamer-binding transcription factor 4, OCT4, is transcribed by the *Pou5f1/Oct4* gene. OCT4 is

part of the POU family of transcription factors which includes Pit-1, Oct1 and Unc-86. All POU members contain a conserved POU domain which is responsible for binding to DNA. The two subdomains of OCT4, POU-specific and POU homeodomain, allow for the fine-tuned binding of chromatin targets. OCT4 can also work alone or in conjunction with other transcription factors to promote transcription (Rizzino & Wuebben, 2016). OCT4 plays an extensive role in embryo development. It is present in the oocyte, morula, inner cell mass, epiblast, primitive endoderm and germ cells (Palmieri et al., 1994; Xiao et al., 2016). In conjunction with NANOG and SOX2, OCT4 plays an important role as a core transcription factor in pluripotency maintenance and formation (Patra, 2020). OCT4 was thought to play an essential part in reprogramming due to the limited success of iPSC formation when removing it from the *OSKM* cocktail (Y. Li et al., 2011; Zhu et al., 2010). Eventually, it was shown that OCT4 could be replaced by homologs or removed entirely when using a lentiviral delivery system (Jerabek et al., 2014; Nakagawa et al., 2008; Velychko et al., 2019). Although *Oct4* can be removed from *OSKM* reprogramming, endogenous OCT4 expression is still required for pluripotency maintenance (Velychko et al., 2019). Based on the necessity of either endogenous OCT4 activation or exogenous OCT4 delivery for successful reprogramming, exogenous *Oct4* is often kept in the different variations of the reprogramming cocktails alongside *Sox2* (Buganim et al., 2014; Feng et al., 2009; Gao et al., 2013; Han et al., 2010).

Sex-determining region Y (SRY)-box 2, SOX2 is part of the SRY-related family of transcription factors. SRY-related transcription factors contain a conserved high mobility group domain which mediates binding to DNA. SOX2, like OCT4, is important in early

preimplantation development. Although *Sox2* has a similar expression pattern to *Oct4* until the second cell fate decision, it has its own unique role in development (Frum et al., 2013; Wicklow et al., 2014). During early embryo development, SOX2 is one of the earliest markers of the inner cell mass. SOX2 works in conjunction with NANOG and OCT4 to promote pluripotency but works alone to promote FGF4 expression and pluripotency maintenance (Wicklow et al., 2014). Later in development SOX2 plays important roles in the formation of the extraembryonic ectoderm, nervous system, foregut and brachial arches (Rizzino & Wuebben, 2016). Like OCT4, SOX2 also plays an important role in pluripotency maintenance and self-renewal alongside NANOG (S. Liu et al., 2015; Patra, 2020).

Unlike OCT4 and SOX2, KLF4 is not a member of the three core pluripotency transcription factors (OCT4, SOX2 and NANOG) (Orkin et al., 2008; Patra, 2020; Shanak & Helms, 2020). However, KLF4 has been shown to prevent ESC differentiation and support the core pluripotency factors by promoting NANOG and SOX2 (Macarthur et al., 2012; Niwa et al., 2009; Orkin et al., 2008; P. Zhang et al., 2010). KLF4, Krüppel-like factor 4, is part of the SP/KLF family of transcription factors. KLF4 is a zinc finger transcription factor that plays a dual role as a transcriptional activator and repressor depending on the target (Ghaleb & Yang, 2017; Manini, 2008). KLF4 was first identified as a factor associated with growth arrest but has since been shown to more broadly play a role in regulating proliferation, apoptosis and homeostasis along with pluripotency promotion (X. Zhang et al., 2016).

Like KLF4, c-MYC is not a member of the three core pluripotency transcription factors (Orkin et al., 2008; Patra, 2020; Shanak & Helms, 2020). However, c-MYC plays an important role in promoting the larger pluripotency network (Huangfu, Osafune, et al., 2008; Orkin et al., 2008). c-MYC is encoded by the proto-oncogene family which belongs to the superfamily of basic helix-loop-helix DNA-binding proteins. Once partnered with Myc-associated protein X (Max), c-MYC can bind to the E-box region of target genes allowing for transcription to occur (Yoshida, 2018). c-MYC broadly plays important roles in cell growth, metabolism, proliferation, and down regulation of differentiation. Although it is mainly thought of as a proto-oncogene, it does play a role in embryo development and stem cell self-renewal (X. Zhang et al., 2016). c-MYC deletions are embryonic lethal between E9.5-10.5. c-MYC removal in pluripotent cells leads to differentiation (Yoshida, 2018). Although c-MYC does play a role in pluripotency and embryo development, it is a proto-oncogene. This creates concern for its use in iPSC formation that is aimed at patients therapy development (Maekawa et al., 2011). Fortunately, *OSKM* reprogramming can occur without *c-Myc* but reprogramming occurs at a reduced efficiency (Nakagawa et al., 2008).

### ***What is known about the mechanism of OSKM somatic cell reprogramming***

Much is known about how OCT4, SOX2, KLF4 and c-MYC operate in normal physiology, development, and disease formation. But how these four factors work together to reverse differentiation remains an active area of research. It was initially thought that *OSKM* operated quickly in a few cells to induce pluripotency and that the remaining weeks of reprogramming were to allow iPSCs to grow into a useable number of cells. However, the

many studies exploring chromatin state and gene transcription show that *OSKM* reprogramming is not that simple. Instead, *OSKM* reprogramming is a two-step process marked by the downregulation of somatic genes followed by the upregulation of pluripotency genes. This complex process is coordinated by transcriptome regulation and nucleosome remodeling (Huang et al., 2015; Knaupp et al., 2017; Soufi et al., 2012).

Within the first couple days of reprogramming, somatic genes are down regulated. Chromatin immunoprecipitation sequencing (ChIP-seq) and assay for transposase-accessible chromatin sequencing (ATAC-seq) data suggest that OSK alone, and in combination, can bind to active somatic gene enhancers or co-bind to prebound somatic transcription factors. This results in disrupted somatic gene transcription (Chronis et al., 2017; Soufi et al., 2012). Once somatic gene transcription is downregulated, chromatin must be closed at non-pluripotent gene sites. Evidence suggests that OSK can interact with histone deacetylases to close chromatin at somatic genes (D. Li et al., 2017). During this time, c-MYC binds to genes involved in apoptosis and senescence. Early in reprogramming, c-MYC has also been found to bind to genes not associated with either somatic gene repression or pluripotent gene activation and is suspected to play a large role in unwanted ectopic gene expression. Although c-MYC displays unproductive effects early in reprogramming, these deleterious effects are compensated for by the ability of c-MYC to increase the binding efficiency of OSK later in reprogramming through the induction of constitutively active promoters and indirect modification of epigenetically switched promoters (Banito et al., 2009; Soufi et al., 2012; Zviran et al., 2021).

Once somatic genes are repressed, OSKM must activate pluripotency genes in closed chromatin. It is believed that OSK works alone and in combination to access closed chromatin and bind to their target nucleosomes (J. Chen et al., 2016; Chronis et al., 2017; Knaupp et al., 2017; D. Li et al., 2017; Malik et al., 2019; Soufi et al., 2012; Zviran et al., 2021). c-MYC alone cannot bind to closed chromatin but can bind with OSK in combination to enhance binding efficiency (Soufi et al., 2012). OSKM can recognize a diverse range of DNA-binding domains. OCT4 can identify partial motifs which increases the number of DNA sites that it can engage. SOX2 has a high affinity for nucleosome binding given the unique bended confirmation of closed DNA (Soufi et al., 2015). KLF4 has a lower affinity for binding closed chromatin but appears to non-specifically bind nucleosomes as a searching mechanism (Soufi et al., 2015). This searching mechanism explains the observations that early in reprogramming, binding of OS to DNA is often facilitated through KLF4 binding (Knaupp et al., 2017). OSK binding to closed chromatin is observed as soon as 48 hours after the initiation of reprogramming, but the activation of core pluripotency genes like *Nanog* is not observed until later in reprogramming. These specific pluripotency genes appear to be mainly activated through OS enhancer binding and not through KM (Chronis et al., 2017; Malik et al., 2019).

Together, OSKM appear to be working in a stepwise manner to downregulate somatic genes and turn on pluripotency genes. However, because OSK can access closed chromatin, there is a wide array of ectopic gene expression that occurs before a final pluripotency state is achieved (Cacchiarelli et al., 2015; J. Chen et al., 2016; Raab et al., 2017; Soufi et al., 2012; Takahashi et al., 2014). This would suggest that reprogramming

is not as simple as a reversal of development and that instead OSKM acts in a unique non-physiological manner to help cells attain a pluripotent fate (Raab et al., 2017). Similar to chromatin states, gene and protein expression also proceeds through phases which include initiation, maturation and stabilization (González & Huangfu, 2016; Hansson et al., 2012; Schwarz et al., 2018).

During the initiation phase of reprogramming on days 0-6, there is an upregulation in genes important in mesenchymal to epithelial transitions, proliferation and stress response genes (González & Huangfu, 2016; Hansson et al., 2012; Mikkelsen et al., 2008). It is suspected that the upregulation in stress response genes acts as a fail-safe mechanism to prevent uncontrolled proliferation induced by c-MYC (Mikkelsen et al., 2008). During the initiation phase, there are global transcriptional changes that occur which affect mRNA, miRNA and lncRNA numbers (González & Huangfu, 2016).

After initiation, follows the maturation phase of gene expression which typically begins between days 9-12. This involves the transient expression of genes associated with different cell fates and developmental processes (Cacchiarelli et al., 2015; J. Chen et al., 2016; González & Huangfu, 2016; Takahashi et al., 2014). During this phase of reprogramming, these cells are in an intermediate phase in which they uniquely express markers not associated with pluripotency but instead with other cell types. During this phase, there is an emergence of several cell populations (González & Huangfu, 2016; Schwarz et al., 2018; Xing et al., 2020). These populations express genes broadly relating to endoderm, mesoderm, neuroectoderm and primitive streak formation (Cacchiarelli et

al., 2015; Mikkelsen et al., 2008; Takahashi et al., 2014). It has been suggested that during this phase, cells are very plastic and may be amenable to cell fate manipulation. At this point, only a few cells appear to be headed toward true pluripotency formation which can explain the low iPSC yield in reprogramming. The remaining cells not fated for pluripotency may remain in this intermediate state or mature on to different cells types (Meissner et al., 2007). This could explain the continued expression of endoderm genes and primitive streak genes present at the end of reprogramming (Cacchiarelli et al., 2015; Takahashi et al., 2014). In addition, during this phase, there is an increase in *Sox2* and *Oct4* expression which suggests that during this phase, cells are able to start producing their own endogenous SOX2 and OCT4 and are becoming less dependent on exogenous OSKM (J. Chen et al., 2016; González & Huangfu, 2016).

For those cells that have started to express pluripotency genes during the maturation phase, they can continue to the stabilization phase. During this phase, cells will continue to settle into their new pluripotent state and express any remaining pluripotency markers, extend telomers and reactivate X-chromosomes (González & Huangfu, 2016).

Although a lot is known about the different chromatin states and genes expressed during reprogramming, it is evident that each reprogramming cell reacts differently to OSKM. This has made it challenging to map out the exact mechanism necessary for cells to achieve pluripotency and not follow a failed trajectory. A reprogramming cell's pathway also appears to be influenced by the starting cell types (Polo et al., 2010). For example, blood cells and neuronal stem cells activate endogenous OCT4 or NANOG more quickly

than fibroblasts, respectively (Apostolou & Stadtfeld, 2018). In addition, the ability of OSKM to induce an intermediate state that is associated with the creation of a diverse population of cell types provided the first glimpse into the possibility that OSKM may not be specific at inducing pluripotency. Further investigation into the possibility that OSKM creates non-pluripotent cells revealed that indeed, induced multipotent embryonic-like stem cells are created during *OSKM* reprogramming.

For example, the first discovery of a stable non-pluripotent embryonic-like stem cell occurred with the discover of induced extraembryonic endoderm (iXEN) in 2016 (Parenti et al., 2016) followed by the discovery of induced trophoblast stem cells (iTSCs) in 2020 (Castel et al., 2020; X. Liu et al., 2020) (Fig. 1.2. B). The discovery of stable induced non-pluripotent stem cells has presented a problem with uncovering the mechanism of successful pluripotency. Most previous work uses techniques that look at gene expression or chromatin state over a population and not with in individual cells. This population approach has ignored the possibility that many of these non-pluripotency genes could be important in establishing alternative fates and are not just noise.

### **Section 1.3. The discovery and promise of iXEN**

#### ***Discovery of iXEN***

*OSKM* cell reprogramming produces iPSCs at a low efficiency. The cells that do not go on to complete successful reprogramming are thought to be stuck in an intermediate state of reprogramming. Researchers have tried to evaluate these intermediate cell states by hypothesizing that these cells contain unique pluripotency states that are different than

an typical iPSC (S. Guo et al., 2015; Tonge et al., 2014). However, these studies fail to address how or why an array of unique gene signatures associated with known cell types exist within these intermediate cell states of reprogramming. Interestingly, many of the genes expressed are transient, except for endodermal genes which continue to be expressed to the end of reprogramming (Cacchiarelli et al., 2015; Maekawa et al., 2011; Takahashi et al., 2014). This continued expression of endodermal genes suggests that there is a unique gene expression profile being activated during *OSKM* reprogramming that is worth understanding. In addition, chemical reprogramming, which uses small molecules to turn on pathways that activate *OSKM* (Fig 1.3. A), also reported the expression of a transient state that represented extraembryonic endoderm (Y. Zhao et al., 2015). Lastly, *Fbx15*, which was used as a readout for pluripotency in the discovery of iPSCs (Takahashi 2006, Tokazawa 2003), also has a similar expression level in XEN cells as it does in ESCs (data from Rugg-Gunn et al., 2010) (Fig. 1.4 A) indicating that *Fbx15* is not a specific marker for pluripotency formation. By taking a closer look at colonies that failed to express typical iPSC morphology in *OSKM* reprogramming, it was discovered that some of these colonies were indeed iXEN. These iXEN cells had a similar morphology, expression pattern and differentiation potential as embryo-derived XEN cells (Parenti et al., 2016). iXEN were even capable of contributing to the parietal endoderm in the developing embryo and forming visceral endoderm *ex vivo*. These results have been confirmed through the establishment of stable, chemically induced iXEN cells and the establishment of iXEN in other species (He et al., 2020; Nishimura et al., 2017).

Currently, it is still not understood how OSKM reprogramming creates two distinct embryonic-like stem cells in parallel. However, given the current knowledge on early embryo development, it can be speculated as to how OSKM may facilitate iXEN formation. Functionally, iPSCs are similar to the epiblast and iXEN cells are similar to the primitive endoderm in the embryo. In the embryo, *Oct4* is expressed in, and required for, both epiblast and primitive endoderm cell development (Aksoy et al., 2013; Frum et al., 2013; Le Bin et al., 2014; Niakan et al., 2010; Wicklow et al., 2014). It is not until E4.5 that OCT4 becomes a strict pluripotency factor and is no longer expressed in the primitive endoderm lineage. It has also been shown through the study of *Sox2* null embryos, that SOX2 indirectly promotes primitive endoderm formation (Wicklow et al., 2014). KLF4 has also been shown to be expressed in the primitive endoderm and may play a role in primitive endoderm gene expression (Morgani & Brickman, 2015). Lastly, c-MYC has been shown to regulate primitive endoderm gene expression in pluripotent and somatic cells (Neri et al., 2012; Smith et al., 2010). Together, this evidence suggests that OSKM may play a direct role in the formation of iXEN. The discovery of iXEN alongside iPSC provides a unique opportunity to better understand the role of OSKM in cell fate decisions and how cells must choose one embryonic cell fate over another.

### ***Significance of embryonic and induced XEN***

XEN and iXEN are a type of embryonic stem cell that are *in vitro* models of embryonic primitive endoderm. In humans and mice, primitive endoderm cells are in contact with epiblast cells (Stern & Downs, 2012). This arrangement is important as primitive endoderm transmits key signals to the epiblast that allow for the formation of the anterior-

posterior axis, blood islands and the anterior neural plate in mice (Belaousoff et al., 1998; Stuckey et al., 2011; Thomas & Beddington, 1996). Given primitive endoderm's significance to the developing embryo, having a stem cell model of primitive endoderm allows for the in-depth study and manipulations of genes and the extracellular environment. In addition, researchers have started to incorporate XEN cells into experimental models. It has been shown that signals secreted by XEN cells can increase cardiomyocyte formation *in vitro* (Brown et al., 2010).

Despite the significance of XEN in development and use in research models, to date there have been no reported establishment of human XEN cell lines derived from human embryos (Rossant, 2014) or reprogramming. Instead, all insights into the fundamental processes of primitive endoderm/ XEN cells have been studied using the mouse embryo and mouse embryo-derived XEN cell lines (Moerkamp et al., 2013). Although the mouse model provides unique advantages which include the creation of null alleles, fluorescently tagged proteins, lineage tracing tools and the creation of chimeras and teratomas to evaluate stem cell quality, the significance of primitive endoderm/ XEN cells in human development can only be speculated. The establishment of human iXEN cells will enable future studies that explore environmental or genetic influences on human XEN and how this might impact development (Linneberg-Agerholm et al., 2019). The discovery that OSKM induces iXEN formation provides promise that OSKM can induce iXEN in human cells.

It has been well established that the primitive endoderm plays an important role in signaling, waste and nutrient exchange within the developing embryo. However, mouse studies have suggested that the primitive endoderm can functionally contribute to definitive endoderm lineages by incorporating into the gut tube (Kwon et al., 2008; Nowotschin et al., 2019). This demonstrates that primitive endoderm cells are plastic and not strictly extraembryonic. Thus, XEN may also be used to engineer human tissues of endodermal origin. In addition, it has been shown that canine iXEN cells can be easily differentiated into hepatocyte-like cells, which demonstrates the medical relevance of iXEN cells (Nishimura et al., 2017).

### ***iXEN and iPSCs as a key to understanding OSKM somatic cell reprogramming***

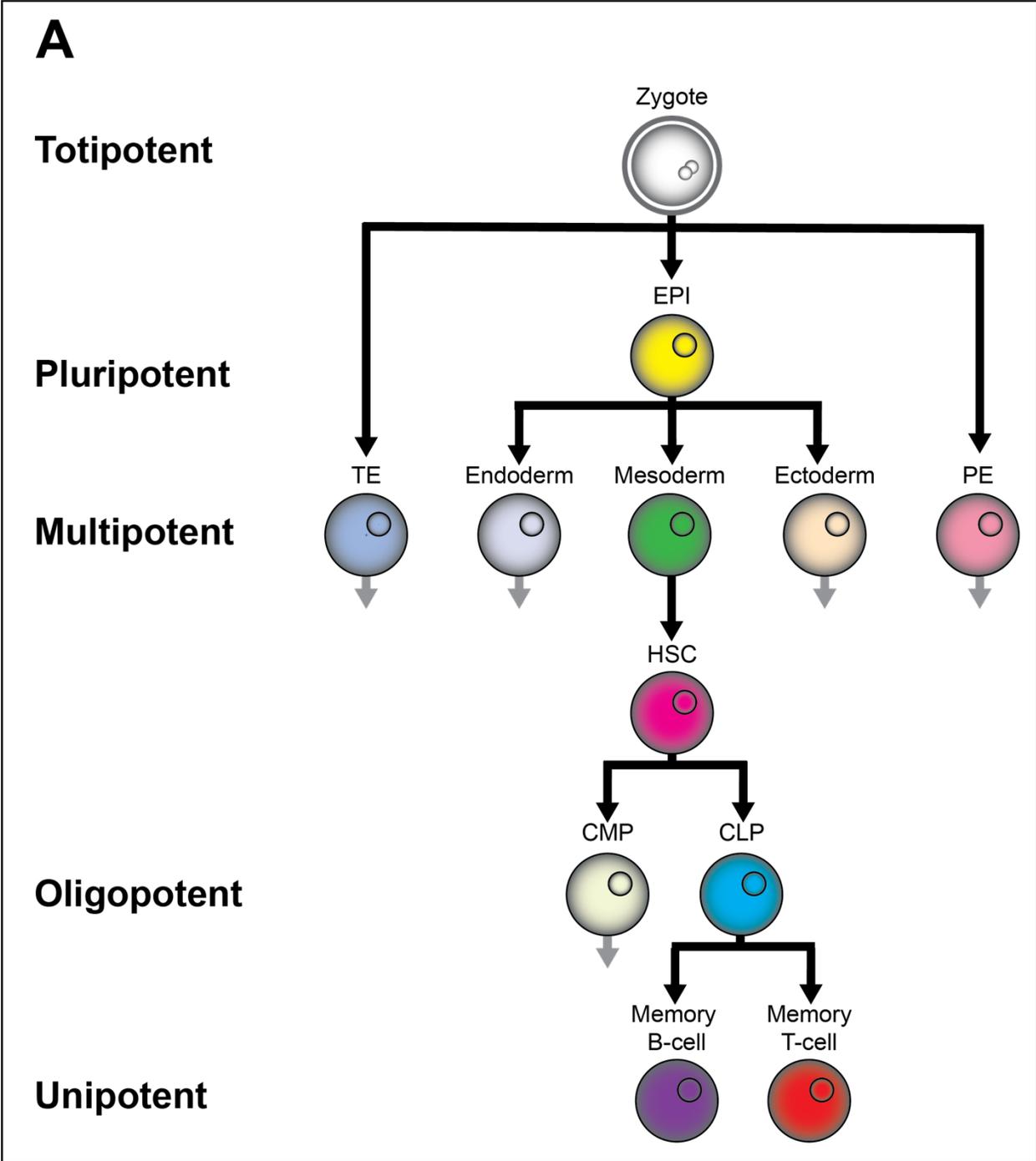
The field of reprogramming has completed remarkable studies that have addressed how OSKM broadly functions to establish a pluripotent state. Interestingly, there is no detailed consensus on how OSKM specifically functions to repress somatic genes and activate only pluripotent genes. This has limited the field's ability to modify the reprogramming process to produce more elite iPSCs at a higher rate. This lack of mechanistic detail was thought to be impaired by the off-target effects generated from OSKM and the heterogeneity that exists within cells. But the recent discoveries of iXEN cells and iTSCs provides an alternative explanation in which OSKM is not a specific inducer of pluripotency but early embryo development. Instead of OSKM generating noise, it may actually be generating alternative fates that are worth pursuing.

Our lab believes that evaluating the formation of iXEN cells could hold the key to unlocking the true mechanism of OSKM. In particular, our lab focuses on iPSC and iXEN formation due to the close relationship of their embryonic counterparts, epiblast and primitive endoderm. In addition, there is a large body of knowledge generated from embryo development studies to help guide our decisions into exploring how OSKM may produce iXEN and iPSCs in parallel. For example, there are many interesting transcription factors that play a role in establishing the progenitor state to epiblast and primitive endoderm or play a dual role in epiblast and primitive endoderm formation. One of the most interesting targets is OCT4 given that it is important in establishing the progenitor state and both epiblast and primitive endoderm. The exploration of these different transcription factors may elucidate how alternative cell states are formed during reprogramming.

In order to begin to tackle how OSKM forms two induced stem cell types in parallel, it is important to identify reliable markers of early iPSC and iXEN formation and begin to evaluate subtle differences in gene expression to get a baseline understanding of what is occurring. From there it will be important to modify the cell environment and reprogramming cocktail to see how this influences one cell fate decision over another. Lastly, given the heterogeneity of reprogramming cells, it is essential to apply single cell methods to better understand subtle differences occurring within cell types.

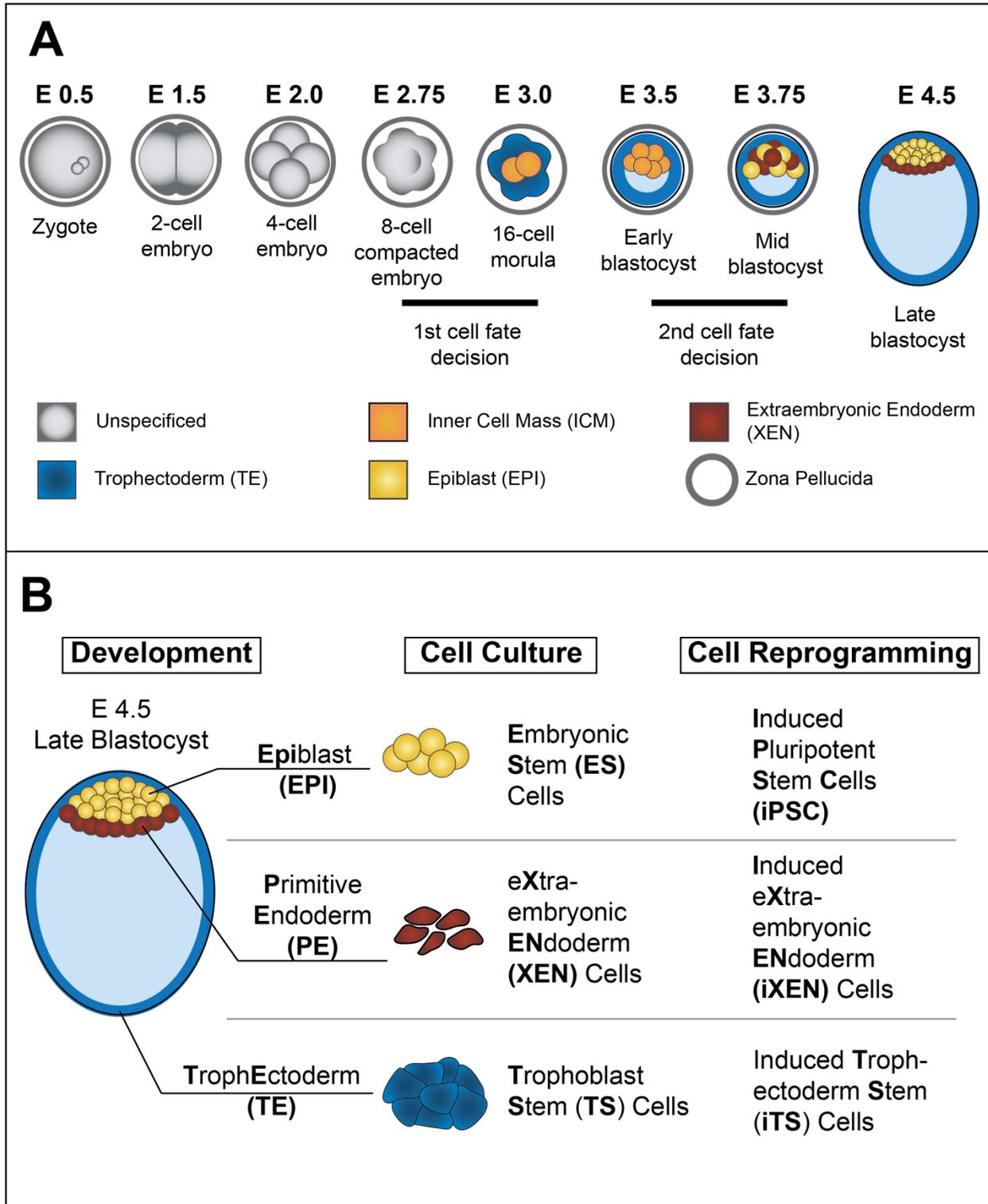
## APPENDIX

Figure 1.1. Stem cell potency.



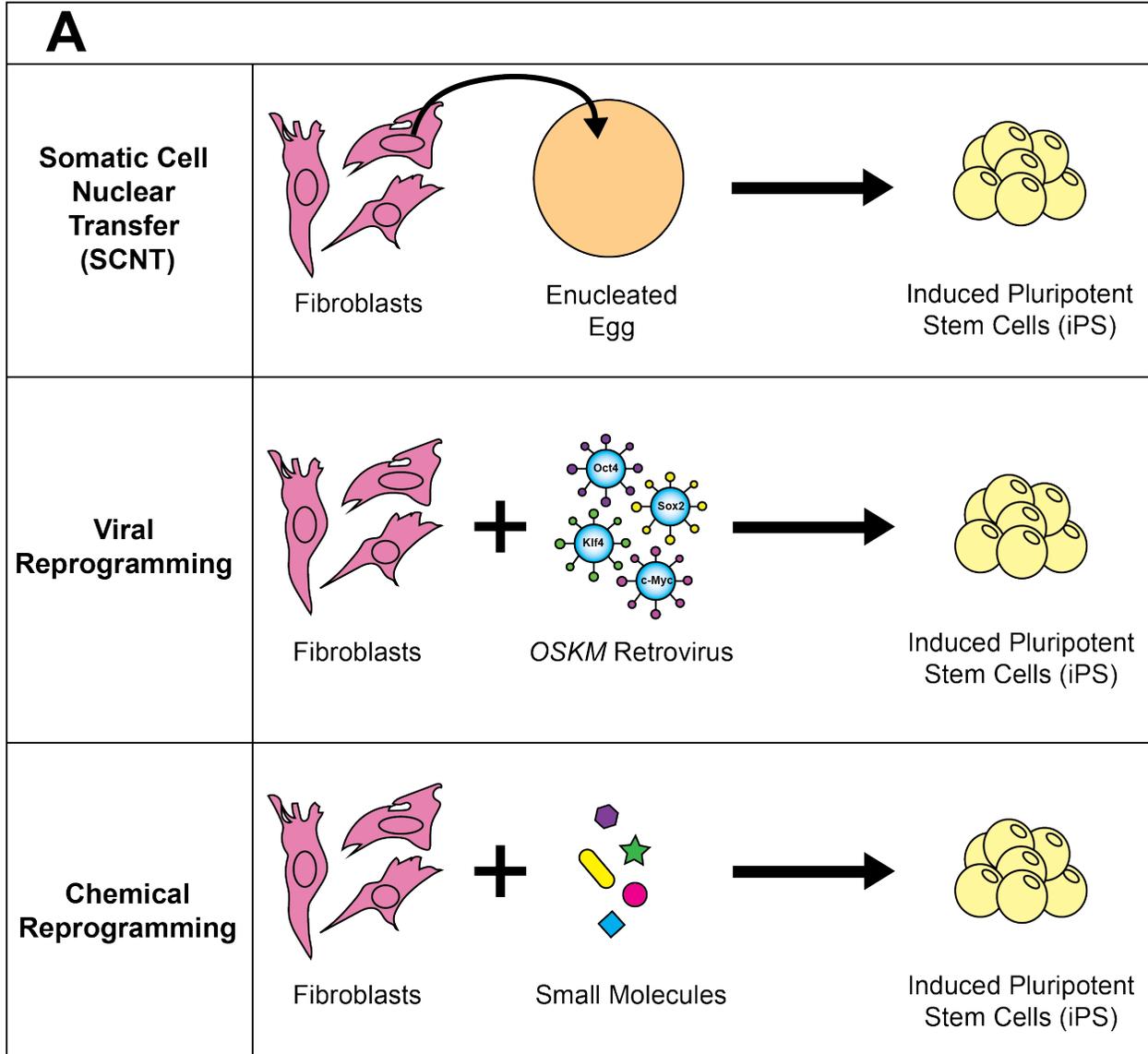
**Figure 1.1. (cont'd).** Stem cell potency is defined by the ability of a cell to differentiate into other lineages. The highest potency is a totipotent cell which can give rise to all embryonic and extraembryonic cell types. As a totipotent cell differentiates it can become multipotent or pluripotent. As a stem cell continues to differentiate, the cell becomes further restricted in its opportunities to take on different cell fates and eventually becomes limited oligopotency and unipotency. The figure highlights one track of differentiation from totipotency to unipotency in which a zygote differentiates into memory B-cell and T-cells. EPI = Epiblast, TE = Trophectoderm, PE = Primitive Endoderm, HSC = Hematopoietic Stem Cell, CMP = Common Myeloid Progenitor, CLP = Common Lymphoid Progenitor.

Figure 1.2. Embryo development and the formation of embryonic stem cells.



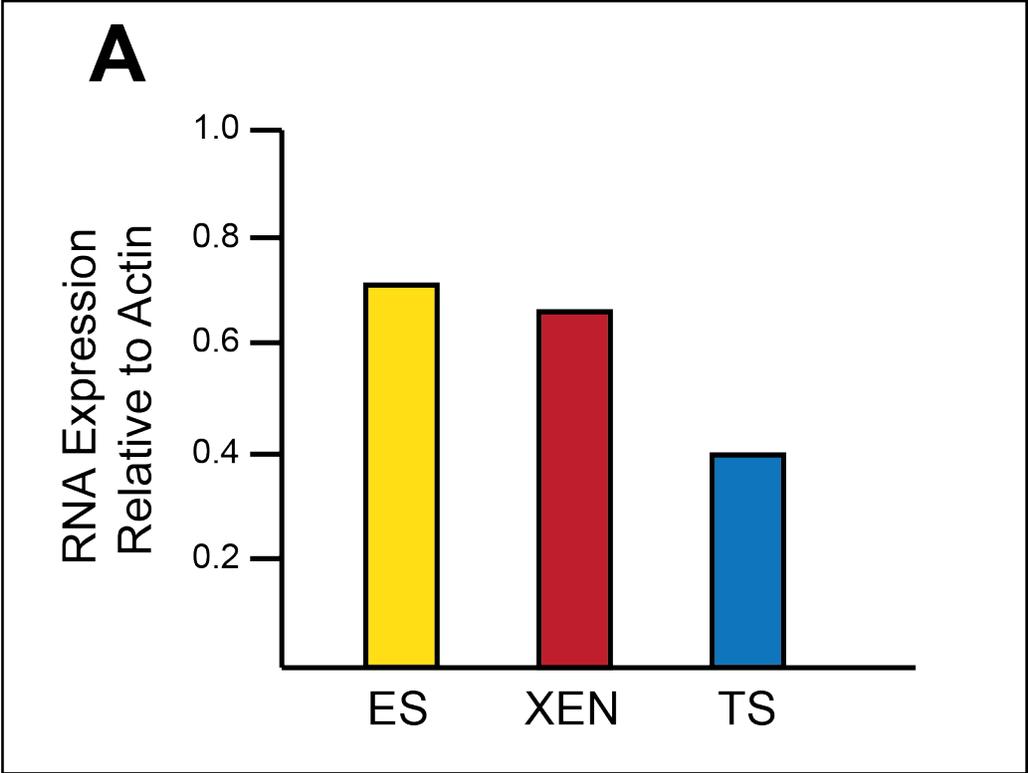
**Figure 1.2. (cont'd).** Early embryo development provides a unique model for understanding the formation of stem cells and studying stem cells as they journey from a totipotent cell (zygote) into pluripotent (EPI) and multipotent cells (PE and TE). A) Early embryo development starting with the zygote until late blastocyst formation. B) Embryonic stem cells can be harvested from the embryo to represent the different cell lineages in the late blastocyst. These embryonic-derived stem cells also have induced stem cell counterparts that are produced in cell reprogramming.

**Figure 1.3. Somatic cell reprogramming.**



The formation of somatic cell reprogramming began with somatic nuclear transfer in which the nucleus of a somatic cell was placed into an enucleated egg to reprogram the somatic nucleus into a pluripotent state. Following somatic cell nuclear transfer was the discovery that exogenous *OSKM* could induce pluripotent cell formation. *OSKM* induction can occur via retroviral delivery or by growing somatic cells in a combination of small molecules to induce the expression of endogenous *OSKM*.

Figure 1.4. *Fbx15* gene expression in embryonic derived stem cell lines.



*Fbx15* was initially used as a marker of pluripotency during the development of OSKM reprogramming. Microarray data, from Rugg-Gunn et al., 2010., shows that *Fbx15* is expressed in embryonic stem (ES) cells, extraembryonic endoderm (XEN) cells and trophoblast stem (TS) cells showing that it is not a specific marker of pluripotency.

## CHAPTER 2.

Utilization of molecular and cellular techniques to distinguishing between  
extraembryonic endoderm and pluripotent stem cells during OSKM somatic cell  
reprogramming

Authors: Moauro A<sup>1</sup> and Ralston A<sup>2\*</sup>

- 1) Graduate Program in Physiology and Osteopathic Medicine, Michigan State University, East Lansing, MI, 48823
- 2) Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48823

Published as:

Moauro A<sup>1</sup> and Ralston A\*. (2022). Distinguishing between iXEN and iPSC in OSKM Somatic Cell Reprogramming. *Methods in Molecular Biology*. Springer.

A. Moauro wrote the manuscript and assembled the figures.

All authors edited the manuscript.

The previously published work has been modified to include the following:

Sections - 2.9, 2.10, 3.8 and 3.9

Figures – 2.3 and 2.4

## **Abstract**

Mouse somatic cell reprogramming using *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) induces formation of two stem cell types: induced pluripotent stem cells (iPSC) and induced extraembryonic endoderm stem cells (iXEN). Since both stem cell types routinely arise alongside one another during reprogramming, it is critical to distinguish between both cell types to ensure that the desired cell population is selected and analyzed. This chapter details, from start to finish, how to reprogram mouse embryonic fibroblasts (MEFs) using retrovirus and how to distinguish between iXEN and iPSC at the colony and single cell levels.

## Introduction

Somatic cell reprogramming using transcription factors *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) has long been recognized to produce induced pluripotent stem cells (iPSCs) (Takahashi & Yamanaka, 2006). Like embryonic stem cells (ESC), iPSCs are pluripotent and are capable of forming all three germ layers as well as the germ line. iPSCs provide therapeutic promise for regenerative medicine and enable novel research models related to personalized medicine.

In addition to iPSCs, a distinct stem cell type has more recently been discovered: induced extraembryonic endoderm stem cells (iXEN), which routinely arise alongside iPSC during OSKM reprogramming in mouse somatic cells (Nishimura et al., 2017; Parenti et al., 2016). iXEN are similar to the extraembryonic endoderm (XEN) of the embryo. In mouse, the XEN lineage is important in forming extraembryonic structures such as the yolk sac endoderm, which take part in nutrient exchange. These cells play a signaling role in establishing the anterior-posterior axis, blood islands and anterior neural plate in mice (Belaoussoff et al., 1998; Stuckey et al., 2011; Thomas & Beddington, 1996). Lastly, in mice XEN has been shown to contribute to the organs of the definitive endoderm (Kwon et al., 2008) and iXEN cells have been induced to form hepatocyte-like cells in dogs (Nishimura et al., 2017), highlighting the similarities between extraembryonic and definitive endoderm. Like iPSC, iXEN also has therapeutic promise and provides a novel research alternative to the use of embryos for studying XEN.

Since both iPSC and iXEN routinely arise in OSKM reprogramming, it is important to distinguish between the cell types. Here, we describe how to reprogram cells and identify putative iPSC and iXEN as they arise. The first step in distinguishing between the two cell types is to sort out putative iPSC and iXEN by colony picking or single cell sorting. Once putative cell types have been collected, additional analysis is recommended to confirm the cellular phenotype. Confirmatory analysis should be completed by evaluating transcript and protein levels of specific markers related to iPSC and iXEN fates. For example, Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR) and RNA-seq can be used for transcriptional read outs. Immunofluorescent imaging and flow cytometry analysis can be completed for protein expression read outs.

The protocols below provide details, from start to finish, on how to collect starting somatic cells (MEFs), how to reprogram cells using virus, and provides recommendations for distinguishing between iXEN and iPSCs at both colony and single-cell levels.

## **Materials**

### ***Media Preparation***

1. 293T Cell Medium: DMEM, 15% Fetal bovine serum (FBS), 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 0.1 mM Beta-mercaptoethanol
2. ESC Medium: DMEM, 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 1 mM Sodium pyruvate, 100 U/mL Penicillin/streptomycin, 15% Knock out serum replacement, 10 ng/mL Leukemia Inhibitory Factor (LIF)

3. MEF Medium: DMEM, 2 mM Glutamax, 100 U/mL Penicillin/streptomycin, 10% Fetal bovine serum (FBS)
4. Reprogramming Medium 1: DMEM, 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Fetal bovine serum FBS, 10 ng/mL Leukemia Inhibitory Factor (LIF)
5. Reprogramming Medium 2: DMEM, 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Knock out serum replacement (KOSR), 10 ng/mL Leukemia Inhibitory Factor (LIF)
6. XEN Medium: DMEM, 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 1 mM Sodium pyruvate, 100 U/mL Penicillin/streptomycin, 15% FBS

### ***Preparing Mouse Embryonic Fibroblasts (MEFs)***

1. 0.25% Trypsin-EDTA
2. MEF medium (see above)
3. PBS
4. 6-well flat bottom tissue culture treated polystyrene plates
5. Sterile 4.5" dissecting scissors
6. Sterile fine point forceps
7. Dissecting microscope

## ***Preparing Replication-incompetent Retroviruses for Overexpression of OSKM for Reprogramming***

1. HEK 293T cells (below passage 12)
2. 293T medium
3. MEF medium without antibiotics
4. Lipofectamine 2000
5. Opti-Mem
6. 16 µg pMXs plasmid with gene of interest (the original plasmids used to establish OSKM reprogramming (Takahashi & Yamanaka, 2006) can be found below)
  - a. pMXs-Oct3/4 (Addgene Plasmid #13366)
  - b. pMXs-Sox2 (Addgene Plasmid #13367)
  - c. pMXs-Klf4 (Addgene Plasmid #13370)
  - d. pMXs-c-Myc (Addgene Plasmid #13375)
7. 8 µg pCL-Eco packaging plasmid (Addgene Plasmid #12371)
8. 100 mm tissue culture treated polystyrene plate
9. 0.45 µm filter

## ***Viral Titer of Retrovirus***

1. MEFs
2. MEF medium
3. Replication-incompetent retrovirus
4. 0.25% Trypsin-EDTA
5. Polybrene

6. TRIzol
7. Chloroform
8. Reverse transcription kit
9. qPCR machine
10. Primers specific to viral genome (sequences provided in Table 2.1)
11. qPCR fluorescent detector such as Sybr Green

### ***OSKM Viral Reprogramming***

1. MEFs
2. MEF medium
3. Reprogramming medium 1
4. Reprogramming medium 2
5. Replication-incompetent retrovirus
6. Polybrene
7. Tissue culture treated polystyrene dishes

### ***Picking and Passaging iXEN and iPSC Colonies***

1. Reprogrammed cells
2. DPBS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$
3. 0.25% Trypsin-EDTA
4. Reprogramming medium 1
5. Stereomicroscope

6. Fine point forceps
7. 96 well tissue culture treated polystyrene
8. Tissue culture treated polystyrene plates

### ***Confocal Imaging of Reprogramming Cells***

1. Reprogrammed cells grown on confocal grade tissue culture treated plastic 4-wells, for example ibidi #80426
2. 0.1% Gelatin
3. 4% Formaldehyde in PBS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$
4. Pre-Block Solution (0.5% Triton X-100 in PBS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ )
5. Blocking Solution (0.1% Triton X-100 and 10% FBS in PBS, PBS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ )
6. Primary and secondary antibodies with conjugated fluorophores (see Table 2.2)
7. 0.001 mg/mL DAPI (final concentration)
8. Confocal Microscope

### ***Cell Sorting of Reprogramming Cells***

1. MEF cells carrying a fluorescent reporter allele of interest
2. 0.2  $\mu\text{m}$  filtered Cell Sorting Buffer (PBS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ , 1mM EDTA, 25 mM HEPES pH 7.0, 2% FBS)
3. Irradiated MEFs (E13.5 MEFs exposed to 6000 rads of  $\gamma$ -irradiation) (Andras Nagy, Marina Gertsenstein, 2006)
4. 0.25% Trypsin-EDTA

5. Reprogramming medium 1 or XEN medium
6. 96 well tissue culture treated polystyrene
7. 40 µm filter
8. 0.001 mg/mL DAPI (final concentration)
9. Flow cytometer capable of sorter

### ***RNA-seq of Passaged Cell Lines***

1. Cells for analysis
2. 70% EtOH
3. RNase removal spray
4. Trizol
5. Chloroform
6. Centrifuge
7. Isopropanol
8. 75% EtOH
9. RNase free H<sub>2</sub>O
10. QIAGEN RNeasy kit
11. Qubit
12. Qubit broad range detection kit
13. Qubit high sensitivity detection kit
14. Invitrogen TURBO DNA-free kit
15. TapeStation/ Bioanalyzer
16. Sequencer capable of RNA-seq

## ***scRNA-seq of Reprogramming Cells***

1. Reprogramming sample
2. 6 well tissue culture treated polystyrene
3. DPBS without Ca<sup>2+</sup> and Mg<sup>2+</sup>
4. 1mM EDTA diluted in DPBS
5. 0.25% Trypsin-EDTA
6. Reprogramming medium 1
7. scRNA-seq Cell Buffer (25mM HEPES pH 7.0, 2% FBS in DMEM)
8. 40 mm filter
9. Stereomicroscope
10. Hemacytometer
11. Trypan Blue Stain
12. Ice
13. Sequencer capable of scRNA-seq

## **Methods**

### ***Preparing Mouse Embryonic Fibroblasts (MEFs)***

Reprogramming cells using OSKM can be completed using several cell types; however, many researchers choose to use MEFs. MEFs are a great choice because the cells are easy to collect, easy to work with and provide a good yield of cells. In addition, many experiments require a specific cell genotype such as knock ins, knock outs and fluorescent reporters. It is easy to make MEFs in house by breeding mice to produce the desired genotype. Our lab often uses fluorescent reporters such as GFP tagged to a gene

interest. For example, *Nanog-GFP<sup>+/-</sup>* (Maherali et al., 2007) helps in visualizing putative iPSCs.

1. Sacrifice a pregnant female mouse 13 days after the morning plug was observed (E13.5).
2. Swab the abdomen of the mouse with 70% ethanol. Make a large incision in the shape of a U on the mouse's abdomen. Grab the body wall with fine forceps and make an incision with scissors. Remove both horns of the uterus by cutting at the oviduct and cervix, and then place the uterus in a dish with PBS.
3. Cut open the uterus longitudinally to expose all of the embryos. Remove all extraembryonic tissue (yolk sac and placenta) and transfer each embryo to a dish with fresh PBS. Embryos can be placed in individual wells of a 6-well dish or in a 10 cm plate with embryos arranged in a clock pattern around the dish.
4. Decapitate the embryo and remove all the internal organs of each embryo with forceps leaving behind only skin, muscle, and bone (See Note 1). Place the desired embryo tissue in a new well of a 6-well dish with 1mL PBS (See Note 2).
5. Remove the PBS and add 0.5 mL Trypsin-EDTA per embryo. Mince the embryos with sterile scissors. Once minced as much as possible, incubate the embryo-trypsin mixture at 37 °C for 5 min.
6. Mince the mixture again and use a P1000 pipettor to break up the tissue further by gently pipetting up and down (See Note 3).
7. Once the tissue is disaggregated, add 2 mL MEF medium to each well to quench the Trypsin digestion.
8. Add the embryo mixture and 8 mL MEFs medium to a 15 mL conical tube.

Centrifuge cells at 200 xg for 4 min. Carefully remove the medium and add 10 mL fresh MEF medium. Plate the cells onto a gelatin coated 10 cm dish and incubate at 37 °C in 5% CO<sub>2</sub>.

9. Replace the medium every day until cells are confluent. Once confluent, freeze down the cells or continue on for reprogramming. These cells are passage #1.

### ***Preparing Replication-incompetent Retroviruses for Overexpression of OSKM for Reprogramming***

In order to reprogram cells, many researchers use replication-incompetent retrovirus to deliver transcripts *Oct4*, *Sox2*, *Klf4* and *cMyc*. Making retrovirus is easy and produces a high yield. The protocol below details how to produce Moloney Murine Leukemia Virus (MMLV) derived retrovirus (See Note 4).

1. Plate 293T cells in 293T medium onto a 10 cm plate. Change medium every other day and split cells once they reach ~80-90% confluence. Passage the cells twice before using for transfection, splitting at a ratio of 1:6 each passage. Make sure cells are no higher than passage 12 on the day of transfection.
2. On the day of plasmid transfection, cells should be 70-80% confluent. For transfection, set up two tubes per virus and wait for 5 minutes for the contents of each tube to mix:
  - a. Tube A: 1.5 mL Opti-MEM (See Note 5) with 50 µL Lipofectamine.
  - b. Tube B: 1.5 mL Opti-MEM with 16 µg pMXs gene plasmid and 8 µg pCL-Eco packaging plasmid (See Note 6).
3. Next, mix the contents of the tubes A and B, and then incubate at room

temperature for 20 minutes. During this time, replace the medium on 293T cell plates with 12mL fresh MEF medium that contain no antibiotics.

4. Add the 3 mL transfection solution (tubes A + B) in a dropwise manor to the 293T cells. Place cells in the incubator at 37 °C with 5% CO<sub>2</sub>.
5. After 18-24 hours, replace the medium with 10mL fresh MEF medium.
6. After another 24-30 hours, collect conditioned medium and filter through a 0.45 µm filter or spin medium at 200 x g for 4 minutes to remove any cells. Divide virus into 1 mL aliquots and immediately store at -80 °C until the virus is ready to be use for reprogramming or titers.

### ***Viral Titer of Retrovirus***

Once retrovirus has been prepared, it is important to titer the virus to determine the quantity of virus produced. This added step ensures each reprogramming experiment uses the same quantity of virus allowing for consistent results independent of virus preparations.

1. Plate MEFs onto gelatin coated wells at a density of 50-100 cells/mm<sup>2</sup>. Prepare 4 wells of cells per virus to be titered.
2. Dilute virus of interest using DMEM to create 4 different virus concentrations (See Note 7) to be tested and add 4 mg/mL of Polybrene to each of the final virus solution. Add viral mixture to cells and incubate in the cell culture incubator at 37°C with 5% CO<sub>2</sub>.
3. After 48 hours, collect the infected MEFs and harvest the RNA using TRIzol and chloroform (See Note 8).

4. Reverse transcribe the RNA to create cDNA and quantify viral transcript levels by qPCR (primer sequences are provided in Table 1).
5. Standard curves can be created using serial dilutions of the retroviral plasmids. The concentration of virus can then be calculated using the standard curve.

### ***OSKM Viral Reprogramming***

Once virus and MEFs are prepared, viral reprogramming can proceed (See Note 9). The protocol below details how to infect cells with retrovirus and how cells should be maintained through the 23-day protocol (Figure 2.1A).

1. Day -1: plate MEFs (passage 3 or lower) at a density of 50-100 cells/mm<sup>2</sup> onto gelatin coated dishes. The lower the density, the less likely the cells will be overgrown by the end of reprogramming.
2. Day 0: mix together all OSKM viruses (1x10<sup>8</sup> transcripts/mL/cell) and 4 mg/mL of Polybrene. Using DMEM, bring the volume to the final well volume (e.g. 2 mL for one 6-well). Add viral mixture to cells and incubate for 24 hours in the cell culture incubator at 37 °C with 5% CO<sub>2</sub>.
3. Day 1: replace virus medium with MEF medium and incubate for another 24 hours in the incubator.
4. Day 2 and 4: replace medium with Reprogramming medium 1.
5. Day 6: replace medium with Reprogramming medium 2 and then every other day until the end of the experiment (~21 days) (See Note 10).

### ***Picking and Passaging iXEN and iPSC Colonies***

Once colonies have started to emerge during reprogramming, colonies can be picked and expanded. These colonies can then be analyzed for iXEN and iPSC markers and appropriate cell morphology.

1. Once colonies have formed (See Note 11) and are ready for picking, wash well with PBS and replace with fresh PBS.
2. Using a microscope and fine point forceps, trace the colony of interest (See Note 12) to separate it from the underlying fibroblasts.
3. Set a P20 pipettor to 5 mL, use it to lift the colony, and then transfer it onto a gelatin coated well of a 96-well dish containing 30 mL trypsin. Once colonies are picked, incubate the 96-well at 37 °C for 3 minutes.
4. Quench trypsin with 200 mL medium (medium should contain FBS) of choice. Gently pipet cell suspension up and down to break up the colony.
5. Replace medium every other day until cells are confluent. Once confluent, transfer cells to a 24-well. Continue to passage the cells for at least 11 passages (Parenti et al., 2016) to ensure cells have settled into their final cell fate.
6. After 11 passages, cell lines should be analyzed using various methods (qPCR, morphology, and imaging) to determine whether they are iPSC or iXEN.

### ***Fluorescently Activated Cell Sorting (FACS) of Reprogramming Cells***

FACS allows for single or bulk sorting using differential marker expression. Specifically, single cell sorting allows for individual cell analysis which is useful for cell fate determination or single cell RNA-seq. Single cell techniques can provide better resolution

and detect subtle differences between cells which are lost in bulk analysis. The below protocol details how to sort single cells and expanding those single cells to form cell lines. Before starting, we recommend having a MEF line that expresses a fluorescent reporter that distinguishes between iPSC and iXEN. For example, *Nanog-GFP<sup>+/-</sup>* (Maherali et al., 2007) MEF lines can be used to select putative iPSC.

1. Cells will be sorted into 96-wells plates. The day before the sort, plate  $8 \times 10^6$  cells/plate of irradiated fibroblasts onto 1-2 96-well plates (See Note 13).
2. On the day of the sort, replace the 96-well plate medium with 300  $\mu$ L Reprogramming medium 1. Wrap the plates in parafilm and place on ice.
3. Harvest confluent cells with trypsin (incubation at 37 °C for 4 min) and wash twice with PBS. These cells should contain a fluorescent marker read out such as GFP (See Note 14) for sorting.
4. Wash collected cells with Cell Sorting Buffer and resuspend to  $4-7 \times 10^6$  cells/mL (cell concentration may vary depending on the instrument). Filter samples using a 40  $\mu$ m filter right before sorting. Keep cells on ice until you are ready to run the sort.
5. Right before running the sort, add 1  $\mu$ g/mL of DAPI (sample can only sit in DAPI for 20 minutes). DAPI-positive cells indicate dead cells.
6. Proceed to sort living cells (See Note 15). For single cell resolution, we recommend sorting 1 living cell into a single 96-well. Once a plate is filled, replace parafilm and place the plate back on ice.
7. Once all cells are sorted, remove parafilm and place plates in the incubator at 37°C with 5% CO<sub>2</sub>.

8. After 6 hours, replace medium with 100  $\mu\text{L}$  of desired medium.
9. Replace medium every 2 days and passage cells once colonies have formed.

### ***Confocal Imaging of Reprogramming Cells***

Fluorescent imaging using confocal microscopy (See Note 16) allows for higher resolution imaging of multiple markers in combination with morphology while limiting background fluorescence and light scatter. Confocal microscopy provides spatial information on how different markers are expressed through out a colony and how individual colonies express markers in comparison to one another. Confocal imaging allows for the analysis of multiple colonies on a whole plate which retains dynamic spatial reprogramming information. In comparison, single cell sorting or colony picking only provides information about individual cells or individual colonies respectively.

1. Complete reprogramming on gelatin coated confocal grade tissue culture treated 4-well dishes.
2. At the desired time point, carefully aspirate cell culture medium and fix cells for 10 minutes at room temperature with 200  $\mu\text{L}$  4% formaldehyde.
3. Carefully aspirate 4% formaldehyde and add 280  $\mu\text{L}$  of Pre-blocking Solution and incubate for 30 minutes at room temperature.
4. Carefully aspirate pre-blocking solution. Add 280  $\mu\text{L}$  of Blocking Solution and incubate for 1 hour at room temperature.
5. Carefully aspirate blocking solution and add 280  $\mu\text{L}$  of diluted primary antibodies to wells. Seal and incubate at 4  $^{\circ}\text{C}$  overnight.

6. Carefully aspirate primary antibodies solution. Add 280  $\mu$ L of Blocking Solution and incubate for 30 minutes at room temperature.
7. Carefully aspirate the Blocking Solution and add diluted 280  $\mu$ L of secondary antibodies for 1 hour at room temperature in the dark.
8. Carefully aspirate secondary antibodies solution. Add 280  $\mu$ L of Blocking Solution and incubate for 30 minutes at room temperature in the dark.
9. Aspirate the blocking solution and stain with 280  $\mu$ L DAPI (See Note 17) for 5 minutes at room temp in the dark.
10. Aspirate DAPI and add 280  $\mu$ L PBS. Keep cells in the dark until they are ready for imaging (See Note 18). Examples of imaged colonies can be found in Figure 2.3.

### ***RNA-seq of Passaged Cell Lines***

Once colonies have been picked and stable cell lines have been established, it is important to verify the cell types. These samples should be compared to positive and negative controls such as ESC, XEN, and fibroblast cells. A great way to look for subtle differences between samples is to perform RNA sequencing (RNA-seq) on the newly established cell lines and look for transcriptional differences.

1. Collect one 6-well dish of cells, pellet, and place cells in  $-80$  °C freezer until ready to extract RNA.
2. Clean bench top and materials with 70% EtOH followed by an RNase removal spray.
3. Remove cell samples from the freezer and perform a Trizol extract to collect the RNA. To start, add 1mL Trizol to all samples in the fume hood.

4. Next add 200  $\mu$ l chloroform to each sample and shake vigorously until sample reaches an opaque pale pink.
5. Incubate samples for 2 min at room temp.
6. Centrifuge samples at 12,000 xg at 4 °C for 15 min.
7. Carefully remove the upper clear phase to a fresh tube. Leave behind all traces of the white and pink layers. It is better to leave some of the upper phase behind than accidentally extract the white and pink phase; quality is more important than quantity.
8. To the clear phase, add 500  $\mu$ L isopropanol and mix to precipitate the RNA. Incubate at room temp for 10 min.
9. Centrifuge samples exactly as before. Carefully remove and discard most of the isopropanol.
10. Add 1 mL 75% EtOH to wash remaining salts from pellet. Vortex and centrifuge as before. Remove all traces of liquid, air-dry briefly, and dissolve pellet in 100  $\mu$ L RNase free H<sub>2</sub>O.
11. Run samples through the RNeasy kit from QIAGEN to clean up any impurities left over from the Trizol extraction.
12. Once samples have been run through the RNeasy kit, dilute samples 1:10 in RNase free H<sub>2</sub>O.
13. Use the Qubit broad range RNA kit to determine the concentration of RNA in the sample. Use this concentration to calculate the amount of RNA you will need to make a 200 ng/ $\mu$ L concentration of RNA in a total volume of 100  $\mu$ L.
14. With the 200 ng/ $\mu$ L sample, use the DNA TURBO kit to remove any traces of

DNA (it is recommended to use 2  $\mu$ L of TURBO and incubate at 20 minutes at 37 °C).

15. Qubit sample again using the BR RNA kit to determine the sample concentration. Use this concentration to prepare samples to be run on a Bioanalyzer/ TapeStation for quality control and create a final sample to be used for sequencing.
16. Use Qubit one last time to confirm the concentration of the quality control sample and sequencing sample. Use the RNA high sensitivity kit for the quality control sample and RNA broad range kit for the sequencing sample.
17. For quality control testing on a TapeStation it is recommended to have a RINe scores >8.
18. Libraries should be prepared using the correct RNA amount and library preparation kit.
19. Libraries should be sequenced to the appropriate depth with the standard being around 50-90 million with 50bp pair-end end reads per sample.
20. Before mapping, adapter sequences should be removed. Trimmed raw sequencing reads can then be aligned to the appropriate reference genome followed by read counting. Sequence quality need to be evaluated before and after read mapping. Transcripts with low abundance (at least 10 counts per million in at least 3 samples) should be removed from the full data set.
21. Once the transcripts have been processed, additional filtering, MDS plots, heat maps, differential gene expression analysis and volcano plots can be generated to compare samples.

### ***scRNA-seq of Reprogramming Cells***

Reprogramming samples contain a heterogeneous collection of cells in which some cells have undergone a complete transformation into induced stem cell types or even partial reprogramming. As the research community continues to understand the establishment of pluripotency and how to influence cell fate, there is a great need for in-depth analysis of individual cells. One of the best ways to obtain in depth transcriptional analysis of individual cells is through single cell RNA sequencing (scRNA-seq). This allows for the identification of populations of similar cells in a reprogramming sample as well as transcripts of interest.

1. Prepare enough reprogramming sample to fill one 6-well dish. This will provide enough sample for processing. For reprogramming samples or samples with a lot of cell debris and death, passage the sample at 1:1 or 1:2 the day before. Dead cells and cell debris will not replate after passing leading to a significantly cleaner sample. Other techniques such as magnetic-activated cell sorting (MACS) or density gradient centrifugation can also be used to clean up samples.
2. On the day of analysis, wash the well 1x with DPBS.
3. Incubate sample in 1mM EDTA (diluted in DPBS) and incubate at room temperature for 2-4 min.
4. Remove EDTA and wash cells 2x with DPBS.
5. Add 0.25% Trypsin-EDTA and incubate at room temperature for 1-2 min. Remove excess trypsin and incubate cells at 37 °C in 5% CO<sub>2</sub> for 2-4 min (until cells start to slide off the plate).

6. Immediately add reprogramming 1 medium (or other cell medium that contains FBS) to the plate to quench the trypsin.
7. Spin the cell-media mixture at 200 xg for 4 min.
8. Remove media and resuspend cells in an appropriate volume (recommend 500-1000 mL for one confluent 6-well dish of cells) in scRNA-seq Cell Buffer.
9. Filter sample through a 40 mm filter to remove any large clumps of cells. If clumps persist after filtering, gently pipette the sample up and down a couple times.
10. Count cells under the microscope using a hemacytometer to determine cell concentration, the percentage of dead cells and percentage of clumps. Trypan Blue can be used at a 1:2 dilution to determine cell death. Samples should have cell clumping of less than 1% and cell viability of at least 80%. Examples of cell debris and cell death along with acceptable and poor samples can be found in Figure 2.4.
11. Dilute sample to the necessary concentration needed for scRNA-seq. This dilution will be based on the number of cells to be analyzed and target recovery. Keep samples on ice until ready to run on the instrument.
12. Sample should be ran using the correct instrument protocol to create libraries and perform sequencing.
13. One samples have been sequences, base calling is performed, and samples are demultiplexed.
14. Sample is then alignment to the appropriate reference transcriptome followed by cell detection and UMI counting. Only cells with <10% of reads coming from

mitochondrial genes, >10,000 UMIs, and >1000 detected genes were included in analysis.

15. Using cell cycle markers, S-phase and G2M-phase scores can be obtained for each cell. To mitigate the influence of the cell cycle on clustering results, re-normalized raw UMI counts can be completed to regress the percent of reads coming from mitochondrial genes.
16. Cell clusters were determined using the first 30 principal components and the creation of cluster trees. The appropriate resolution was chosen by picking the resolution in the tree that provides stable clusters while still accounting for the appropriate cluster diversity.
17. Cluster enriched genes can then be identified using a log fold change threshold > .25 expressed in at least 1% of the cells in either the cluster of interest or all other cells. P-values should be calculated using the Wilcox rank-sum test and corrected for multiple comparisons with the Bonferroni method. Genes with an adjusted p-value <.01 are considered cluster enriched.
18. Once clusters have been identified, GSEA can be performed to determine the identity of cell clusters. This can be done using various methods including the comparison of the dataset to a published dataset with known cell identity, the determination of top GO terms or pathways expressed in each cluster or more complex analysis tools such as Copybara to identify cell fate and cell transitions when the output is unknown.

## Notes

### *Methods Section*

1. Preserving the anatomy of the embryo is key to making sure all viscera are removed. The more the embryo morphology is disrupted, the more difficult it will be to interpret the anatomy.
2. The head can be saved and used for genotyping, for example when establishing MEFs which carry a specific reporter or allele of interest.
3. Disaggregate tissue as much as possible. This will help the cells proliferate.
4. MMLV derived retrovirus will only infect mouse cells. In comparison lentivirus can infect human and mouse cells and extra precaution must be used when working with lentivirus. Both lentivirus and MMLV derived retrovirus are replication-incompetent retroviruses. This means that both viruses lack the genes that allow infected cells to produce more virus.
5. DMEM can be substituted for Opti-MEM.
6. The same protocol can be used to make lentivirus. In order to make lentivirus use your preferred lentivirus envelope plasmid, packaging plasmid and gene plasmid at a ratio of 2:3:4. For example, our lab uses 2.52  $\mu\text{g}$  PMD2.g, 3.78  $\mu\text{g}$  psPAX2 and 10  $\mu\text{g}$  of gene plasmid when infecting a 10 cm plate.
7. When determining viral titers, we test three different concentrations of virus. This is to ensure that we get an accurate  $C_p$  value and that we are not maxing out the detection of the instrument or using a viral quantity that is too low for detection. For a 6-well transfection, we test the following viral volumes: 0  $\mu\text{L}$  retrovirus, 50  $\mu\text{L}$  retrovirus, 100  $\mu\text{L}$  retrovirus, and 200  $\mu\text{L}$  retrovirus. These

volumes can be adjusted up or down depending on the viral yield.

8. An alternative method of titrating virus is to use a fluorescent protein-expressing virus instead of quantifying transcript levels qPCR. A fluorescent virus can be made in parallel to the virus of interest and infected at different concentration onto MEFs. After 48 hours, fluorescent MEFs are then counted under a fluorescent microscope to record the percentage of fluorescent cells. This percentage can then be used to back calculate the transduction units per mL (TU/mL).  $TU/mL = (\text{Number of cells transduced} \times \text{Percent fluorescent}) / (\text{Virus volume in mL})$ .
9. Reprogramming can be completed with a doxycycline inducible system if MEFs carry *Rosa26-M2rtTA<sup>+/-</sup>* and *mCol1a OSKM<sup>+/-</sup>* (Carey et al., 2010). Instead of adding virus at Day 0, 2 µg/mL doxycycline can be added to the medium and all subsequent medium changes.
10. Colonies will start to emerge at day 7. iPSC colonies look like embryonic stem cell (ESC) colonies, and are generally small in size, well circumscribed and dome-shaped in morphology. iXEN colonies appear flatter with ragged borders and are ~3x larger than iPSCs (Figures 2.1 B and 2.1C).
11. Colonies can be observed as early as 7 days. Clear morphological differences can be observed between iXEN and iPSCs by 14 days.
12. It can be difficult to remove colonies as they are often well attached to the underlying fibroblasts. If the underlying fibroblast layer starts to peel off the dish when removing the colony an alternative technique can be used. Rather than

- taking the whole colony, the surface of the colony can be scraped off and moved to the 96-well.
13. The number of plates used for sorting can be adjusted up or down depending on the cell sorting yield and the number of cells needed for downstream analysis. Note that one 6-well of reprogramming cells is enough to fill 1-2 96 well plates.
  14. Instead of using a mouse line that contains a fluorescent reporter, cell surface marker staining can be performed.
  15. Make sure to have proper controls. These include a live sample with no fluorescence (negative control) to determine background fluorescence, a live sample expressing the fluorophore of interest (positive control) for gating, and a permeabilized and fixed sample with DAPI for live/dead exclusion.
  16. Note that confocal imaging of colonies can be challenging for a number of reasons. As the colonies become large, it is harder to image them, especially iXEN. It is recommended to try imaging at different time points and different magnifications.
  17. Other nuclear stains can be used, such as DRAQ5. It is important to look at the instrument's lasers and available fluorescent antibodies to determine the best nuclear stain and fluorophore panel for the experiment.
  18. When imaging colonies, make sure to test antibody specificity. For an iPSC positive control, mouse embryonic stem cells (ESC) lines can be used. For an iXEN positive control, mouse XEN lines can be used.

### ***Final Thoughts on Analyzing iXEN and iPSC***

Having the techniques to analyze emerging reprogramming cells is important for early and proper detection. When it comes to iPSC and iXEN, morphology and marker expression are very important in distinguishing these two cell types.

As discussed previously in the note section of “OSKM Viral Reprogramming” and demonstrated in Figure 2.1B, iXEN and iPSC colonies have very distinct morphologies from one another. As cells continue to be passaged to form cell lines, iPSC and iXEN begin to look like their ESC and XEN line counterparts respectively. iPSC lines continue form tightly clustered colonies of rounded small cells with an unpolarized epithelium (Figure 2.2 A). By contrast, iXEN lines are mesenchymal, larger than ESCs and are often less rounded in appearance and more geometric (Figure 2.2B).

Morphology is only one marker of cellular identity and may vary depending on the model organism. Follow-up with marker expression should always be performed. Evaluating marker expression can be performed at the protein level using fluorescent imaging with confocal imaging, flow cytometry or western blotting. In addition, RNA levels can be quantified using qRT-PCR or RNA-seq. The type of analysis that is chosen will depend on the lab’s set up and which techniques the lab feels is right for their research. Despite the type of marker analysis that is done, the markers that are used often remain the same. For iPSC, the core markers for pluripotency when looked at in combination are NANOG, SOX2 and OCT4 (Takahashi & Yamanaka, 2006). For iXEN, the core extraembryonic

endoderm markers when evaluated in combination are GATA6, GATA4, SOX7 and SOX17 (Nishimura et al., 2017; Parenti et al., 2016; Y. Zhao et al., 2015).

As a final note, before any analysis is performed, it is important to test all antibodies and primers using appropriate positive and negative controls, especially since there are many commercially available options. For a positive iPSC control, we recommend using a mouse embryonic stem cell (ESC) line. For a positive iXEN control, we recommend using a mouse extraembryonic endoderm (XEN) cell line. Both lines can be derived from blastocysts (Behringer et al., 2014). Tables 1 and 2 contain a list of qRT-PCR primers and antibodies that our lab has found to be reliable.

## APPENDIX

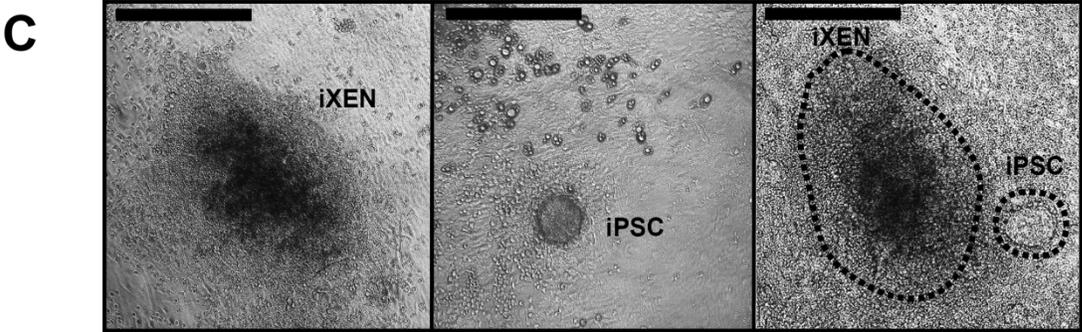
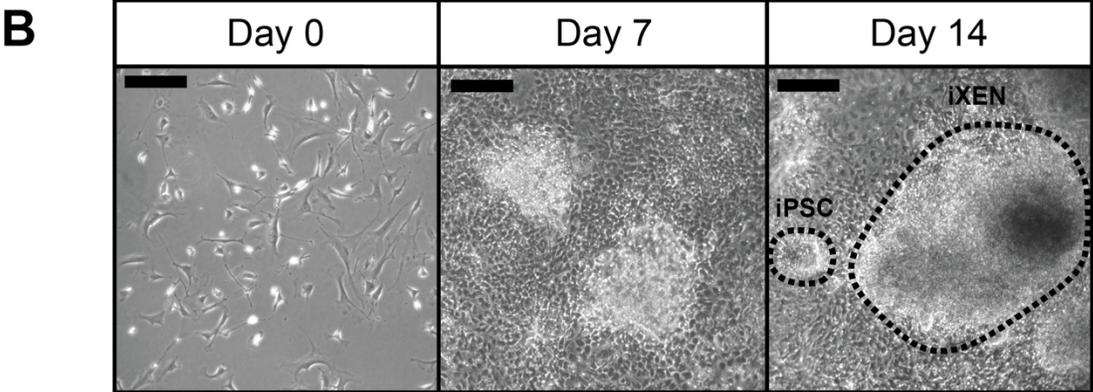
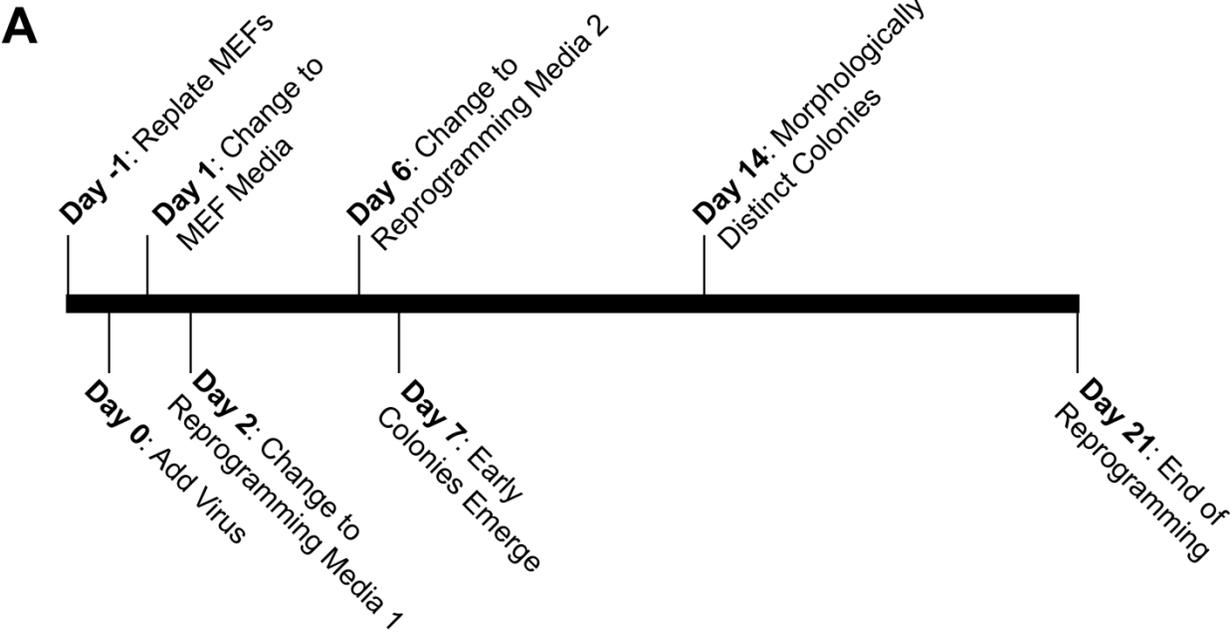
**Table 2.1. qRT-PCR primers for detecting endogenous and viral transcripts**

| <b>Gene Target</b> | <b>Forward Sequence (5' to 3')</b> | <b>Reverse Sequence (5' to 3')</b> |
|--------------------|------------------------------------|------------------------------------|
| <i>Oct4</i>        | GTTGGAGAAGGTGGAACCAA               | CCAAGGTGATCCTCTTCTGC               |
| <i>Nanog</i>       | ATGCCTGCAGTTTTTCATCC               | GAGGCAGGTCTTCAGAGGAA               |
| <i>Sox2</i>        | GCGGAGTGGAAACTTTTGTCC              | CGGGAAGCGTGTACTTATCCTT             |
| <i>Gata6</i>       | ATGCTTGCGGGCTCTATATG               | GGTTTTCGTTTCCTGGTTTG               |
| <i>Gata4</i>       | CTGGAAGACACCCCAATCTC               | ACAGCGTGGTGGTGGTAGT                |
| <i>Sox7</i>        | GGCCAAGGATGAGAGGAAAC               | TCTGCCTCATCCACATAGGG               |
| <i>Sox17</i>       | CTTTATGGTGTGGGCCAAAG               | GCTTCTCTGCCAAGGTCAAC               |
| <i>ActinB</i>      | CTGAACCCTAAGGCCAACC                | CCAGAGGCATACAGGGACAG               |
| <i>Viral Oct4</i>  | GAACCTGGCTAAGCTTCCAA               | ACTTCCTTTCCACTCGTGCT               |
| <i>Viral Sox2</i>  | AACCAAGACGCTCATGAAGAA              | GCTGTAGCTGCCGTTGCT                 |
| <i>Viral Klf4</i>  | CTGAACAGCAGGGACTGTCA               | GTGTGGGTGGCTGTTCTTTT               |
| <i>Viral cMyc</i>  | GCCCAGTGAGGATATCTGGA               | ATCGCAGATGAAGCTCTGGT               |

**Table 2.2. Antibodies for fluorescent imaging**

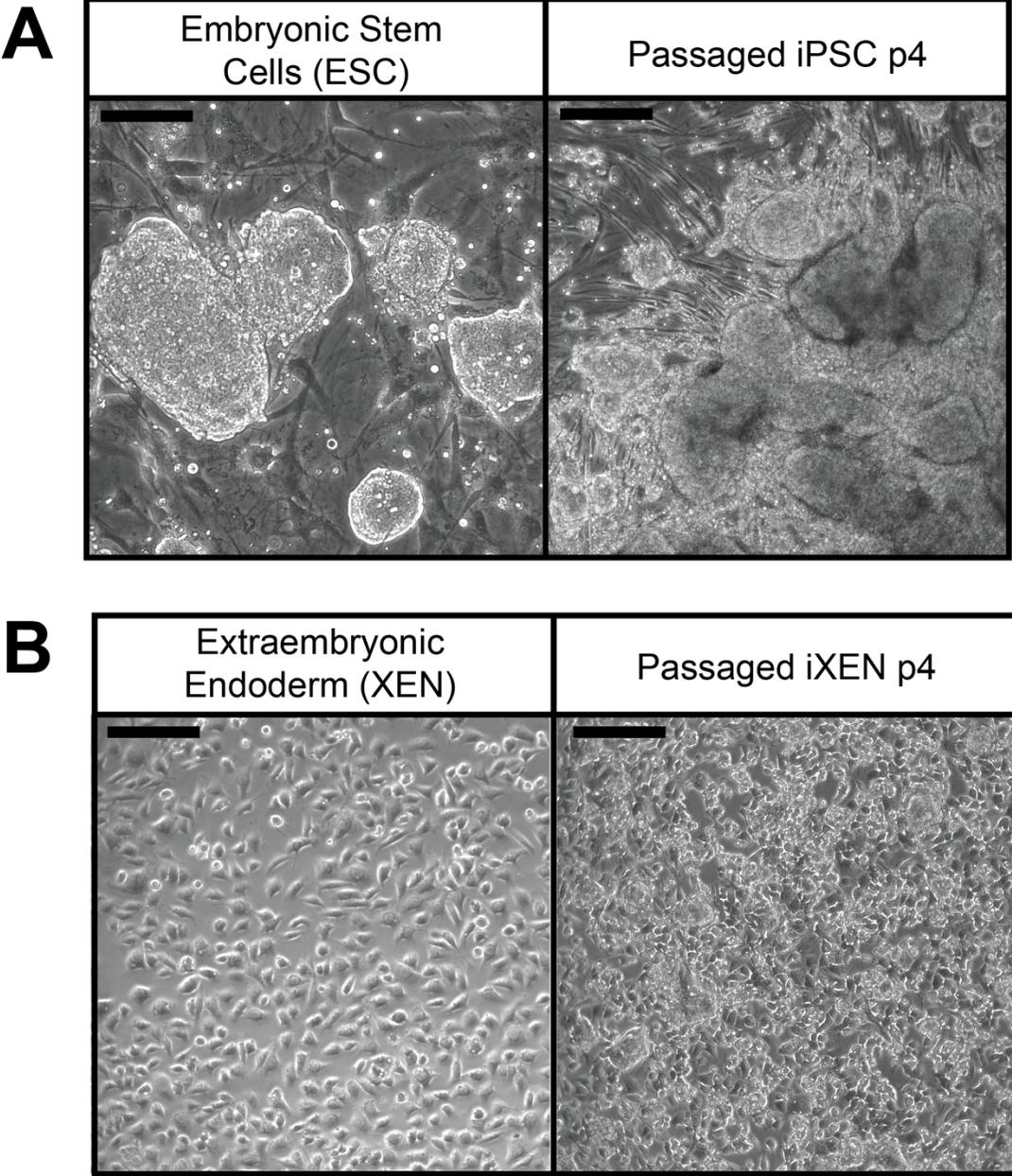
| <b>Antigen</b>      | <b>Antibody Source</b>                |
|---------------------|---------------------------------------|
| SOX17               | R&D Systems (AF1924)                  |
| SOX7                | R&D Systems (AF2766)                  |
| GATA6               | R&D Systems (AF1700)                  |
| GATA4               | Santa Cruz Biotech (sc-1237)          |
| OCT4                | Santa Cruz Biotechnology (sc-5279)    |
| SOX2                | Neuromics (GT15098)                   |
| NANOG               | Reprocell (RCAB0002P-F)               |
| Anti-Mouse IgG 488  | Jackson Immuno Research (715-545-140) |
| Anti-Rabbit IgG 488 | Invitrogen (A10040)                   |
| Anti-Rabbit IgG 647 | Jackson Immuno Research (711-606-152) |
| Anti-Goat 546       | Invitrogen (A11055)                   |

Figure 2.1. iXEN and iPSC formation in OSKM reprogramming.



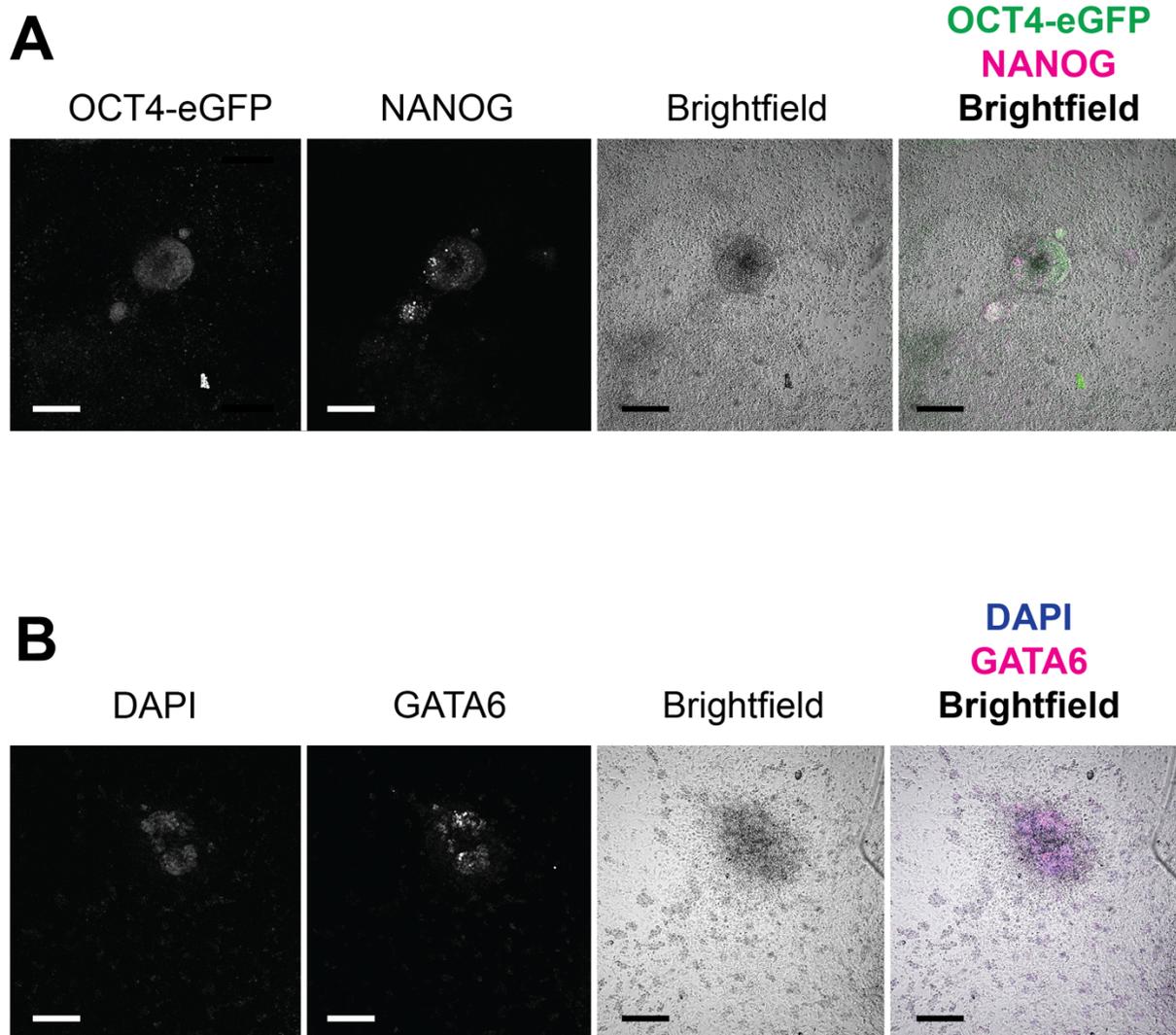
**Figure 2.1. (cont'd).** A) Timeline of reprogramming. B) Colonies of unknown identity first emerge around day 7 with distinct morphologies becoming clear by day 14. C) Days 14 to 21 yield iXEN and iPSC colonies. iXEN colonies are larger with ragged borders. iPSC colonies are smaller with well-defined borders. Bars are equal to 200  $\mu\text{m}$ .

Figure 2.2. Passaged iXEN and iPSC morphology.



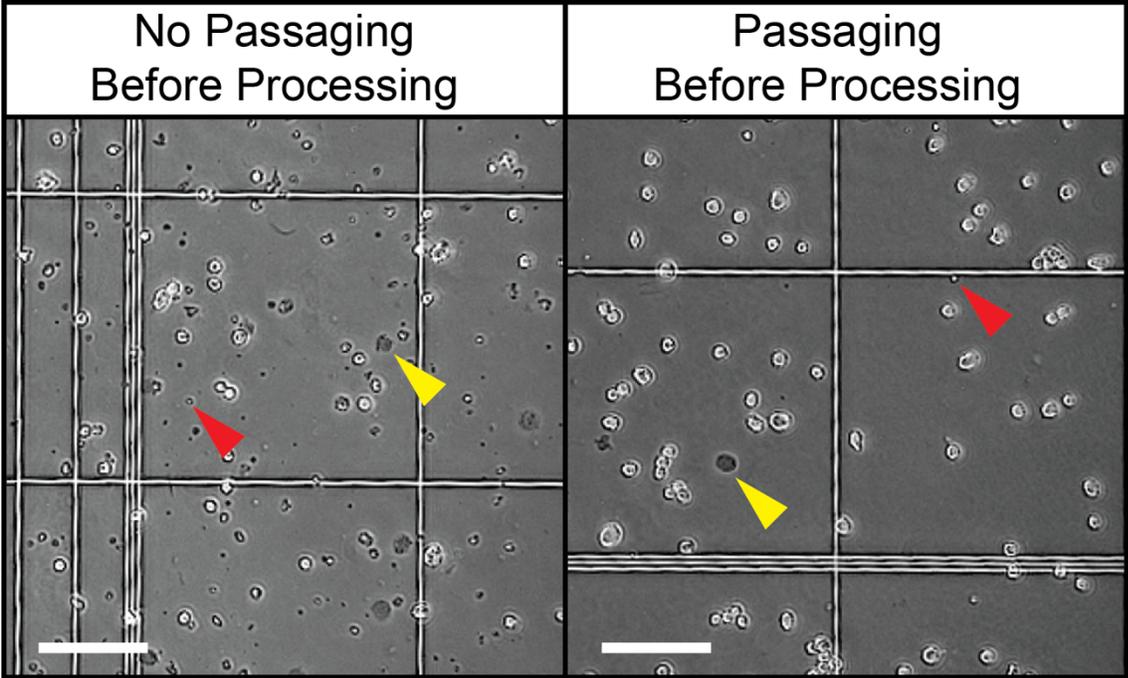
**Figure 2.2. (cont'd).** As colonies are passaged, they acquire morphologies similar to their embryo-derived stem cell counterparts. A) iPSC cells appear epithelial and grow in dome-like colonies. B) iXEN cells appear mesenchymal and individual cells are geometric in appearance. Scale bars are equal to 200  $\mu\text{m}$ .

Figure 2.3. Confocal images of colonies on day 14 of reprogramming.



A and B) Representative colonies of iPSC and iXEN respectively. All images were captured on an inverted Olympus FluoView microscope. The colonies do not appear to uniformly express markers iPSC and iXEN specific markers. This could be due to the plane of focus and differences in gene expression among reprogramming cells. Scale bars are equal to 200  $\mu\text{m}$ .

**Figure 2.4. Examples of cells on a hemacytometer before scRNA-seq library preparation.**



Images show the difference between samples prepared for scRNA-seq in which one sample was passaged the day before submitting (right) and the other sample was not passaged before submitting (left). The sample on the right has less cell debris (red arrow) and cell death (yellow arrow). Images are taken on a Leica upright light microscope. Scale bars are equal to 100  $\mu\text{m}$ .

## CHAPTER 3.

A Closer examination of fluorescent reporters NANOG, OCT4, GATA6 and GATA4 during somatic cell reprogramming reveals unexpected expression in multiple colony types

Moauro A, Kruger R, O'Hagan D and Ralston A\*.

A. Moauro wrote the chapter, assembled the figures, and performed the experiments in Figures 3.1, 3.2, 3.3, 3.4 and S3.1 A, S3.2 C.

Kruger R and O'Hagan D completed the experiments in Figures S3.1B-C and S3.2A-B.

All authors reviewed and edited the chapter.

Chapter is ready for submission to the Journal of Cellular Reprogramming.

## **Abstract**

Somatic cell reprogramming was first developed to create induced pluripotent stem cells (iPSCs). The most commonly used induction method is through the exogenous delivery of transcription factors *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*). To understand how *OSKM* gives rise to iPSCs, several researchers have studied chromatin state, transcription factor binding sites and gene expression during this dynamic process. While trying to uncover the mechanisms that allows for *OSKM* to give rise to iPSCs, it was discovered that other unique stem cell populations arise during reprogramming including induced extraembryonic endoderm stem (iXEN) cells. As the field continues to explore how *OSKM* functions and how changing various conditions affects cell fates during this process, it is important to identify reliable markers that can label cell types of interest. In this study we examined how different fluorescent reporter lines (NANOG, OCT4, GATA6 and GATA4) alone and in combination can identify iXEN or iPSC colonies as they form. We observed a disagreement between fluorescence and colony morphology in all tested reporter lines. We found that morphology in conjunction with the fluorophore expression pattern must be used to reliably identify specific induced stem cell colonies.

## Introduction

Somatic cell reprogramming using various methods including viral transfection or small molecule induction, has been shown to produce induced pluripotent stem cells (Takahashi & Yamanaka, 2006; Y. Zhao et al., 2015). As researchers work to uncover how these methods activate a pluripotent network, it's been observed that other interesting cell populations arise alongside iPSCs during the reprogramming process. These populations include multipotent induced stem cells termed induced extraembryonic endoderm (iXEN) and induced trophoblast stem cells (iTSCs) (Castel et al., 2020; He et al., 2020; X. Liu et al., 2020; Parenti et al., 2016).

Given that reprogramming with *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*) yields a mixed population of cells, it is important to develop methods for tracking the formation of different cell types. If we can identify early cell types as they form, we can better track how they arise in parallel and how they change throughout the 21-day process and beyond. In addition, as the field develops more specific cocktails with better yields, reporter lines will provide an excellent quick readout for successful reprogramming.

Other methods for identifying cells include evaluating the gene or protein expression of select markers. However, many of these methods require lysing or cell fixation. This inadvertently kills the cell and renders a simple snapshot of what is occurring during this dynamic reprogramming process, making it impossible to track the progress of a single cell over time. The use of fluorescent reporters allows for live cell tracking overtime and avoids killing cells. Ideally the fluorophore would be a direct readout for the formation of

the desired cell type such as iPSCs, iXEN or iTSCs. To use fluorescent reporter lines, we must first identify reliable markers that are specific to our cell type of interest.

Many studies use OCT4-eGFP as a readout for successful reprogramming (Dos Santos et al., 2014; Huangfu, Maehr, et al., 2008; Judson et al., 2009; Shi et al., 2008; X. Y. Zhao et al., 2009). This is because OCT4 is an important transcription factor that is necessary to maintain pluripotency (Nichols et al., 1998; Niwa et al., 2000). However, OCT4 is also an important transcription factor in the embryonic development of extraembryonic endoderm (Frum et al., 2013; LeBin et al., 2014). Given that reprogramming produces iPSC and iXEN, it is reasonable that OCT4-eGFP may also be expressed in both cell populations. Luckily NANOG-GFP and NANOG-eGFP lines (Maherali et al., 2007; Xenopoulos et al., 2015) are also available for use as a fluorescent reporter for reprogramming cells and many researchers have taken advantage of these lines to track iPSC formation (Brambrink et al., 2009; Buganim, Faddah, Cheng, Itskovich, Markoulaki, Ganz, Klemm, van Oudenaarden, et al., 2012; Pour et al., 2014; Tsubooka et al., 2009; Xiao et al., 2016). Unfortunately, the use of only NANOG limits the field's ability to identify other cell types such as iXEN and iTSCs. By not using a fluorescent reporter we risk the overestimation of successful iPSC formation or other colony types.

In order to identify reliable reprogramming markers, we focused our search on markers that could differentiate iPSC from iXEN colonies. In embryonic development, pluripotent cells and extraembryonic endoderm cells arise from the same progenitor cell population called the inner cell mass. Given this close relationship in embryonic development, it is

possible that iPSC and iXEN may be challenging populations to distinguish between which emphasizes the need for identifying reliable markers. To do this, we evaluated three single reporter lines and two double fluorescent reporter lines to determine their reliability in identifying putative or early iPSCs and iXEN colonies. We accomplished this task by looking at fluorescent protein expression of colonies relative to morphology to determine if the two corresponded to the same cell type. We then followed up our initial findings by picking iXEN and iPSC colonies at the end of reprogramming (day 21) to create stable cell lines to confirm cell identity. We found that although fluorescent markers are consistently expressed in stable stem cell lines, morphology and fluorescent reporters are often at odds during the reprogramming process.

## **Materials & Methods**

### ***Creation of *Nanog*<sup>mCherry</sup> Mouse***

The *Nanog*<sup>mCherry</sup> mouse was created via a CRISPR/Cas9-mediated knock-in of *mCherry* at the *Nanog* locus using the established protocol from Yang et al., 2013. The p.59995 plasmid was obtained through Addgene. Prior to plasmid use and post-modification, the plasmid was sequenced using Sanger Sequencing and purified using a spin column. The injection of the CRISPR/Cas9 protein, guide RNA and plasmid were performed in C57Bl/6 zygotes and transferred into pseudo-pregnant female mice with the help of the Michigan State University Transgenic Core. Our founder population was immediately bred to CD-1 mice to obtain a CD-1 genetic background. For genotyping, we use the *Nanog*<sup>mCherry</sup> primers in mouse lines and *mCherry* primers for embryos. Primers for genotyping can be found in the Table 2.

### ***Immunofluorescence and Confocal Microscopy***

Embryos were fixed with 4% formaldehyde (Polysciences) for 10 min, permeabilized with 0.5% Triton X-100 (Millipore Sigma), then blocked in 10% fetal bovine serum (Hyclone) with 0.1% Triton X-100 at 4°C overnight. The following day, embryos were incubated in primary antibody at 4°C overnight. Primary antibodies used were goat-anti-Sox17 (R&D, AF1924, 1:2000 dilution) and goat-anti-Sox2 (Neuromics, GT15098, 1:2000 dilution). The next day, embryos were washed for 30 min in block, then stained for 1 hr with donkey-anti-goat Alexa488 secondary antibody (Invitrogen, A-11055, 1:400 dilution). Following staining, embryos were again washed for 30 min in block, then stained for 10 min in DRAQ5 (Cell Signaling Technology, 4084; 1:400 dilution). Nanog-mCherry embryos were imaged after fixing and co-staining or live imaged. Imaging was performed using an Olympus FluoView FV1000 Confocal Laser Scanning Microscope system with a 20× UPlanFLN objective (0.5 NA) and 3× digital zoom. For each embryo, z-stacks were collected with 5 µm intervals between optical sections. Optical sections are displayed as an intensity projection over the z-axis. Figures were prepared using FIJI, Adobe Photoshop, and Adobe Illustrator.

### ***Mouse Strains***

The following alleles were maintained in a CD-1 background; *Gata4*<sup>H2B-eGFP</sup> (Simon et al., 2018), *Gata6*<sup>tm1Hadj/J</sup> (Freyer et al., 2015), *Nanog*<sup>mCherry</sup>, and *Pou5f1*<sup>tm2Jae</sup> (Lengner et al., 2007). All animal work conformed to the guidelines and regulatory standards of Michigan State University Institutional Animal Care and Use Committee.

### ***Mouse Embryonic Fibroblast (MEF) Preparation***

MEF lines were established from E13.5 embryos. After harvesting, head and viscera were removed and subject to DNA extraction and genotyping to confirm MEF genotype. Individual embryos were dissociated and plated on gelatin in MEF Medium [DMEM, 10% Fetal Bovine Serum (Hyclone), Pen/Strep (10,000 units each)] and grown at 37°C with 5% CO<sub>2</sub> until confluent. Once confluent, MEF line were stored in liquid nitrogen.

### ***Reprogramming***

All MMLV-derived retrovirus was produced by transfecting 293T cells with pCL-ECO and pMXs plasmids. pMXs plasmids contained either *Oct4*, *Klf4*, *Sox2* or *cMyc* cDNAs (Addgene #13366, #13367, #13370 and #13375). Transfected 293T cell supernatant was harvested 48 hours later. mCherry virus was made in conjunction with all viral preps and used to infect CD-1 MEFs to determine viral titers. Viral preps were stored at -80 °C. For retroviral reprogramming (Takahashi and Yamanaka, 2006) MEFs are plated the day before at a density of 50-100 cells/mm<sup>2</sup>. Virus is then added at a MOI of 1 with Polybrene and incubated for 24 hrs. The following day medium was replaced with MEF medium, followed by Reprogramming Medium 1 [DMEM (Invitrogen), 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Fetal bovine serum (FBS; Hyclone), 10 ng/mL Leukemia Inhibitory Factor (LIF)] on days 2 and 4. Medium was then replaced with Reprogramming Medium 2 [DMEM (Invitrogen), 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Knockout Serum

Replacement (KOSR; Invitrogen), 10 ng/mL Leukemia Inhibitory Factor (LIF)] on day 6 and every other day there after until the end of the experiment.

### ***Colony Counting and Lab Images***

At various time points until day 20 of reprogramming, the number of iPSC and iXEN colonies were counted based on morphology. In addition, the presence of fluorescent markers were detected using a Lumen Prior 200 connected to an inverted Leica microscope at 10X magnification. Colony images were taken at 20 days post infection.

### ***RNA isolation and qPCR***

RNA was harvested using 1:6 chloroform to Trizol (Invitrogen). 1 µg RNA was reverse transcribed to create cDNA using QuantiTect Reverse Transcription Kit (Qiagen), following manufacturer instructions. For qPCR, cDNA was amplified using a Lightcycler 480 (Roche) according to manufacturer's guidelines. The amplification efficiency of each primer pair (refer to Primers & Oligos Table 1) was measured by generating a standard curve from appropriate cDNA libraries using extraembryonic endoderm (XEN) cells and embryonic stem (ES) cells. All reactions were completed in quadruplicate.

## **Results**

### ***Nanog-2A-mCherry fluorescent reporter mouse line***

To begin our exploration into finding ideal fluorescent markers for the identification of putative iPSCs and iXEN colonies, we chose to start by testing commonly used pluripotency markers. We began by comparing the expression of NANOG to OCT4. To

accomplish this goal, we needed to establish a reporter line that could express both fluorescent markers simultaneously as all published lines only carry green fluorescent proteins. To do this, we established a reporter line with a red fluorescent protein. We created a *Nanog-2A-mCherry* reported line based on published work by Yang et al., 2013, who used CRISPR to create transgenic knock-ins in embryos. Using the published protocol, plasmid and guide RNA from Yang et al., 2013, we were able to recreate a *Nanog-2A-mCherry* mouse also known as *Nanog<sup>mCherry</sup>*.

*Nanog<sup>mCherry</sup>* was created by inserting *mCherry* in frame at the stop codon of the *Nanog* locus to create a translational fusion protein (Fig. 3.6A). Genotyping reveals our knock-in band at the expected size (Fig. 3.6B). Confocal imaging of fixed and stained E3.75 embryos demonstrates that NANOG-mCherry is specific to cells of the pluripotent epiblast due to the colocalization of mCherry with SOX2 and complementation with extraembryonic endoderm marker SOX17 (Fig. 3.6C). NANOG-mCherry fluorescence is decreased with fixation but proves to be a bright marker during live imaging which is critical for the identification of colonies during reprogramming (Fig. 3.7A and B).

### ***NANOG-mCherry and OCT4-eGFP expression alone cannot reliably identify putative iPSC colonies***

Once NANOG-mCherry was proven to be a specific marker for pluripotency in embryos, it was important to test NANOG's specificity during reprogramming. We reprogrammed *Nanog<sup>mCherry/+</sup>* mouse embryonic fibroblasts (MEFs) with *OSKM* retrovirus and began tracking colony formation and fluorescence at day 8 post infection. By day 8, we observed

several colonies starting to emerge with some displaying distinct iPSC or iXEN morphology. iPSC colonies are round compacted domes with smooth borders where iXEN colonies appear flatter and more spread out with rough edges (Parenti et al., 2016). It is not until day 11 that our morphology counts consistently yield a ratio of 1 iPSC colony to every 3 iXEN colonies which is the expected ratio established by Parenti et al., 2016. Additionally, we observe a third main colony type that displays morphological characteristics of both iPSC and iXEN colonies which we have termed Mixed colonies. These Mixed colonies display sections of domed cell-clusters with smooth edges that cascade into more flattened colonies with rough undefined borders (Fig. 3.1 C and D). These Mixed colonies are produced at a similar ratio to iPSCs in which for every 1 iPSC colony there is 1 Mixed colony observed. Little is known about these Mixed colonies. Given their interesting morphological characteristics, we chose to keep track of their formation but acknowledge the need for additional studies to truly understand their potential.

When tracking fluorescence, NANOG-mCherry is expressed by day 11 and detected in up to 40% of colonies (Fig. 3.1B and C). Interestingly, we observed a decrease in the number of fluorescent colonies by the end of reprogramming which may be due to the large colony size and limited ability of the laser light to penetrate and excite these colonies. Alternatively, this could be a true decrease in fluorescent expression in which some colonies only transiently express NANOG-mCherry. By day 20, we observed that the number of NANOG-mCherry expressing colonies matched the number of iPSC morphology colonies. This would seemingly indicate that NANOG-mCherry is a good

reporter for iPSC colony formation by the end of reprogramming (Fig.3.1B). If NANOG-mCherry were a good marker for iPSC formation, this would not explain why iPSC morphology numbers do not match with NANOG-mCherry fluorescence throughout reprogramming.

Taking a closer look at NANOG-mCherry positive colonies reveals that NANOG-mCherry is not restricted to iPSC colonies but is also observed in other colony types (Fig. 3.1C and D). Each colony type (iPSC, iXEN and Mixed) expressed NANOG-mCherry. This indicates that NANOG-mCherry alone is not a specific marker for early iPSC colony formation. Although NANOG-mCherry may not be specific for putative iPSC colonies, the expression pattern among iPSC, iXEN and Mixed colony types is different which can be used as an advantage in identifying true iPSC colonies. In iPSC colonies, NANOG-mCherry is expressed evenly throughout the colony, while iXEN and Mixed colonies express NANOG-mCherry in diffuse patches (Fig. 3.1D). In addition, Mixed colonies tend to only express NANOG-mCherry in areas that are more consistent with iPSC morphology. This emphasizes the importance of using morphology and fluorescence in conjunction to identify true iPSC colonies.

An ideal pluripotency marker would be restricted to only iPSC colonies. Since there is ectopic expression of NANOG-mCherry in iXEN and Mixed colonies, this indicates a disagreement between fluorescence and colony morphology (Fig. 3.1B). It is known that reprogramming is a dynamic process, and that genetic expression changes frequently in iPSC cells until stable cell lines are well established (Polo et al., 2010). It is possible that

NANOG-mCherry is a reliable marker for pluripotency, but reprogramming colonies are too infantile to reliably express NANOG. To determine if NANOG-mCherry becomes restricted to pluripotent cells, colonies were picked and expanded to a passage 12 or greater to create stable cell lines. Cell lines were created using the methods from Parenti et al., 2016 which shows that iPSC and iXEN colonies can be picked to create embryonic competent cell lines that are indistinguishable from their embryo-derived stem cell counterparts. Stable cell lines show that NANOG-mCherry is indeed exclusively expressed in iPSC cell lines and not expressed in iXEN cell lines. These iPSC cell lines expressing NANOG-mCherry also express appropriate morphology and pluripotent markers relative to embryo-derived stem cell controls termed extraembryonic endoderm stem (XEN) cells and pluripotent embryonic stem (ES) cells (Fig. 3.1E and F) (Evans & Kaufman, 1981; Kunath et al., 2005; Martin, 1981).

Next, we chose to evaluate OCT4-eGFP with NANOG-mCherry to determine if two pluripotency reporters in combination could better identify putative iPSC colonies without identifying iXEN or Mixed colonies. Overall, we observed that there are slightly fewer double fluorescent expressing colonies relative to single reporter expression (Fig. 3.1D and 3.2A). However, OCT4-eGFP and NANOG-mCherry double expression was not only observed in iPSC colonies but also Mixed and iXEN colonies (Fig 3.2A and B). This indicates that together these markers are not able to better identify putative iPSC colonies than they are individually. Interestingly, we detected a similar expression pattern between OCT4-eGFP and NANOG-mCherry in the different colony types indicating that the same cells are likely expressing both markers (Fig. 3.2B).

## ***GATA6-H2B-Venus is an unreliable and non-specific reporter for putative iXEN colony identification***

Currently there are no widely used fluorescent reporters in the reprogramming field that track iXEN formation. Some of the most popular markers of XEN include GATA6, GATA4, SOX7, SOX17 and PDGFR $\alpha$ . To date only three of these markers have an available fluorescent reporter mouse line which include *Gata6*<sup>H2B-Venus</sup>, *Gata4*<sup>H2B-eGFP</sup> and *Pdgrf $\alpha$* <sup>eGFP</sup> (Freyer et al., 2015; Hamilton et al., 2003; Simon et al., 2018). Unfortunately, PDGFR $\alpha$  is not a specific marker for XEN formation as the protein is a general mesenchymal marker and widely expressed in several cell types including fibroblasts (Hamilton et al., 2003). It would be challenging to find new PDGFR $\alpha$  expression in iXEN when background fibroblasts also express PDGFR $\alpha$ . However, this still leaves two GATA reporters to be evaluated.

Like in previous reprogramming with *Nanog*<sup>mCherry</sup>, we observe the formation of all three colony types at the expected ratios when reprogramming with *Gata6*<sup>H2B-Venus/+</sup> MEFs. When evaluating GATA6 we detect Venus expression around day 14 post infection. At most, GATA6-H2B-Venus is expressed in 30% of all colonies (Fig. 3.3A). If GATA6 is a reliable marker for iXEN colony formation, we would expect that all GATA6-H2B-Venus expressing colonies would display only an iXEN morphology. Surprisingly we observed that all three colony types (iPSC, iXEN and Mixed) express GATA6-H2B-Venus (Fig. 3.3B and C). In addition, GATA6 is nonuniformly expressed within all three colony types in which not every cell appears to be expressing Venus (Fig. 3.3C). Unlike NANOG-

mCherry, we were not able to speculate the colony type based on the pattern of Venus expression alone.

Together, this information would suggest that GATA6 is not a reliable marker for iXEN colony formation. Like NANOG-mCherry, GATA6 appears to be ectopically expressed in iPSC and Mixed colonies. To test if GATA6-H2B-Venus becomes restricted to iXEN cells upon cell maturation, putative colonies were picked and passaged to create stable cell lines. After passaging, all *Gata6*<sup>H2B-Venus/+</sup> cell lines expressed the appropriate morphology and key extraembryonic endoderm genes. However, GATA6-H2B-Venus was only expressed in some iXEN lines despite qPCR showing *Gata6* expression (Fig. 3.3D and E). It would appear that the iXEN lines with no Venus expression are preferentially expressing the wild type *Gata6* allele over the *Gata6-H2B-Venus* allele. This mismatch has also been observed in Freyer et al., 2015, and was speculated to be due to a monoallelic expression of *Gata6*. This potentially monoallelic expression means that several GATA6 expressing colonies are going undetected by fluorescence and producing an inaccurate readout.

***GATA4-H2B-eGFP is minimally expressed during reprogramming but does not express in putative iPSC colonies***

When evaluating *Gata4*<sup>H2B-eGFP</sup> reprogramming MEFs, we observed all three colony types at the expected ratios and observed eGFP expression as early as day 11. However, GATA4-H2B-eGFP expression is reduced and only expressed in 10% of the total colonies at most (Fig. 3.4A). Interestingly, GATA4-H2B-eGFP is only expressed in iXEN or Mixed

morphology colonies and not in iPSC colonies (Fig. 3.4B and C) indicating that it may be more specific than GATA6-H2B-Venus at the cost of identifying less putative iXEN colonies. Like GATA6-H2B-Venus, GATA4-H2B-eGFP has a non-uniform expression pattern within iXEN and Mixed colonies in which not every cell of the colony appears to be expressing eGFP. Instead eGFP is expressed in patches throughout the colony regardless of whether the colony has an iXEN or Mixed morphology. The decreased GATA4-H2B-Venus expression in total colonies relative to GATA6-H2B-Venus expression could be explained by embryonic development. As the embryo develops and forms extraembryonic endoderm, it is known that GATA4 turns on after GATA6 (Artus et al., 2011; Plusa et al., 2008). It is possible that early putative iXEN colonies have not yet begun to express GATA4 and need more time to mature.

To test if GATA4-H2B-eGFP is a reliable marker for iXEN, putative colonies were picked and passaged to create stable cell lines. After passaging, all  $Gata4^{H2B-eGFP/+}$  iXEN cell lines expressed eGFP in conjunction with appropriate morphology and markers (Fig. 3.4D and E). Unlike  $Gata6^{H2B-Venus/+}$ , all  $Gata4^{H2B-eGFP/+}$  iXEN lines expressed GATA4-H2B-eGFP by passage 5 (Fig. 3.4F).

Although GATA6-H2B-Venus and GATA4-H2B-eGFP on their own did not prove to be reliable markers for iXEN colony formation, it is possible that these markers may be more reliable when used in combination with another fluorescent reporter. We chose to evaluate GATA4-H2B-eGFP with NANOG-mCherry given that GATA4-H2B-eGFP was reliably turned on in all iXEN lines where GATA6-H2B-Venus failed to become a reliable

reporter in stable iXEN lines. Given that GATA4-H2B-eGFP is a late marker of extraembryonic endoderm formation, there were fewer colonies to evaluate. We were able to observe a small portion of double positive colonies. All double positive colonies were of mixed morphology despite single fluorescent evidence that GATA4-H2B-eGFP and NANOG-mCherry are present in iXEN colonies (Fig. 3.4G and H). This lack of double fluorescence in iXEN colonies could be due to the small sample size of double positive colonies in which increasing the sample size could lead to the possible finding of a double positive iXEN colony. Alternatively, iXEN colonies do not express both GATA4-H2B-eGFP and NANOG-mCherry simultaneously. In addition, the mixed colonies that expressed both reporters had distinct fluorophore expression patterns for GATA4-H2B-eGFP and NANOG-mCherry that displayed minimal overlap. Together this would suggest that cells which express GATA4-H2B-eGFP do not express NANOG-mCherry (Fig. 3.4H). The use of GATA4-H2B-eGFP in combination with NANOG-mCherry has proven to be a more specific reporter line as double fluorescence is only observed in Mixed colonies. Where single fluorescent colonies were more likely to be iXEN or iPSC colonies.

## **Discussion**

When looking at fluorescence we observe that the morphology of a colony does not always match with the fluorescent reporter read out during reprogramming but that this mismatch disappears in stable cell lines (Fig. 3.5A and B). When using NANOG-mCherry, we saw its expression in iPSCs, iXEN and Mixed colonies. Fortunately, there was a distinct difference in expression pattern between colonies. In iPSC colonies, the entire colony expressed mCherry whereas iXEN and Mixed colonies only expressed mCherry

in patches. The use of NANOG-mCherry with OCT4-eGFP revealed similar data in which the double expression was not selective for only iPSC colonies. In addition, the expression of NANOG-mCherry and OCT4-eGFP was almost identical indicating that these two markers are most likely expressed in the same cells. NANOG-mCherry and OCT4-eGFP, both alone and together, are not exclusively expressed in iPSC colonies during cellular reprogramming and cannot be used as a stand-alone indicator of colony identity. We therefore recommend using the expression pattern of the fluorophore to better identify true iPSC colonies.

When evaluating GATA6-H2B-Venus as a marker for iXEN formation, we saw very similar results to NANOG-mCherry in which all three colony types express Venus. Unlike NANOG-mCherry we did not detect a difference in the expression pattern of Venus between colony types. In addition, we found that not every iXEN cell line expressed Venus despite expressing *Gata6*. One concern is that *Gata6* may be monoallelic in extraembryonic cell types which is novel and presents broader implications in gene regulation. These findings should be followed up with additional studies to look into chromatin state using CHIP-seq and allele specific targeted sequencing (AST-seq) in different tissue types to determine if GATA6 is truly monoallelic (Nag et al., 2013). Alternatively, this could be artifact due to the creation of the *Gata6*<sup>H2B-Venus</sup> line in which *Gata6* is knocked out and replaced by *Venus*. This is doubtful given that GATA6-H2B-Venus is a reliable marker during cardiac and gut endoderm formation (Freyer et al., 2015). Altogether, we strongly recommend that morphology be considered if using GATA6-H2B-Venus to identify putative iXEN colonies.

GATA4-H2B-eGFP was less widely expressed and more specific than GATA6-H2B-Venus. GATA4-H2B-eGFP was not expressed in any iPSC colonies and was only observed in iXEN or Mixed colonies. Unfortunately, the number of colonies expressing GATA4-H2B-eGFP was low relative to NANOG-mCherry and GATA6-H2B-Venus. Since GATA4-H2B-eGFP identifies iXEN and Mixed colonies we still suggest using morphology in conjunction with eGFP expression to identify putative iXEN colonies.

Interestingly, when pairing GATA4-H2B-eGFP with NANOG-mCherry we detect both fluorophores in Mixed colony types only. In addition, the colonies that express both fluorophores have a different expression pattern indicating that the same cells do not express both markers simultaneously. This is further supported by previous reprogramming studies in which *Gata4* overexpression has been shown to repress *Nanog* during *OSKM* reprogramming (Serrano et al., 2013). The expression of both GATA4 and NANOG within mixed colonies suggests mixed colonies are a heterogeneous group of cells that have possibly taken on separate cell identities. This is an interesting finding in which single cell methodology such as flow cytometry and scRNA-seq can help us determine if NANOG and GATA4 are truly not co-expressed at a single cell level and what other genes might be expressed in these positive cell types.

Our study shows that NANOG, OCT4, GATA6 and GATA4 are not reliably expressed in their respective iPSC or iXEN colonies which leads to several interesting points worth resolving in the future. First, this study has reaffirmed the dynamic nature of the reprogramming process (Cacchiarelli et al., 2015; Chronis et al., 2017; Knaupp et al.,

2017; D. Li et al., 2017; Raab et al., 2017; Soufi et al., 2012, 2015; Takahashi et al., 2014). It appears that even late into reprogramming colonies are still settling into their final cell fate which explains the lack of fluorescence in colonies of interest and possible ectopic expression. Fortunately, this dynamic nature extends the window in which cells can be studied for understanding how cell fate networks are turned on and how ectopic gene expression is downregulated. Future studies should include looking at the change in gene expression from the time of colony picking to the establishment of a stable cell line.

Next, we also observed that NANOG and GATA6 are expressed in all colony types. Understanding if this expression is ectopic or important to reprogramming will help elucidate this finding. When more closely examining the literature behind GATA6, evidence exists for both scenarios. Studies have shown that exogenous GATA6 plus SOX2, Klf4 and c-MYC are sufficient to produce iPSCs (Shu et al., 2015a) while others have shown GATA6 hinders iPSC formation (Mikkelsen et al., 2008; Serrano et al., 2013). In addition, when evaluating GATA6 expression in embryonic formation it is observed that GATA6 is present in the precursor cell population that gives rise to both pluripotent and extraembryonic endoderm cells (Artus et al., 2011; Plusa et al., 2008). If reprogramming follows development, it is possible that GATA6 does play a role in establishing iPSCs. Like GATA6, NANOG is also expressed in the same precursor cell population that gives rise to pluripotent and extraembryonic endoderm cells in the embryo (Strumpf et al., 2005). The same logic may be applied to NANOG expression. Better single cell resolution studies to examine a cells transcriptome will help explain this conundrum. In addition,

knocking out *Gata6* or *Nanog* can be completed to determine how iPSC and iXEN colony formation is impacted.

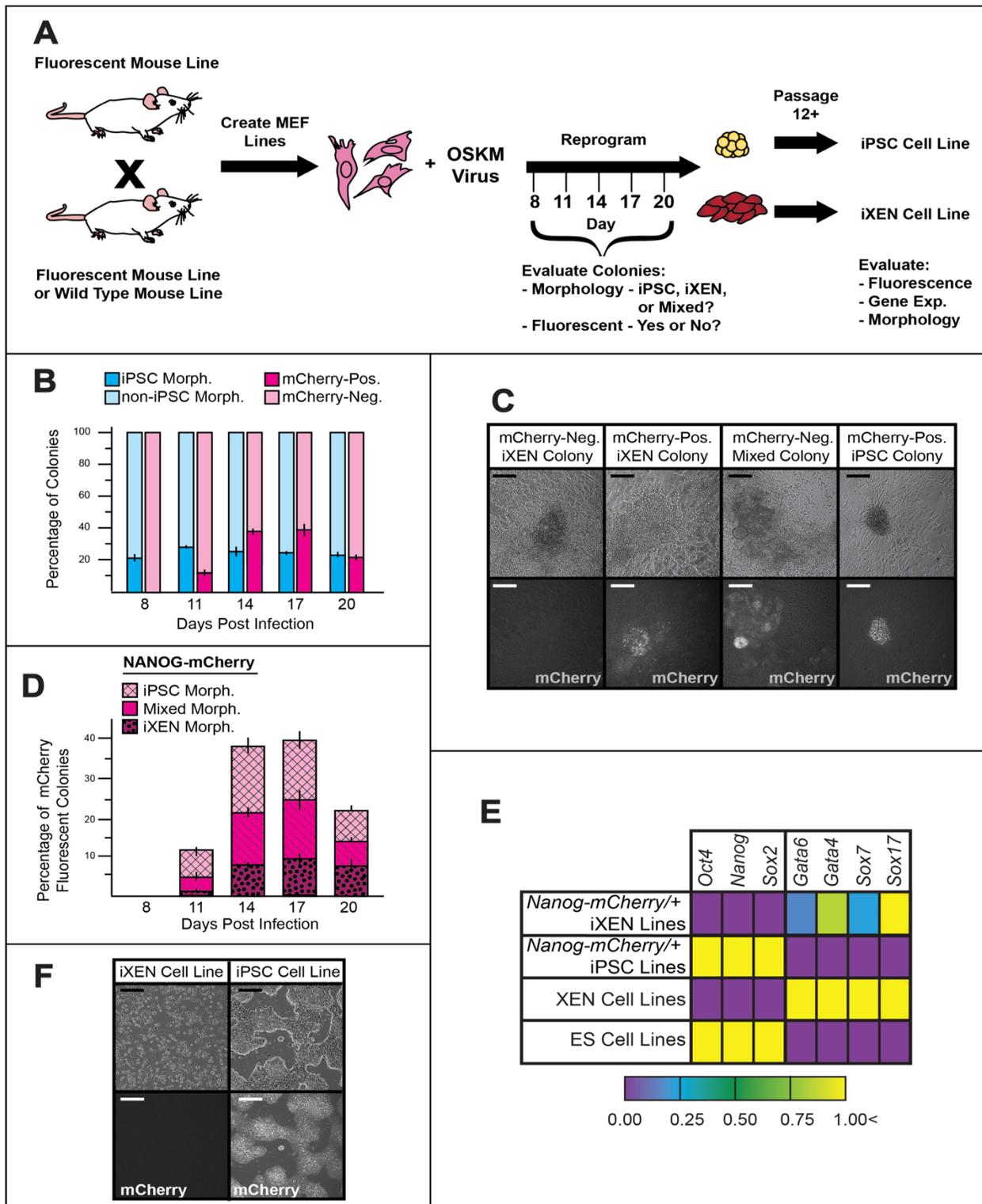
Lastly, our study found a significant amount of mixed morphology colonies in all reprogramming experiments. It is easy to discount these colonies as failed or partial reprogramming, but they display unique characteristics of iPSC and iXEN colonies regarding morphology and fluorescent marker expression. They even uniquely express NANOG-mCherry and GATA4-H2B-eGFP in which both fluorophores do not appear to be expressed in the same cells within the colony. Mixed colonies may contain a unique population of cells that have the propensity to take on iPSC or iXEN fates. Additional studies will need to be done to test the potential of Mixed colonies. These colonies can be picked and expanded in selective ES or XEN media to determine if they can form both iPSC and iXEN cell lines.

### **Acknowledgements**

We would like to thank Anna-Katerina Hadjantonakis and her lab for providing us *Gata4*<sup>H2B-eGFP</sup> and *Gata6*<sup>tm1Hadj/J</sup> mice, Shinya Yamanaka and his lab for the development of the Addgene plasmids. Lastly, we would like to thank the Michigan State University Transgenic Core for their help in creating the *Nanog*<sup>mCherry</sup> mouse.

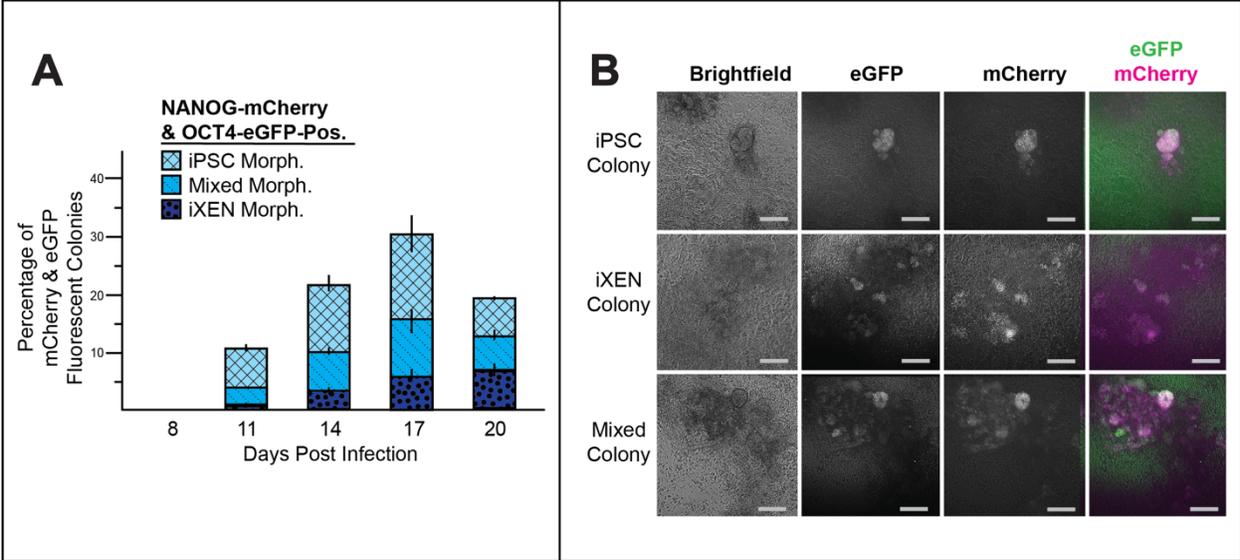
## APPENDIX

**Figure 3.1. NANOG-mCherry expression during reprogramming shows specificity in cell lines but not colonies.**



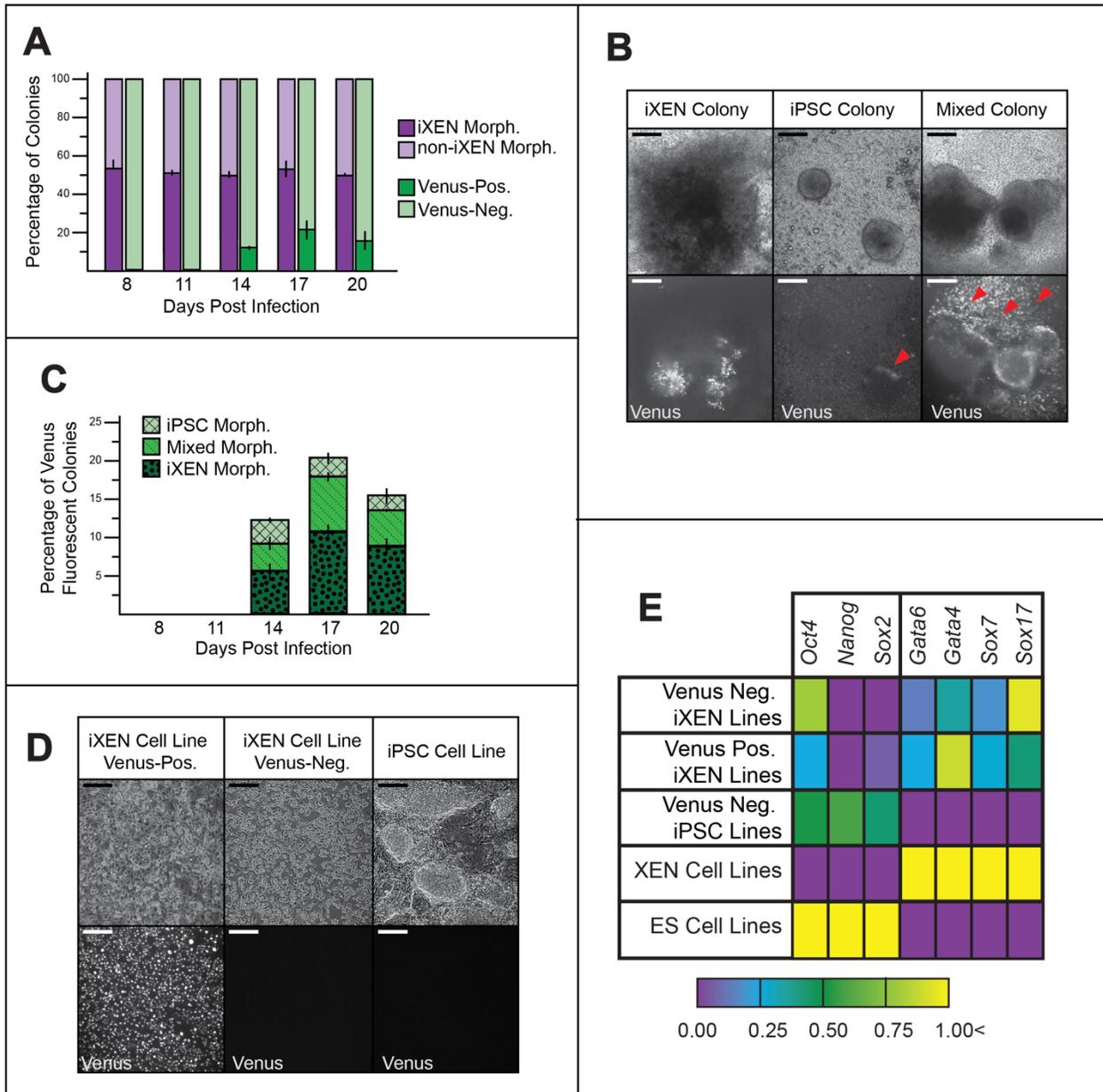
**Figure 3.1. (cont'd).** A) Workflow of experiments including the creation of MEF lines, evaluation of putative colonies during the 21-day reprogramming process and picking of colonies at the end of reprogramming to produce stable induced stem cell lines. B) Total mCherry expression is higher than expected during days 14 and 17 of reprogramming relative to observed iPSC morphology colonies but normalizes by day 20. n = 3. C and D) Further evaluation of mCherry expressing colonies reveals that iPSC along with iXEN and Mixed colonies express mCherry. E) qPCR confirms that established iXEN and iPSC lines express appropriate markers. qPCR marker expression is relative to positive control lines XEN or ES. Error bars are calculated using standard error. n = 3 lines. F) Cell lines created from iPSC and iXEN colonies demonstrate that mCherry becomes restricted to all iPSCs lines. Scale bar = 200  $\mu$ m. Exp. = expression, Pos. = Positive, Neg. = Negative, Morph. = Morphology.

**Figure 3.2. Reprogramming with NANOG-mCherry & OCT4-eGFP reporters show the same fluorescence pattern and specificity.**



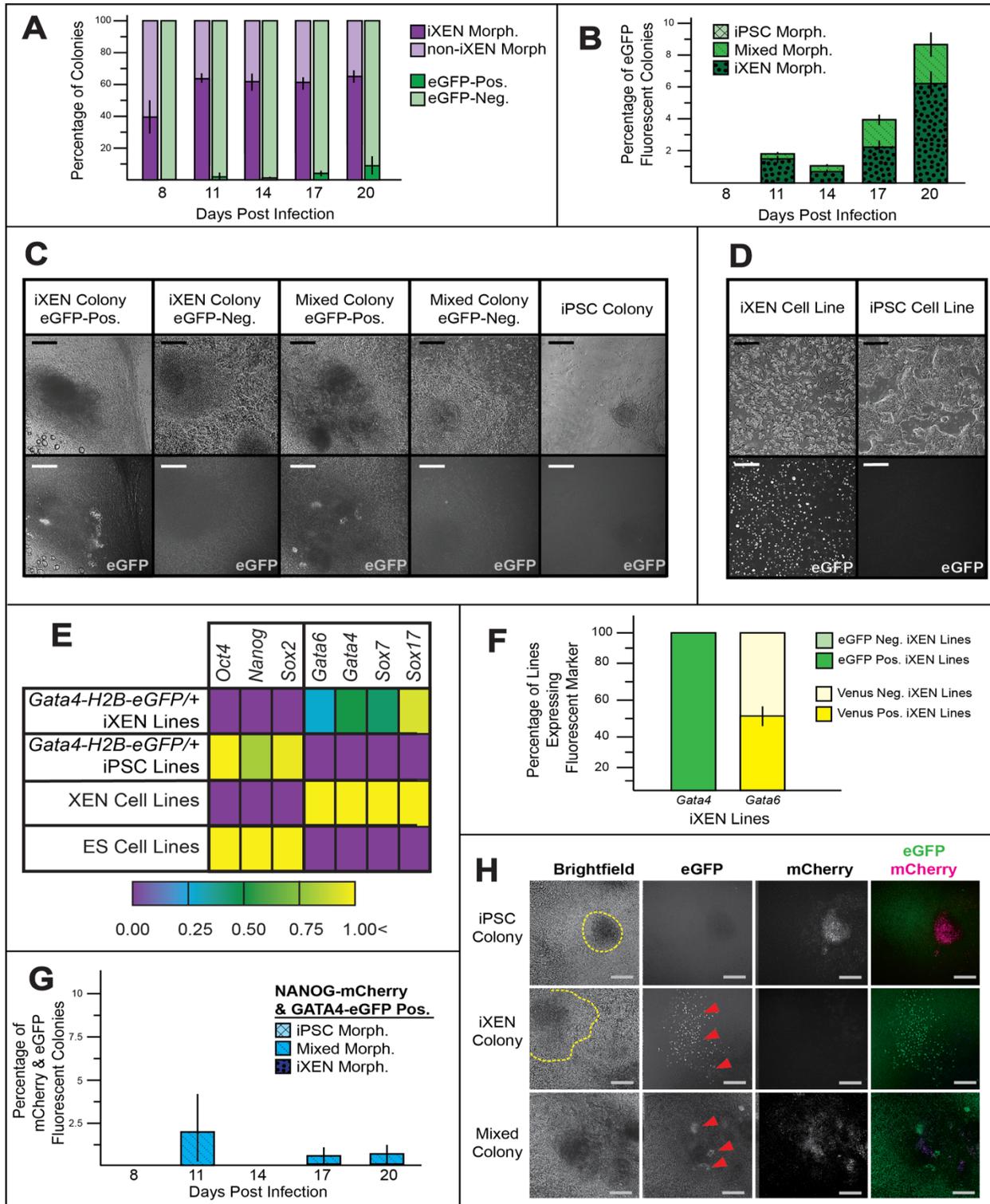
A and B) Reprogramming with double reporters *Nanog*<sup>mCherry/+</sup> and *Oct4*<sup>eGFP/+</sup> shows double fluorescence in all colony types and a similar expression pattern between NANOG and OCT4. Error bars are calculated using standard error. Scale bar = 200  $\mu$ m. Error bars are calculated using standard error. n = 3.

**Figure 3.3. GATA6-H2B-Venus expression during reprogramming shows specificity in cell lines but not colonies.**



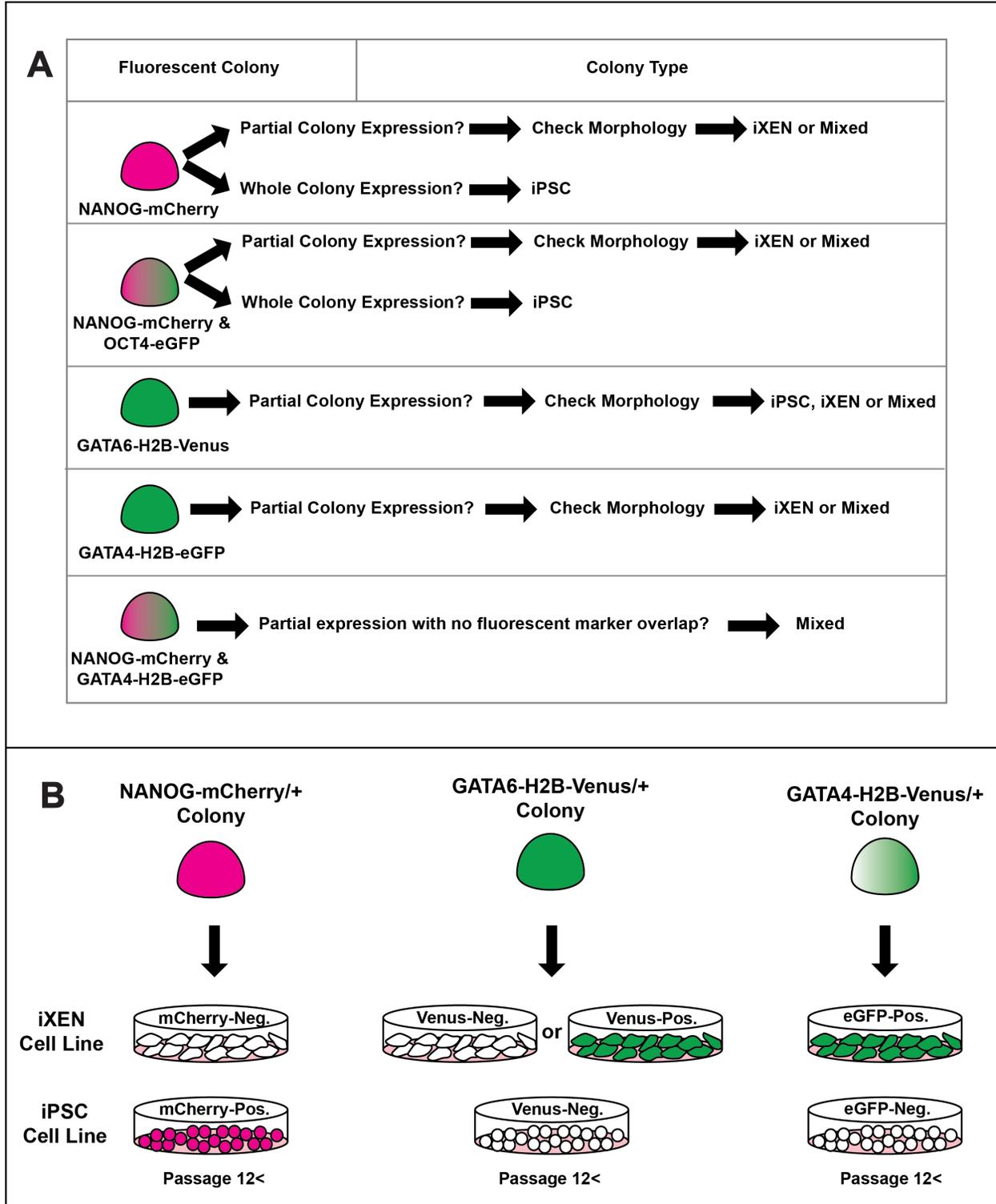
**Figure 3.3. (cont'd).** A) Total Venus positive colonies is low relative to observed iXEN morphology colonies (n=3). B and C) Further evaluation of Venus expressing colonies reveals that iXEN along with iPSC and Mixed colonies all express Venus. D) Cell lines created from iXEN and iPSC demonstrate that Venus becomes restricted to iXEN lines but that not all iXEN lines express Venus. E) Despite a lack of Venus expression in all iXEN lines, qPCR shows appropriate marker expression in iXEN lines (n=4). qPCR marker expression is relative to positive control lines XEN or ES. Scale bar = 200  $\mu$ m. Error bars are calculated using standard error.

**Figure 3.4. GATA4-H2B-eGFP expression during reprogramming shows specificity in cell lines but not colonies.**



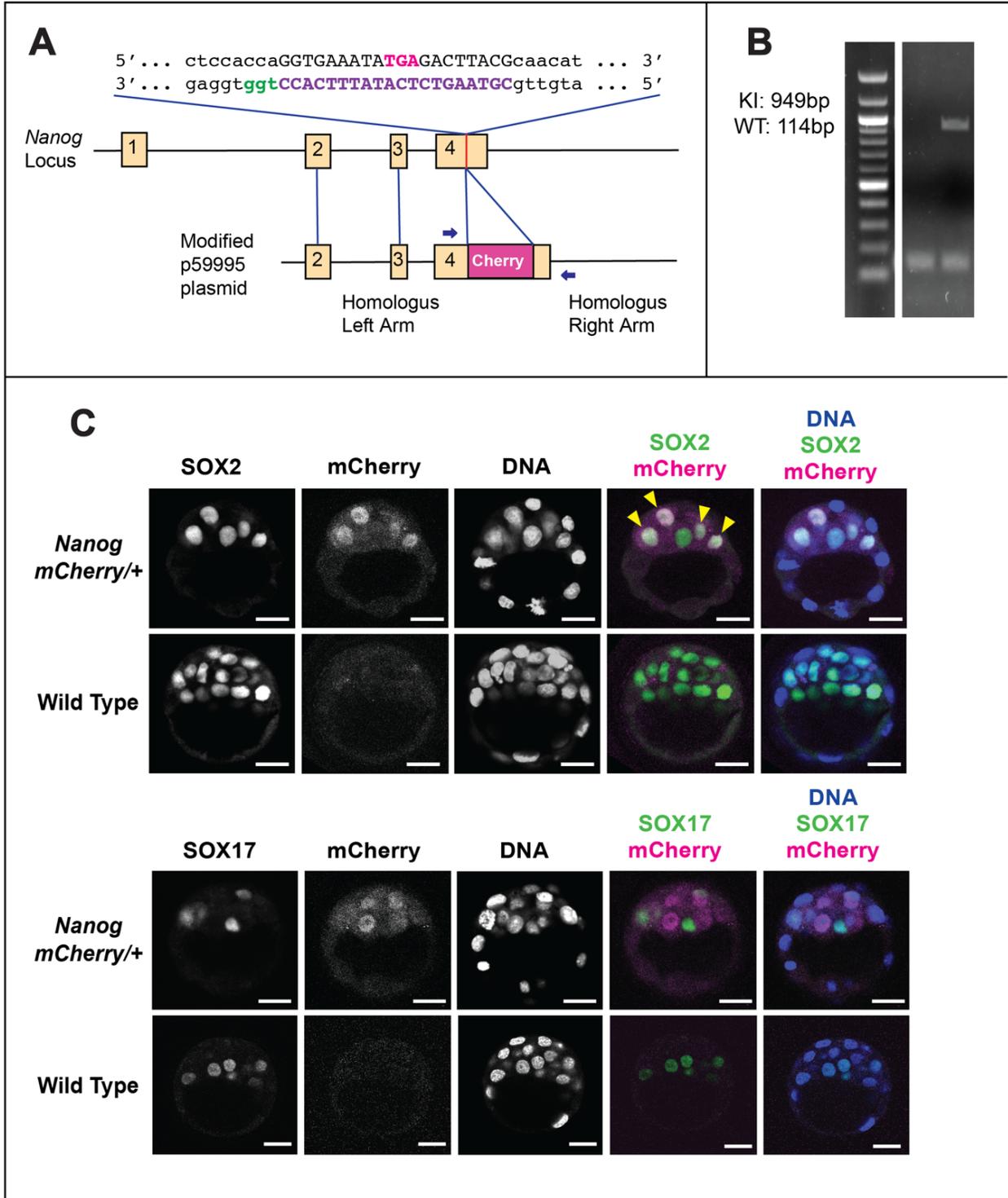
**Figure 3.4. (cont'd).** A) Total eGFP expression is low relative to observed iXEN morphology colonies (n=3). B and C) Further evaluation of eGFP expressing colonies reveals that iXEN along with Mixed colonies express eGFP but that eGFP is not observed in iPSCs. D) Cell lines created from iXEN and iPSC colonies demonstrate that GATA4-eGFP continues to be specific to iXEN lines and that all iXEN lines express eGFP. E) qPCR confirms that established iXEN and iPSC lines express appropriate markers. qPCR marker expression is relative to positive control lines XEN or ES (n=3). F) All *Gata4*<sup>H2B-eGFP</sup> iXEN lines expressed eGFP where only half of the *Gata6*<sup>H2B-Venus</sup> iXEN lines expressed Venus. G and H) Reprogramming with double reporters *Nanog*<sup>mCherry</sup> and *Gata4*<sup>H2B-eGFP</sup> shows very few colonies with double marker expression and the colonies that express both markers are of Mixed morphology. Scale bar = 200  $\mu$ m. Error bars are calculated using standard error.

**Figure 3.5. Fluorescent reporter summary.**



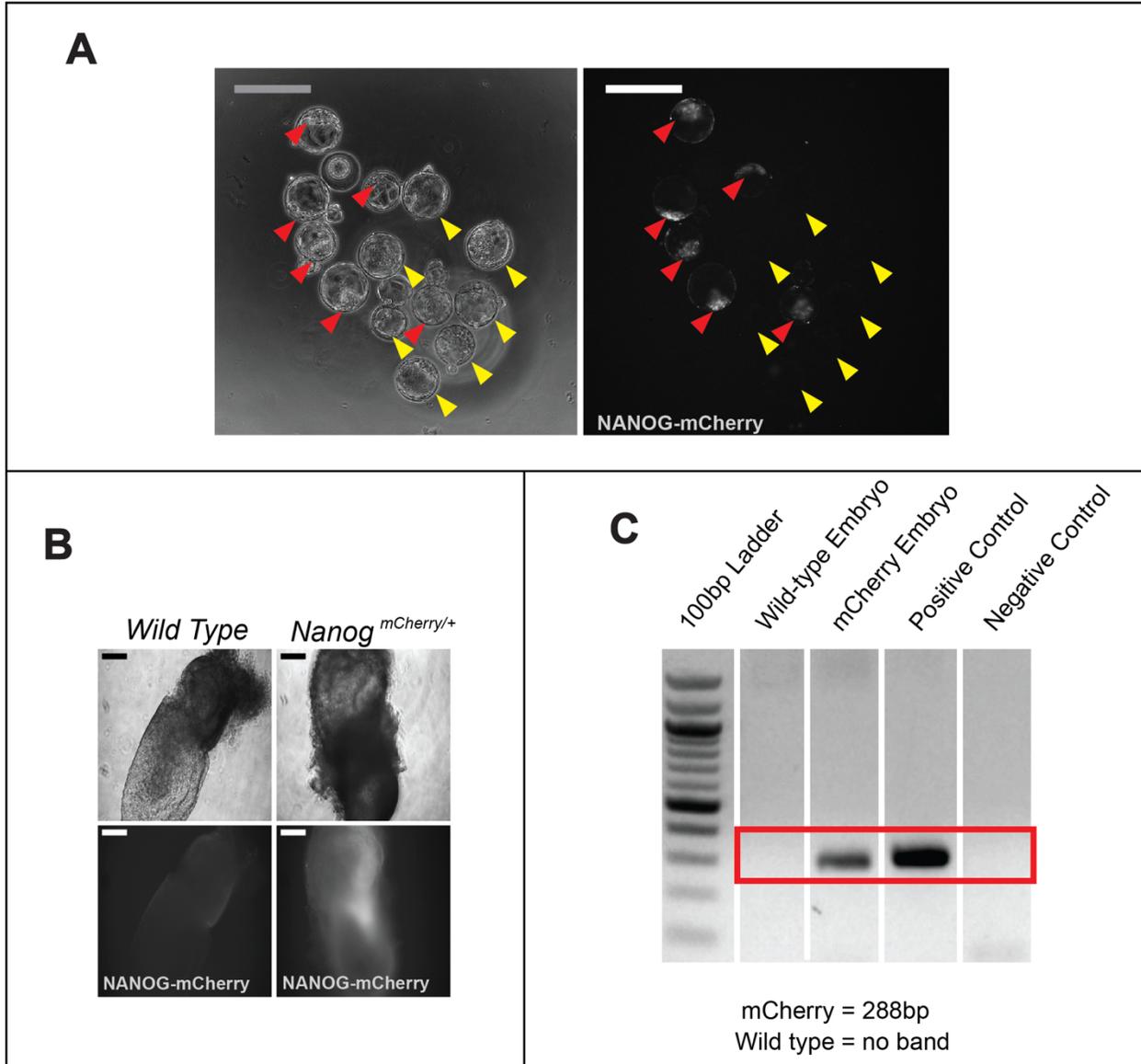
**Figure 3.5. (cont'd).** A) When using fluorescence as a readout for a specific colony it is important to look at the expression pattern of the fluorophore as this can be a strong indication of the colony type, some fluorophores are ubiquitously expressed in all colony types (iPSC, iXEN and Mixed colonies) so further confirmation of colony type via morphology is required. B) When picking colonies to create stable cell lines we observe that the fluorescent reporter becomes restricted to the appropriate cell line. Special consideration must be taken with GATA6-H2B-Venus as not all iXEN cell lines express Venus.

**Figure 3.6. *Nanog*<sup>mCherry/+</sup> reporter creation and testing.**



**Figure 3.6. (cont'd).** A) Schematic of mCherry knock-in shows mCherry inserted in frame at the stop codon of the *Nanog* locus. Magenta represents the stop codon, green the PAM sequence and purple the sgRNA target sequence. Arrows represent the genotyping primer locations. B) Genotyping shows the expected knock in size for mCherry. C) Confocal images taken of E3.75 embryos show the co-localization of NANOG-mCherry with pluripotency marker SOX2 and complementation to primitive endoderm marker SOX17. Yellow arrows indicate marker overlap. Scale bar = 10  $\mu$ m.

**Figure 3.7. *Nanog*<sup>mCherry/+</sup> embryos.**



A and B) Live images of embryos at E4.5 and E7.5 in wild type and knock-in embryos. Yellow arrows indicated embryos without NANOG-mCherry expression and red arrows indicate embryos with NANOG-mCherry expression. C) Genotyping of embryos using *mCherry* primers show that genotyping the knock-in from embryos is possible. Scale bar = 200  $\mu$ m.

**Table 3.1. qPCR primers for detecting endogenous transcripts**

| <b>Gene Target</b> | <b>Forward Sequence (5' to 3')</b> | <b>Reverse Sequence (5' to 3')</b> |
|--------------------|------------------------------------|------------------------------------|
| <i>Oct4</i>        | GTTGGAGAAGGTGGAACCAA               | CCAAGGTGATCCTCTTCTGC               |
| <i>Nanog</i>       | ATGCCTGCAGTTTTTTCATCC              | GAGGCAGGTCTTCAGAGGAA               |
| <i>Sox2</i>        | GCGGAGTGGAAACTTTTGTCC              | CGGGAAGCGTGTACTTATCCTT             |
| <i>Gata6</i>       | ATGCTTGCGGGCTCTATATG               | GGTTTTTCGTTTCCTGGTTTG              |
| <i>Gata4</i>       | CTGGAAGACACCCCAATCTC               | ACAGCGTGGTGGTGGTAGT                |
| <i>Sox7</i>        | GGCCAAGGATGAGAGGAAAC               | TCTGCCTCATCCACATAGGG               |
| <i>Sox17</i>       | CTTTATGGTGTGGGCCAAAG               | GCTTCTCTGCCAAGGTCAAC               |
| <i>ActinB</i>      | CTGAACCCTAAGGCCAACC                | CCAGAGGCATACAGGGACAG               |

**Table 3.2. Genotyping primers**

| <b>Gene Target</b>                         | <b>Forward Sequence (5' to 3')</b> | <b>Reverse Sequence (5' to 3')</b> |
|--|------------------------------------|------------------------------------|
| <i>Nanog<sup>mCherry</sup><sub>y</sub></i> | CCACTAGGGAAAGCCATGCGC<br>ATTT      | GGAAGAAGGAAGGAACCTGGC<br>TTTGC     |
| <i>mCherry</i>                             | AGGACGGCGAGTTCATCTAC               | TGGTGTAGTCCTCGTTGTGG               |
| <i>Pou5f1<sup>eGFP</sup></i>               | CCAAAAGACGGCAATATGGT               | CAAGGCAAGGGAGGTAGACA               |
| <i>Pou5f1 Wild Type</i>                    | TGCCAGACAATGGCTATGAG               | CAAGGCAAGGGAGGTAGACA               |
| <i>Gata6<sup>H2B-Venus</sup></i>           | CCAGGGAGCTCTGAGAAAAAG              | CCTTAGTCACCGCCTTCTTG               |
| <i>Gata6 Wild Type</i>                     | CCAGGGAGCTCTGAGAAAAAG              | GTCAGTGAAGAGCAACAGGT               |
| <i>Gata4<sup>H2B-eGFP</sup></i>            | GTTTCTGCTTTGATGCTGGA               | TGCTCAGGTAGTGGTTGT                 |
| <i>Gata4 Wild Type</i>                     | GTTTCTGCTTTGATGCTGGA               | CGGAGTGGGCACGTAGAC                 |
| <i>Smad4 Wild Type</i>                     | TAAGAACCACAGGGTCAAGC               | TTCCAGGAAAAACAGGGCTA               |

## CHAPTER 4.

OCT4 is expressed in cells fated for extraembryonic endoderm formation during somatic cell reprogramming

Moauro A, Hickey S, Halbisen M, Ralston A\*.

A Moauro completed all experiments, prepared the figures, and wrote the chapter.

S Hickey and M Halbisen completed the scRNA-seq and RNA-seq analysis respectively.

A Ralston helped to write the results and edit the chapter.

Chapter is in preparation for submission to the Journal of Stem Cell Reports.

## **Abstract**

Retroviral reprogramming of mouse fibroblasts using the overexpression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) leads to the formation of two distinct stem cell types in parallel: induced pluripotent stem cells (iPSC) and induced extraembryonic endoderm stem (iXEN) cells. Both iPSC and iXEN cells can proliferate and differentiate in a lineage-appropriate manner, indicating that they are authentic stem cell lines that are equivalent to embryo-derived stem cells. The selection of OSKM as factors for the reprogramming cocktail was founded in embryonic knowledge. Each factor plays a known role in establishing or maintaining pluripotency. When re-examining these embryonic studies, we find that OCT4 plays a dual role in establishing both a pluripotent and an extraembryonic endoderm state. Our findings show that during reprogramming, endogenous OCT4 is expressed in non-iPSC colonies indicating that OCT4 may have a dual role in cell fate establishment. Further evaluation of OCT4 expressing reprogramming cells show that some cells do indeed express an extraembryonic endoderm gene signature. By single cell sorting endogenously expressing OCT4 positive reprogramming cells we find that some are fated for stable iXEN cell line formation. Together our data suggests that like in embryonic development, OCT4 is expressed in iXEN cells and is not a specific marker of pluripotency during reprogramming.

## Introduction

Somatic cell reprogramming using *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) has long-been recognized to produce induced pluripotent stem cells (iPSC) (Takahashi & Yamanaka, 2006). However, an additional, distinct stem cell type has more recently been discovered termed induced extraembryonic endoderm stem (iXEN) cells, which routinely arises in parallel to iPSC during OSKM reprogramming (Nishimura et al., 2017; Parenti et al., 2016).

However, it has yet to be understood why the same four factors produce iXEN alongside iPSCs. To understand this process, we can draw from our knowledge of embryo development in which iPSCs and iXEN model the epiblast (EPI) and primitive endoderm (PE) respectively. During development, the EPI will form the fetus, while the PE contributes to essential extraembryonic tissues and plays an important role in signaling and nutrient exchange (Belaousoff et al., 1998; Stern & Downs, 2012; Stuckey et al., 2011; Thomas & Beddington, 1996).

In the embryo, *Oct4/Pou5f1* is expressed in, and required for, both EPI and PE development. It is not until later in development that *Oct4* becomes restricted to the pluripotent EPI (Aksoy et al., 2013; Frum et al., 2013; Le Bin et al., 2014; Niakan et al., 2010; Wicklow et al., 2014). Since *Oct4* is initially expressed in both EPI and XEN, this raises the possibility that *Oct4* may be acting as a key factor in the formation of both iPSC and iXEN cell lineages during OSKM reprogramming. This possibility is concerning given that during somatic cell reprogramming, the expression of OCT4-eGFP is often used to

monitor or quantify the emergence of iPSCs (Dos Santos et al., 2014; Huangfu, Maehr, et al., 2008; Judson et al., 2009; Shi et al., 2008; X. Y. Zhao et al., 2009). To address the role of *Oct4* in iXEN formation, we sought to evaluate if exogenous *Oct4* is necessary for iXEN formation and if endogenous *Oct4* is expressed in cells fated for iXEN formation.

## **Materials & Methods**

### ***Mouse Strains***

The following alleles were maintained in a CD-1 background: *Nanog*<sup>mCherry</sup>, and *Pou5f1*<sup>tm2Jae</sup> (Lengner et al., 2007). All animal work conformed to the guidelines and regulatory standards of Michigan State University Institutional Animal Care and Use Committee.

### ***Mouse Embryonic Fibroblast (MEF) Preparation***

MEF lines were established from E13.5 embryos. After harvesting, head and viscera were removed and DNA was extracted and genotyped. Individual embryos were dissociated and plated on gelatin in MEF Medium [DMEM, 10% Fetal Bovine Serum (Hyclone), Pen/Strep (10,000 units each)] and grown at 37°C with 5% CO<sub>2</sub> until confluent. Once confluent, MEF line were stored in liquid nitrogen.

### ***Reprogramming***

All MMLV-derived retrovirus and lentivirus was produced by transfecting 293T cells with pCL-ECO and pMXs plasmids or pMD2.g, psPAX2 and FUW plasmids respectively. pMXs plasmids contained either *Oct4*, *Klf4*, *Sox2* or *cMyc* cDNAs (Addgene). FUW

plasmids contained either *rtta*, *Oct4*, *Klf4*, *Sox2* or *cMyc* housed under the control of a tetracycline on promoter (Addgene). Transfected 293T cell supernatant was harvested 48 hours later. mCherry virus was made in conjunction with all viral preps and used to infect CD-1 MEFs to determine viral titers. Viral preps were stored at -80 °C. For retroviral and lentiviral reprogramming (Takahashi and Yamanaka, 2006) MEFs are plated the day before at a density of 50-100 cells/mm<sup>2</sup>. Virus was then added at a MOI of 1 with Polybrene and incubated for 24 hrs. The following day medium was replaced with MEF medium, followed by Reprogramming Medium 1 [DMEM (Invitrogen), 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Fetal bovine serum (FBS; Hyclone), 10 ng/mL Leukemia Inhibitory Factor (LIF)(Millipore Sigma)] on days 2 and 4. Medium was then replaced with Reprogramming Medium 2 [DMEM (Invitrogen), 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Knockout Serum Replacement (KOSR; Invitrogen), 10 ng/mL Leukemia Inhibitory Factor (LIF)] on day 6 and every other day there after until the end of the experiment. When reprogramming with lentivirus, all media was supplemented with 2mg/mL of doxycycline. Cells were grown at 37°C and 5% CO<sub>2</sub>. At the end of reprogramming cells were manually picked and expanded to a passage of 12< to create stable cell lines.

### ***Colony Counting and Lab Images***

At various time points until day 20 of reprogramming, the number of iPSC and iXEN colonies were counted based on morphology. In addition, the presence of fluorescent

markers was detected using a Lumen Prior 200 connected to an inverted Leica microscope at 10X magnification.

### ***Immunofluorescence and Confocal Microscopy***

Cells were plated the day before staining on confocal grade plastic (ibidi) and fixed with 4% formaldehyde (Polysciences) for 10 min, permeabilized with 0.5% Triton X-100 (Millipore Sigma), then blocked in 10% fetal bovine serum (Hyclone) with 0.1% Triton X-100 at 4°C then incubated overnight in primary antibody at 4°C. Primary antibodies used were SOX17 (R&D, AF1924, 1:800 dilution), NANOG (Reprocell, RCAB002P\_F, 1:400 dilution), OCT4 (Santa Cruz, sc-5279, 1:100 dilution) and GATA6 (R&D, AF1700, 1:100 dilution). The next day, cells were washed for 30 min in block, then stained for 1 hr with donkey-anti-rabbit Alexa647 (Jackson ImmunoResearch, 711-606-152, 1:400 dilution), Bovine-anti-goat DyLight488 (Jackson ImmunoResearch, 805-005-180, 1:400 dilution), and donkey-anti-mouse DyLight649 (Jackson ImmunoResearch, 715-495-150, 1:400 dilution) secondary antibody. Following staining, cells were again washed for 30 min in block, then stained for 5 min in DAPI (Sigma, D9542-1MG; 1:1000 dilution). Imaging was performed using an Olympus FluoView FV1000 Confocal Laser Scanning Microscope system with a 40x UPLFL oil lens 1.30 NA or 60x UPLFL oil lens 1.25 NA. Figures were prepared using FIJI, Adobe Photoshop, and Adobe Illustrator.

### ***qPCR and RNA-seq***

RNA was harvested using 1:6 chloroform to Trizol (Invitrogen). 1 µg RNA was reverse transcribed to create cDNA using QuantiTect Reverse Transcription Kit (Qiagen),

following manufacturer instructions. For qPCR, cDNA was amplified using a Lightcycler 480 (Roche) according to manufacturer guidelines. The amplification efficiency of each primer pair (Table 4.2.) was measured by generating a standard curve from appropriate cDNA libraries using mouse extraembryonic endoderm (XEN) cells and embryonic stem (ES) cells. All reactions were completed in quadruplicate.

For RNA-sequencing, cell lines were cultured for at least three passages in XEN media [70% feeder conditioned media (RPMI (Invitrogen) + 20% FBS (Hyclone) + 100  $\mu$ M beta-mercaptoethanol + 2 mM glutamax + 1mM sodium pyruvate + 50  $\mu$ g/mL penicillin/streptomycin) supplemented with 0.025 ng/mL FGF4 (R&D Systems) and 0.001 U/mL Heparin (Sigma-Aldrich)] or 2i [Reprogramming media 2 supplemented with 3 mM GSK3 $\beta$  Inhibitor (Reprocell) + 1 mM Mek1/2 Inhibitor (Stemgent)] before RNA was harvested using the above protocol. Libraries were prepared from 1  $\mu$ g of RNA using Illumina Stranded mRNA Library Preparation kit, and libraries were sequenced using an Illumina NovaSeq 6000, to a depth of 50-90 million with 50bp pair-end end reads per sample.

Before mapping, adapter sequences were removed with Trimmomatic/0.32 (Bolger et al., 2014), and then trimmed raw sequencing reads were aligned to the UCSC mouse reference genome mm10 assembly (<https://genome.ucsc.edu/>) with hisat2/2.1.0 (Kim et al., 2015, 2019; Pertea et al., 2016; Y. Zhang et al., 2021), and were then counted with HTSeq/0.11.2-Python-3.6.6 (Putri et al. 2021, <https://www.python.org/>). Experimental design parameters, including sample size and sequencing depth were based on prior

analysis (Ching et al., 2014). Sequence quality was evaluated before and after read mapping with FastQC/0.11.7-Java-1.8.0\_162 (Wingett et al. 2018) and mapping rates ranged from 85%-99%. Transcripts with low abundance (at least 10 counts per million in at least 3 samples) were removed from the full data set, including data from this study and from (Ichida et al., 2009) prior to generating the MDS plot with the Limma software package version 3.14.4 (Smyth, 2005) in R version 2.15.1 (R Core Team 2012). Additional filtering (at least 10 cpm in at least 3 XEN, 3 Parenti iXEN, 5 SKM iXEN, 5 OSKM iXEN and 5 OCT4-eGFP iXEN samples), MDS plots and differential gene expression analysis was completed with EdgeR 3.24.3 (Y. Chen et al., 2016; McCarthy et al., 2012; Robinson et al., 2009). Volcano plots were generated using the ggplot2 3.3.0 package (Wickham, 2016). Pairwise Spearman correlations (Glasser and Winter, 1961, Spearman, 2010) were calculated for each sample, and the heatmap.2 function of gplots 3.0.1 (Warnes et al., 2016) were used to generate heat maps. All bioinformatic analyses were performed in R/3.3.1 (R Core Team 2018). Gene annotations were performed using MGI (<http://www.informatics.jax.org/batch>) (Table 1), and gene set enrichment analysis was evaluated using ToppGene (<https://toppgene.cchmc.org/priori-tization.jsp>). Raw and processed RNA sequencing files used in this study will be archived and available from the Gene Expression Omnibus database.

### ***scRNA-seq***

Samples were passaged one day before submission to remove dead cells and cell debris. On the day of analysis, cells were harvested and filtered using a 40µm filter. Submitted samples contained <1% of cell clumps and a cell viability of 95%<. Paired-end libraries

were prepared using the 10x Genomics Single Cell 3' V3.1 kit and sequenced on an Illumina NovaSeq 6000.

Base calling was performed using Illumina Real Time Analysis (RTA), and the output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.20.0. After demultiplexing and FastQ conversion, cellrangercount v6.1.1 was used for alignment to the mm10 reference transcriptome (included with cellranger), cell detection and UMI counting.

5,395 cells were detected by cellranger. When analyzing all reprogrammed cells, only cells with <10% of reads coming from mitochondrial genes, >10,000 UMIs, and >1000 detected genes were included for a total of 3,997 cells with a median UMI per cell of ~28000 and a median number of genes expressed per cell of 5239.

The analysis was completed using R v4.1.0 with tools from Seurat v4.1.0. We normalized the UMI counts using SCTransform, regressing the percent of reads coming from mitochondrial genes and the total UMI counts. After converting the list of human cell cycle markers from Tirosh et al, 2015 (included with Seurat), from the human hgnc symbols to mouse mgi symbols using the biomaRt package, we used the CellCycleScoring function to obtain an S-phase and G2M-phase score for each cell. To mitigate the influence of the cell cycle on clustering results, we re-normalized the raw UMI counts, again using SCTransform, this time regressing the percent of reads coming from mitochondrial genes, the total UMI counts, the S-phase score, and G2M-phase score. To define cell clusters,

we used the FindNeighbors function with the first 30 principal components followed by the FindClusters function with a resolution of 0.80. The clusters were visualized using UMAP performed on the first 30 principal components.

When analyzing *Oct4*-positive cells, we filtered for cells with <10% of reads coming from mitochondrial genes, >5,000 UMIs, and >1000 detected genes were included for a total of 1,874 cells with a median UMI per cell of ~30,000 and a median number of genes expressed per cell of 5567. UMI counts were normalized as described above. To define cell clusters, we used the FindNeighbors function with the first 30 principal components followed by the FindClusters function with a resolution of 0.50. The clusters were visualized using UMAP performed on the first 30 principal components.

In both analyses, cluster enriched genes were identified using the FindAllMarkers. Cluster enriched genes included those with a log fold change threshold > .25 expressed in at least 1% of the cells in either the cluster of interest or all other cells. P-values were calculated using the Wilcox rank-sum test and corrected for multiple comparisons with the Bonferroni method. Genes with an adjusted p-value <.01 were considered cluster enriched.

We compare the enriched genes from the *Oct4*-positive clusters of reprogrammed cells with cluster enriched genes from Mohammed et al.'s scRNA-seq map of mouse from peri-implantation to early gastrulation (Table S1, SC3 Cluster specific genes, Enriched genes

by SC3 clustering for each lineage (All genes)) using a hypergeometric test. P-values were corrected for multiple comparisons using the Benjamini-Hochberg Procedure.

### ***Single Cell Sorting***

Reprogramming cells were harvested at different time points throughout reprogramming and placed through a 40 $\mu$ m filter to remove clumps prior to sorting. Sorts were completed on a BD Influx. Gating was accomplished using reprogramming sample and OCT4-eGFP ES cells (Viswanathan et al., 2003) (Fig. S4.1A). Reprogramming cells that were OCT4-eGFP positive and DAPI negative (cell death marker) were sorted into 96-wells containing irradiated fibroblasts for single cell sorting or plated into a 12 well for bulk collection. Samples were then expanded and grown to a passage 12 or greater before analyzing.

### ***Visceral Endoderm Formation***

*In vitro* differentiation followed previously described techniques (Artus et al., 2012; Paca et al., 2012). Culture dishes were treated with Poly-L-ornithine (Sigma) for 30 minutes at room temperature, and then with Laminin (Sigma) at a concentration of 0.15  $\mu$ g/cm<sup>2</sup>. XEN and iXEN cells were plated at a density of ~11,000 cells/cm<sup>2</sup> in N2B27 Medium [50% DMEM-F12 (Invitrogen) + 50% Neural Basal Medium (Invitrogen) + N2 Medium (Invitrogen, 100x) + B27 (Invitrogen, 50x) + Pen/Strep (10,000 units each), beta-mercaptoethanol (55 mM)], and were cultured overnight at 37°C and 5% CO<sub>2</sub>. On days 2, 4, 6 and 8 the culture medium was replaced with fresh N2B27 + 50 ng/ $\mu$ L BMP4 (R&D Systems).

## Results

### ***SKM induces the formation of iXEN cells and is sufficient to activate endogenous OCT4***

Velychok et al., 2019 showed that when using lentivirus for reprogramming, exogenous *Oct4* is not necessary for the formation of iPSCs. SKM alone is sufficient to activate a pluripotency network. These findings provided a unique opportunity to test the requirement for exogenous *Oct4* in the establishment of iXEN. Given that OCT4 is essential in establishing the PE (Frum et al., 2013; Le Bin et al., 2014), we suspected that SKM alone may not be sufficient to create iXEN. Unlike OCT4, the remaining factors (SOX2, KLF4 and c-MYC) play an indirect role in embryonic PE formation (Morgani & Brickman, 2015; Neri et al., 2012; Smith et al., 2010; Wicklow et al., 2014). We hypothesize the indirect role of SKM in PE formation is not sufficient to activate an extraembryonic endoderm network in somatic cells. To test our hypothesis, we set out to reprogram mouse embryonic fibroblasts (MEFs) with SKM lentivirus.

Similar to Velychko et al., 2019, we observed colonies form in SKM treated MEFs at a reduced efficiency of 28% relative to OSKM. However, we saw a sudden increase in colony formation between day 17 to 20 indicating that SKM reprogramming may be delayed relative to OSKM (Fig. 4.1A). Upon morphological examination we observed the emergence of iPSC and iXEN colonies. iPSC colonies are domed and compact with smooth borders where iXEN cell colonies appear flatter and more dispersed, with jagged edges (Parenti et al., 2016). In the OSKM group, we observe 1 iPSC colony for every 3 iXEN colonies which is consistent with the observations in Parenti et al., 2016.

Interestingly, when examining the SKM treated groups we observed an increase in the number of iPSC colonies relative to iXEN colonies. In SKM treated, the beginning colony ratio was similar to OSKM (1 iPSC colony for every 3 iXEN colonies) but grew to a ratio of 1 to 1 by day 20 (Fig. 4.1B, C). This would indicate that the removal of *Oct4* from reprogramming reduces that amount of iXEN formed while favoring the formation of iPSCs.

Surprised by the discovery of iXEN formation in SKM reprogramming, we sought to take a closer look as to why SKM overexpression is sufficient to create iXEN colonies. Looking at the rationale for as to why SKM is sufficient to form iPSC colonies, we see that SKM is able to activate endogenous OCT4 which is a core pluripotency factor (Velychko et al., 2019). We hypothesized that if SKM is sufficient to activate endogenous OCT4, endogenous OCT4 may be expressed in iXEN colonies and acting to activate an extraembryonic endoderm network. To evaluate for the expression of endogenous OCT4, we reprogrammed *Oct4<sup>eGFP/+</sup>* MEF lines. We observed OCT4-eGFP colonies present in both OSKM and SKM reprogramming (Fig. 4.1D). This indicates that SKM and OSKM are sufficient to induce endogenous *Oct4* expression. A closer examination of OCT4-eGFP colonies shows that not only do iPSC colonies express endogenous OCT4 but so do iXEN colonies (Fig. 4.1E). This would support the idea that *Oct4* is a critical transcription factor in iXEN formation.

We next wanted to evaluate for differences in SKM and OSKM derived iXEN to determine if these cells were of similar quality and potential relative to embryo-derived stem cells.

To do so we expanded colonies to create stable cell lines and compared gene expression, morphology and differentiation potential to embryo-derived PE lines termed, extraembryonic endoderm stem (XEN) cells. Embryo-derived XEN cells are a self-renewing, and multipotent stem cell line that can be derived from the mouse blastocyst (Kunath et al., 2005). As an internal technical control, we also created stable iPSC lines from SKM and OSKM reprogramming and evaluated the gene expression and morphology relative to embryo-derived EPI lines, termed embryonic stem (ES) cells (Evans & Kaufman, 1981; Martin, 1981). We saw that both SKM and OSKM derived iXEN and iPSC lines expressed the appropriate markers and were evenly distributed throughout cell lines relative to embryonic derived stem cell controls (Fig. 4.1F, G, H and S4.1B). We were unable to detect a difference in morphology and gene expression between cell lines indicating that all induced lines are comparable to embryo-derived stem cells.

We then took a closer look at gene expression to uncover subtle transcriptional differences using RNA-seq. We observed a similar gene expression between OSKM and SKM derived iXEN lines that are analogous to XEN and MMLV-derived iXEN lines from Parenti et al., 2016 but distinct from starting fibroblasts (Fig. 4.2. C). A direct comparison of OSKM and SKM derived iXEN lines reveals only four differentially expressed genes (*Psca*, *Clic6*, *Rpl21*, and pseudogene 1 *Rps26*). To our knowledge, these genes are of no known significance to iXEN or XEN cells (Fig. 4.2. D). OSKM and SKM derived iXEN as transcriptionally indistinguishable.

Lastly, we evaluated the developmental potential of SKM and OSKM derived iXEN cell lines using an *in vitro* XEN cell differentiation assay (Artus et al., 2012; Paca et al., 2012). XEN cells treated with BMP4 undergo a mesenchymal to epithelial transition (MET) and elevate expression of markers of visceral endoderm (Bielinska et al., 1999). Both SKM and OSKM derived iXEN underwent MET (Fig. 4.2A) and upregulate visceral endoderm gene expression (Fig. 4.2B). We were unable to detect a difference in differentiation ability between induced lines and embryo-derived lines. This indicates that both SKM and OSKM derived iXEN possess the ability to differentiate in a biologically relevant manner and are comparable to embryo-derived XEN cells.

***A population of cells expressing Oct4 display an embryonic primitive endoderm gene signature during reprogramming***

Intrigued by the discovery of endogenous OCT4 expressing iXEN colonies, we wanted to evaluate OCT4 expression at a single cell level. This was done to determine if indeed OCT4 was expressed in naïve iXEN cells and eliminate the possibility that we had misidentifying colonies based on morphology alone. We performed scRNA-seq on day 17 of OSKM reprogramming. This time point was selected as this day consistently displays a high level of *Oct4* expression. In addition, this time point is near the end of reprogramming which provides a larger population of cells that have begun to active a pluripotent or extraembryonic endoderm network. Given that we did not see a difference in the quality of iXEN cells in SKM compared to OSKM reprogramming and observed that both treatments induced endogenous OCT4 expression, we choose to only analyze OSKM reprogramming cells. One key difference between OSKM and SKM

reprogramming is the higher reprogramming efficiency in OSKM which is critical advantage when performing single cell analysis as this provides a greater capture of reprogramming events.

OSKM reprogramming has been shown to produce a heterogeneous population of cells between partially reprogrammed cells to true induced stem cell populations (Cacchiarelli et al., 2015; He et al., 2020; X. Liu et al., 2020; Nishimura et al., 2017; Parenti et al., 2016). Consistent with this knowledge we observed several distinct clusters (Fig. 4.3A). We set out to identify which cluster best represented our putative populations of iPSC and iXEN cells using common lineage markers. Interestingly, common markers such as *Nanog* and *Gata6* were minimally expressed in this late stage of reprogramming, suggesting that our cell populations still had not fully activated their pluripotent or extraembryonic endoderm network. We were able to identify other markers of pluripotency and extraembryonic endoderm which include *Fgf4* and *Pdgrf $\alpha$* , respectively (Lokken & Ralston, 2016). Based on the high level of expression of *Fgf4*, cluster 4 represents putative iPSCs. The high level of *Pdgrf $\alpha$*  would indicate that cluster 6 are putative iXEN cells. Interestingly, many of the clusters including clusters 4 and 6 expressed *Oct4* indicating that *Oct4* is associated with multiple cell fates (Fig. 4.3B). The expression of *Oct4* in putative iPSC and iXEN clusters is in agreement with the understanding of *Oct4* in embryonic EPI and PE development.

Based on the observation of widespread *Oct4* expression, we then removed all *Oct4* expressing cells from our original dataset and re-clustered them to better identify the

types of *Oct4* expressing cells (Fig. 4.3C). In order to identify the cell types in each cluster, we chose to focus on a time in embryo development in which *Oct4* is highly active which is the early embryo (Patra, 2020). We mapped our *Oct4* expressing cells to published data from Mohammed et al., 2017 which provides an in-depth scRNA-seq analysis looking at embryo development from E3.5 to E6.5 (Fig. 4.3D). This time span is critical as it provides information on gene expression and cell identity from before the formation of EPI and PE to after the differentiation of EPI and PE. Unsurprisingly, several clusters aligned to the epiblast of the early embryo but what was surprising is that different cells clustered to different developmental stages of the EPI. We observed cluster overall of E4.5 EPI in cluster 4 and E5.5 and E6.5 in cluster 3. In addition, we saw the expression of primitive streak (PS) genes throughout several of the clusters. This PS signature has been a common finding in several OSKM reprogramming studies and has been reported to only be transiently expressed throughout the reprogramming process (Cacchiarelli et al., 2015; Raab et al., 2017; Takahashi et al., 2014). Our data indicates that no single cluster best represented a PS gene signature in our *Oct4* expressing cells and that no true PS population exists. It is also possible that this PS signature may be a common feature among reprogramming cells regardless of cell fate. We also observe a strong correlation of PE gene expression in cluster 2 which supports the idea that cells expressing *Oct4* do indeed express primitive endoderm genes. This supports the idea that OCT4 plays a role in activating the primitive gene network. Another interesting finding is the expression of inner cell mass (ICM) genes in cluster 4. This finding is significant as ICM cells differentiate to give rise to both EPI and PE cells in the embryo. If this is a true

ICM-like population, it is possible that iPSC and iXEN are arising from a progenitor state in cluster 4.

### ***Endogenous Oct4 is expressed within iXEN cell colonies during reprogramming***

Although our scRNA-seq data reveals that *Oct4* and primitive endoderm genes are expressed within the same cells, it does not tell us whether exogenous OCT4 or endogenous OCT4 is playing a role in primitive endoderm gene activation. We wanted to further evaluate endogenous OCT4 to determine if cells expressing endogenous *Oct4* are truly fated for iXEN formation. We therefore hypothesized that endogenous OCT4 expression labels both iPSC and iXEN cells during somatic cell reprogramming. To test this hypothesis, we overexpressed OSKM in *Oct4<sup>eGFP/+</sup>* MEFs using a retroviral delivery system.

Consistent with our lentiviral OSKM results we observed iPSC and iXEN colonies. We detected OCT4-eGFP expression as early as day 8 of reprogramming. Interestingly, we observed that there were more eGFP expressing colonies than there were observed iPSC morphology colonies (Fig. 4.4A). Further evaluation of the OCT4-eGFP positive colonies revealed that eGFP is present in non-iPSC colonies including iXEN colonies or a Mixed morphology colony that displays characteristics of both iPSC and iXEN colonies (Fig. 4.4B). This indicates that endogenous OCT4 is not expressed in only pluripotent cells, but cells fated for different identities. It is currently unknown as to what these Mixed colonies are and if they have the potential for both iXEN and iPSC fates or if the Mixed

colony represent a new cell type altogether. Either scenario suggests a novel cell type that is different than a pure pluripotent cell.

Next, we wanted to make sure that we were not misidentifying endogenous OCT4-eGFP colonies based on morphology and that indeed endogenous OCT4-eGFP cells are fated for different cell fates. To rule out the possibility of mistaken identity, we used fluorescently activated cell sorting to select OCT4-eGFP-expressing cells on days 11-17 of reprogramming (n=12 samples and 4 cell lines). eGFP-positive cells were pooled into single wells for each time point, and then allowed to proliferate for several passages in order to form stable cell types and turn on their fated network (Polo et al., 2010). The resulting cell lines exhibited several cell morphologies, consistent with the notion that OCT4 is not a specific marker of pluripotency. This is an expected result as previous findings have shown that *Oct4* labels partially reprogrammed cells as well as pluripotent cells (Buganim, Faddah, Cheng, Itskovich, Markoulaki, Ganz, Klemm, Van Oudenaarden, et al., 2012; Chan et al., 2009).

Notably, we also observed XEN cell morphology among roughly two-thirds of the OCT4-eGFP-derived cell lines (Fig. 4.4C, D), consistent with the notion that *Oct4* is needed for the formation of extraembryonic endoderm (Frum et al., 2013; Le Bin et al., 2014). We then evaluated the expression levels of markers of ES and XEN cells within these cell lines. As expected, *Oct4* was highly expressed in all cell lines (Fig. 4.4E). Expression of the pluripotency markers *Nanog* and *Sox2* appeared reduced in OCT4-eGFP-derived cell lines, relative to ES cells, consistent with the presence of non-pluripotent cell types within

the OCT4-GFP-derived cell lines. Additional expression of several endodermal markers were elevated within the OCT4-eGFP-derived cell lines indicating the presence of iXEN cells. These observations are in agreement with previous findings that endogenous *Oct4* expression is not specific to pluripotent cell colonies (Buganim, Faddah, Cheng, Itskovich, Markoulaki, Ganz, Klemm, Van Oudenaarden, et al., 2012; Chan et al., 2009), and also supports the novel hypothesis that endogenous OCT4 is associated with the formation of iXEN cells.

### ***Endogenous Oct4 is temporarily expressed within iXEN cells***

Because we had observed expression of *Oct4* within diverse cell types during reprogramming, we next evaluated the developmental potential of OCT-eGFP-expressing cells clonally to determine what population of OCT4-eGFP positive cells are fated for iXEN development. We sorted single OCT4-eGFP-positive cells (n=6 cell lines, n=2880 sorted cells) into separate wells, and then allowed time for the single cells to proliferate as clonal cell lines. During this proliferation we observed the growth of single cells into colonies. Some of these colonies produced outgrowths at the colony base that resembled a more mesenchymal cell type providing the first sign of possible iXEN formation (Fig. 4.5A).

After the cell lines had proliferated for multiple passages and had formed stable cell lines, we observed a much greater degree of homogeneous morphology than we had after pooling OCT4-eGFP-positive cells (Fig. 4.5B, C). We did still observe cell lines that had one or more cell type present. Of the observed morphologies, we detected iPSC and

iXEN along with other cell morphologies. Other commonly observed morphologies include rounded cells or cells displaying a cobblestone appearance. We presume that the other cell types detected in clonally derived cell lines could be due to random differentiation or partially pluripotent OCT4-eGFP-expressing cells. More importantly, however, among the clonally derived OCT4-eGFP-derived cell lines that maintained a stable, singular morphology, around half of these exhibited XEN cell morphology (Fig. 4.5D). These observations indicate that endogenous OCT4 is expressed within iXEN cells. Most of these putative iXEN cell lines did not, however, maintain expression of OCT4-eGFP (Fig. 4.5E) after prolonged passaging ( $p > 12$ ), suggesting that endogenous OCT4 is expressed transiently and/or dynamically, during iXEN cell formation. This may also explain why iXEN colonies do not always express OCT4-eGFP and when they do, it is expressed in patches rather than throughout the entire colony.

To further evaluate our clonally-derived OCT4-eGFP-derived iXEN cell lines, we sought to compare the expression of extraembryonic endoderm genes to embryo-derived XEN cell lines. We observed low to no expression levels of pluripotency genes and higher expression levels of endodermal genes in the clonally-derived OCT4-eGFP-derived iXEN cell lines constant with XEN cell lines (Fig. 4.6A, B). Moreover, when evaluated at the single cell level using confocal microscopy, the expression of endodermal genes appeared to be homogenous among clonally derived iXEN cell lines (Fig. 4.6A, B). A closer look into subtle transcriptional differences using RNA-seq reveals that our OCT4-eGFP-derived iXEN cell lines do not express MEF genes, but they do not cluster as closely to embryonic-derived XEN cell lines as our previously established iXEN cell lines

(Fig. 4.6C) (Parenti et al., 2016). Further investigation into the differential gene expression between OCT4-eGFP-derived iXEN cell lines and embryonic-derived XEN cell lines reveals that important XEN marker genes are expressed at a similar level (Fig.4.6D). Toppgene analysis of the top 10 upregulated GO term pathways include the upregulation of terms involving translational processes in OCT4-eGFP iXEN lines. Interestingly, Toppgene analysis revealed the downregulation of extracellular matrix proteins in OCT4-eGFP iXEN lines (Table 4.1). Since single XEN cells are mesenchymal, it is common for these cells to express an array of collagen genes (Kunath et al., 2005). We suspect that some of these differentially expressed genes could be due to subtle differences in growth media such as variations in fetal bovine serum. Alternatively, establishing iXEN colonies from single cells compared to iXEN colonies could have affected the transcriptome between samples.

Finally, we evaluated the developmental potential of the clonal OCT4-eGFP-derived iXEN cell lines using the *in vitro* XEN cell differentiation assay (Artus et al., 2012; Paca et al., 2012). This treatment induced MET (Fig. 4.6E) and upregulation of visceral endoderm gene expression (Fig. 4.6F) within the clonally-derived OCT4-eGFP-derived iXEN cell lines, confirming their developmental potential. Coupled with the self-renewal capacity of these cell lines, we confirm that clonally-derived OCT4-eGFP-derived iXEN cells function as bona fide stem cells.

## Discussion

In a hunt to better understand the role of exogenous *Oct4* in iXEN formation we discovered that SKM and OSKM reprogramming produced iXEN and iPSCs. Comparison of iXEN lines from each reprogramming method showed no difference in markers or differentiation potential. The discovery of iXEN in SKM reprogramming reveals that SKM, like OSKM, is not a specific inducer of pluripotency. However, it is well known that OSKM reprogramming induces a wide array of gene expression during reprogramming which is thought to be due to partial reprogrammed or intermediate cell states (Cacchiarelli et al., 2015; González & Huangfu, 2016; Meissner et al., 2007; Mikkelsen et al., 2008; Raab et al., 2017; Soufi et al., 2012; Takahashi et al., 2014; Xing et al., 2020). More recently, it was discovered that the heterogeneity produced by OSKM reprogramming could be explained by the formation of multiple stable induced stem cells (Castel et al., 2020; He et al., 2020; X. Liu et al., 2020; Nishimura et al., 2017; Parenti et al., 2016). It has yet to be determined to what extent SKM reprogramming follows the same path as OSKM reprogramming. It may be possible that although SKM is not a specific inducer of pluripotency, it may induce a smaller array of unwanted genes.

Interestingly, for iPSCs to form, somatic cells must activate a pluripotency network. A critical component of the pluripotency network is OCT4 (X. Chen et al., 2008; Macarthur et al., 2012; Nichols et al., 1998; Niwa et al., 2000, 2009; Orkin et al., 2008). This must mean that SKM is capable of activating endogenous OCT4. One would expect if OCT4 alone controlled pluripotency formation, that all endogenous OCT4 is expressed in colonies fated for pluripotency. However, SKM and OSKM reprogramming activate the

expression of endogenous OCT4 in iXEN colonies. Using traditional retroviral OSKM reprogramming endogenous OCT4 is activated in iXEN colonies as early as day 11 post viral infection. To ensure that our OCT4-eGFP expressing iXEN colonies were truly iXEN, we single cell sorted reprogramming cells that expressed OCT4-eGFP and allowed the single cells to form clonally-derived stable cell lines. These lines revealed an array of morphologies, but we were able to identify iXEN cells, which was further confirmed through gene expression and differentiation assays showing that indeed endogenous OCT4 is expressed in iXEN cells.

Further analysis into the other observed cell morphologies is needed to identify what role OCT4 is playing in the formation of additional cell types. We suspect that some could be partially reprogrammed cells, but we suspect a large amount are due to differentiation of iPSC or iXEN cells. The cobblestone morphology observed in Fig 4.5B resembles visceral endoderm, which is formed from differentiated iXEN. In addition, several cell lines displayed myocyte morphology by passage 2. This cell type was easily identified by the spindle shaped cells and irregular contractions which we suspect is due to the differentiation of iPSCs (Fig. S4.2D). Additionally, single cell sorting places a lot of strain on cells between pressure, temperature, and media changes. Future studies can be done to avoid the possibility of differentiation by growing cells in media that contain specific small molecules that are specific for the maintenance and growth of only iXEN or iPSC cells.

Once we were able to create stable OCT4-eGFP iXEN cell lines, we discovered that some lines continued to express OCT4 late into passaging. This could mean that OCT4 is indeed an important factor in establishing a primitive endoderm network just like in the embryo (Frum et al., 2013; Le Bin et al., 2014). Alternatively, this OCT4 expression could have been enhanced due to the use of LIF in the growth media, but LIF is a commonly used factor in establishing embryonic XEN cells (Morgani & Brickman, 2015; Niakan et al., 2013; Stirparo et al., 2021). This unique discovery of OCT4 expression in our OCT4-eGFP derived iXEN cells lines provides the possibility that iXEN cells may pass through a E3.75 PE state rather than reprogramming directly to an E4.5 PE state. In embryo development, OCT4 is necessary for the formation of PE at E3.75 but OCT4 is usually repressed in the E4.5 PE. Traditionally, it is not until the E4.5 PE stage that embryo-derived XEN cells have been able to be stably captured. As of late, several papers have tried to capture the E3.75 PE state *ex vivo*. If our OCT4-eGFP expressing iXEN cells can be stably captured, this could provide another *in vitro* model of OCT4-expressing PE-like stem cells to serve alongside primitive extraembryonic endoderm stem cells (pXEN) and primitive endoderm stem cells (PrESCs) (Ohinata et al., 2022; Zhong et al., 2018). The establishment of these models are critical as the exact role of OCT4 in embryonic PE development has yet to be determined. The use of tools such as CUT&RUN or CHIP-seq will hopefully one day elucidate what genes OCT4 promotes in PE-like cells.

Lastly, our scRNA-seq revealed that not only is OCT4 expressed in pluripotent-like cells but a wide array of cell identities that cluster to early embryonic cells such as the inner cell mass, different stages of epiblast, primitive streak, and primitive endoderm. It is

possible that OSKM and OCT4 may not be specific for pluripotency but more broadly early embryo formation. This idea is further supported by the previous discovery of induced trophoctoderm stem cells which are embryonically similar to the trophoctoderm and gives rise to the placenta (Castel et al., 2020; X. Liu et al., 2020; Nishimura et al., 2017; Parenti et al., 2016). The current role as to how OSKM is inducing the trophoctoderm lineage is unknown as in normal embryo development OCT4 acts to repress trophoctoderm formation (Nichols et al., 1998). Further investigation into the necessity of OCT4 in reprogramming using OCT4 knockdowns and knockouts will hopefully help elucidate the potentially broad role of OCT4.

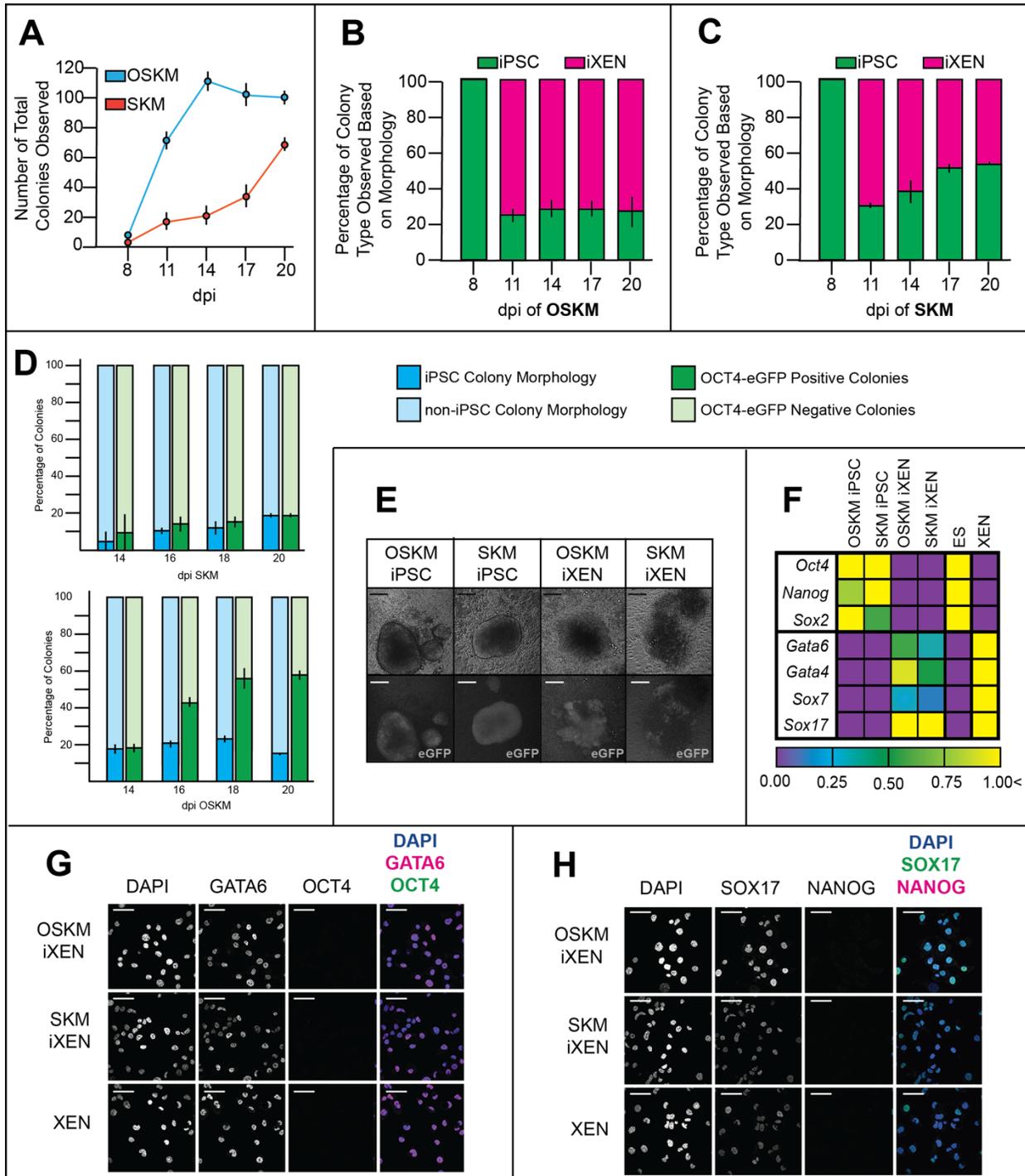
In summary, we have shown without a doubt that OCT4 is no longer a reliable marker of pluripotency during reprogramming. Instead OCT4 is playing a role in extraembryonic endoderm and pluripotency cell fates which is consistent with early embryonic development. We hope that further investigation into this complex system will uncover the exact mechanism capable of inducing iPSC or iXEN fates.

### **Acknowledgements**

We would like to thank Shinya Yamanaka, Dider Trono, Tarjei Mikkelsen and Rudolph Jaenisch for their development of the Addgene plasmids. Lastly, we would like to thank the Michigan State University Flow Cytometry Core for their help in single cell sorting and the Michigan State University Research Technology and Support Facility Genomics Core for their help in RNA-seq and scRNA-seq.

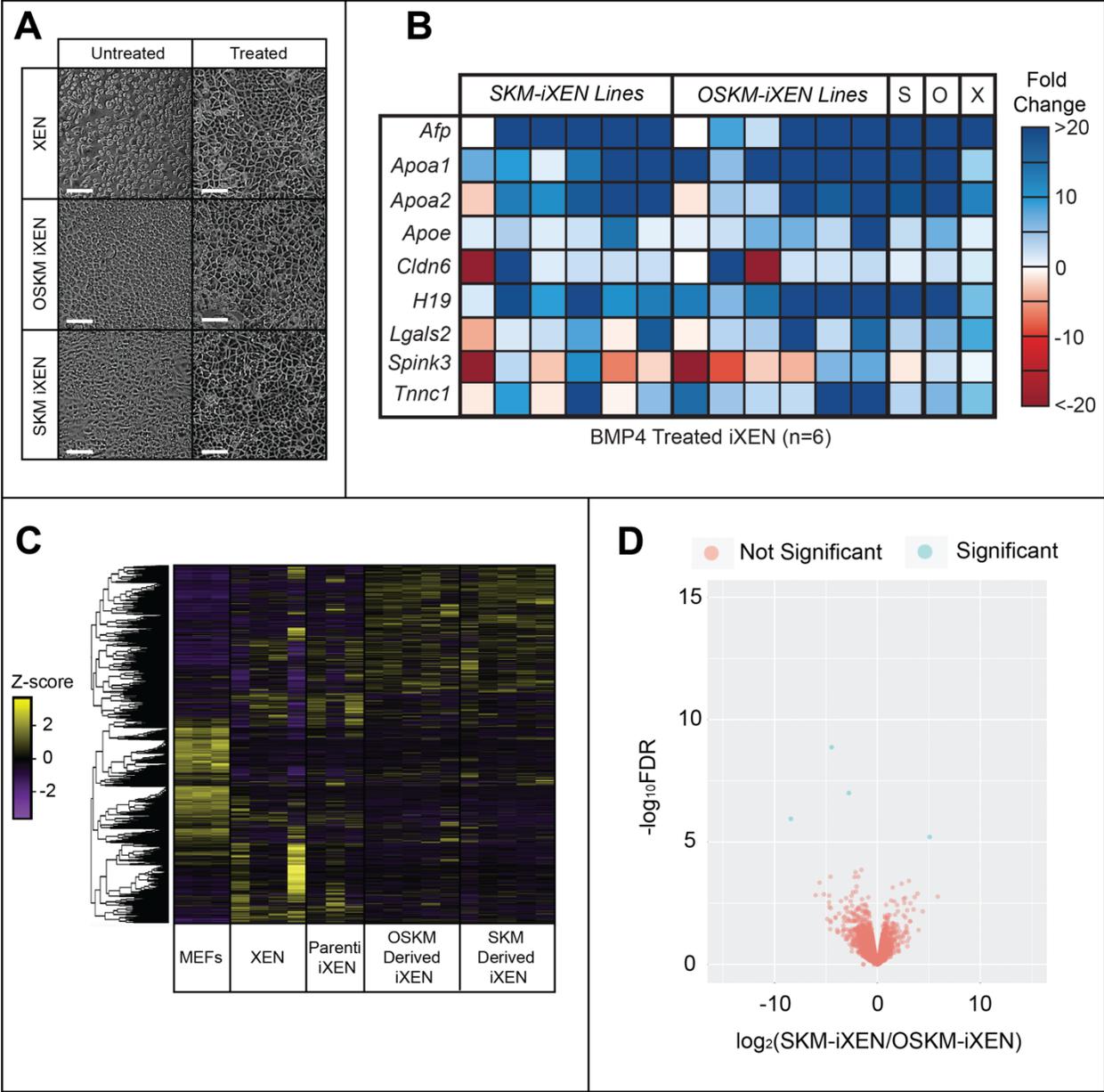
## APPENDIX

**Figure 4.1. OSKM and SKM reprogramming produces iXEN colonies that can be expanded to create stable cell lines.**



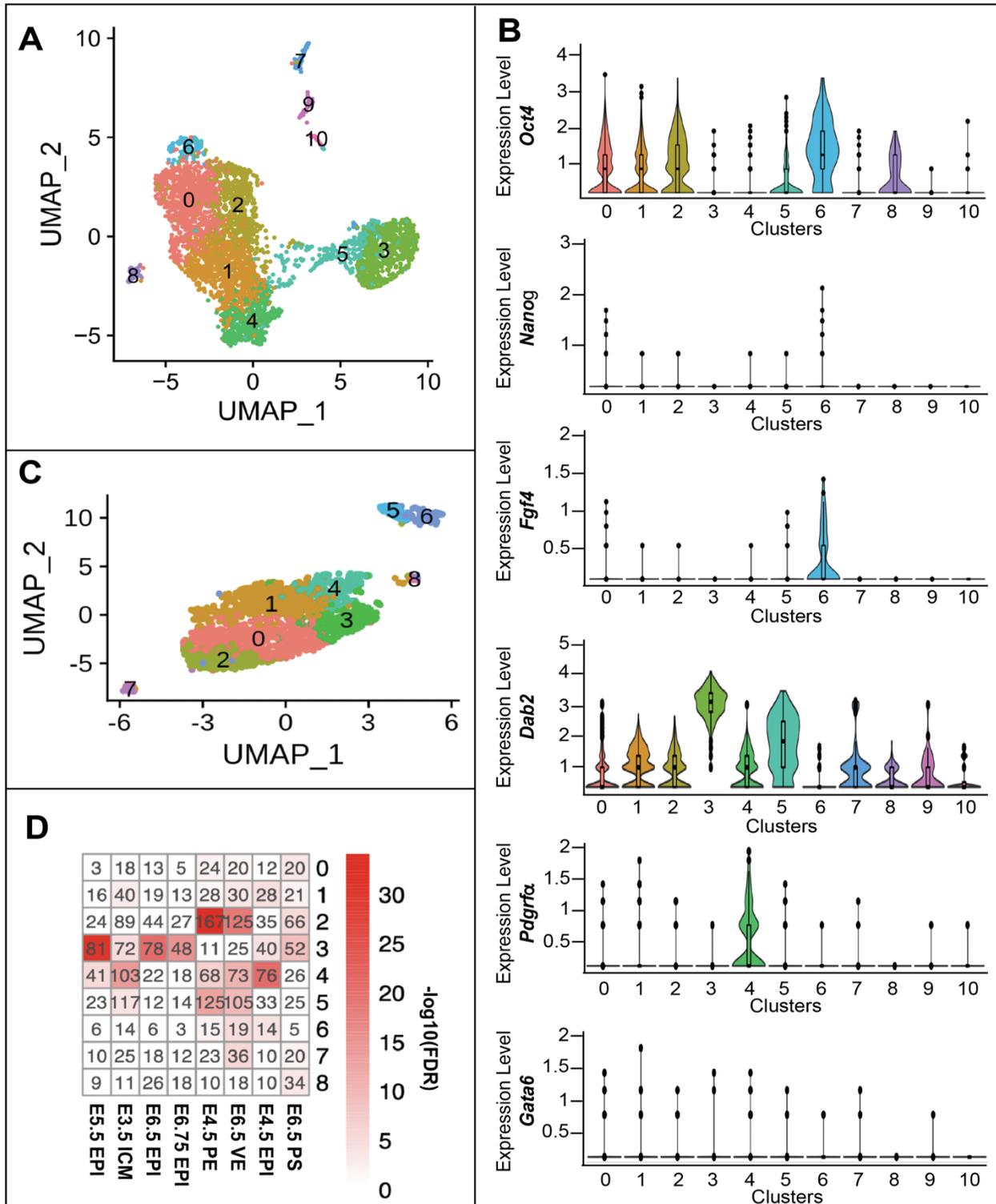
**Figure 4.1. (cont'd).** A) SKM produces colonies at a lower efficiency than SKM. B) OSKM produces iPSC to iXEN colonies at a ratio of 1:3 stably by day 11. C) SKM reprogramming produces iPSC colonies at an increasing rate. D) OCT4-eGFP is expressing in both SKM and OSKM reprogramming. SKM reprogramming produces less OCT4-eGFP positive colonies and produces positive colonies at a rate similar to iPSCs. Scale bar = 100  $\mu$ m. E) Endogenous OCT4 is expressed in both iXEN and iPSC colonies in SKM and OSKM reprogramming. F) Expansion of colonies to create stable iXEN and iPSC lines shows that each line expresses the correct markers and that. G and H) confocal imaging of iXEN lines reveals that primitive endoderm markers are expressed evenly among OSKM and SKM iXEN lines and are expressed at a similar level relative to control embryonic-XEN. scale bar = 50  $\mu$ m. n > 3.

**Figure 4.2. SKM and OSKM derived iXEN produce visceral endoderm and are transcriptionally indistinguishable.**



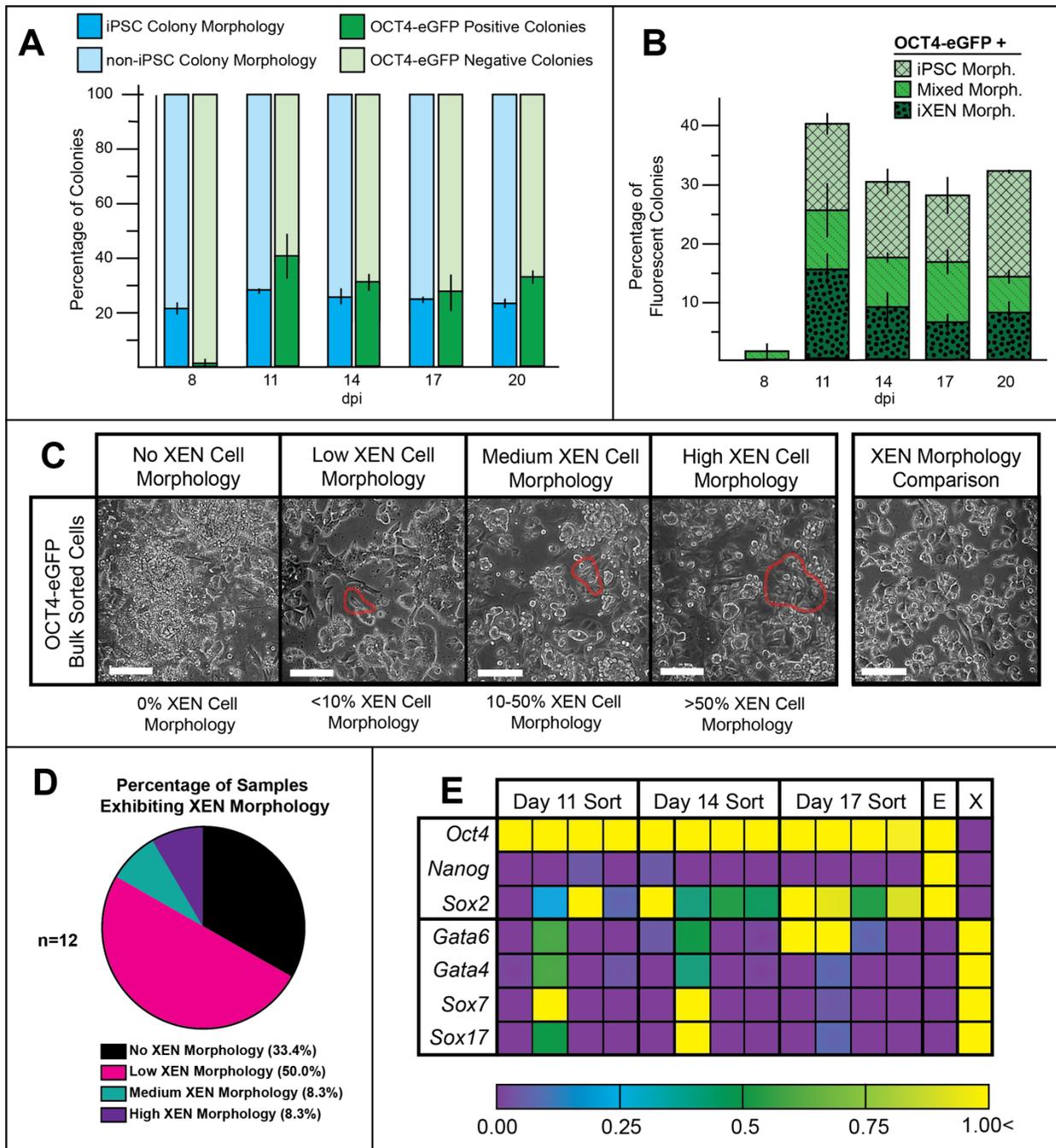
**Figure 4.2. (cont'd).** A) iXEN and XEN cells exposed to BMP4 show morphological changes consistent with visceral endoderm morphology. Scale bar = 100  $\mu$ m and arrows point to visceral endoderm morphology. B) Upregulation of visceral endoderm genes relative to untreated controls using qPCR. SKM iXEN, OSKM iXEN and embryo-derived XEN were treated with BMP4 and compared to untreated controls. C) Heatmap generated from RNA-seq data demonstrates OSKM and SKM derived iXEN lines are similar to each other and are similar to published XEN RNA-seq and Parenti et al., 2016 iXEN but distinguishable from starting fibroblasts. D) Volcano plot comparing differentially expressed genes between OSKM and SKM derived iXEN lines show that lines are undistinguishable with only four genes differentially expressed (n=5).

**Figure 4.3. *Oct4* is expressed alongside early embryo development genes during OSKM reprogramming.**



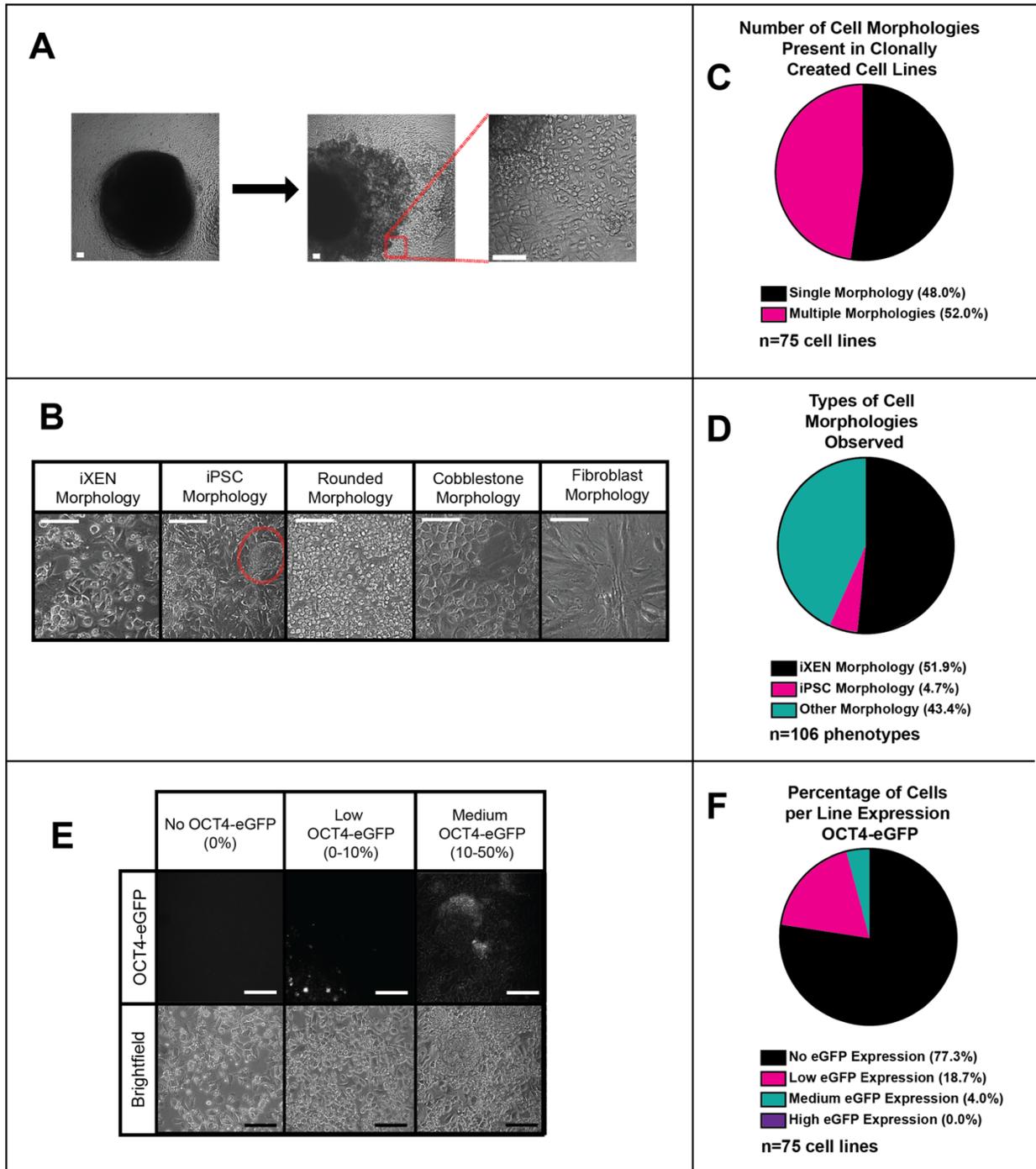
**Figure 4.3. (cont'd).** A) scRNA-seq completed on day 17 of OSKM reprogramming cells reveals several clusters of cells. B) Several clusters are expressing *Oct4*, but expected genes such as *Gata6* and *Nanog* were limited in expression. Expression level determined by log corrected UMI counts. C and D) Clustering of *Oct4* expressing cells and aligning them to early embryo scRNA-seq data from Mohammed et al., 2017 shows that several clusters align to different cell types of the early embryo. In particular, cluster 2 highly expresses PE genes identified from Mohammed et al., 2017. EPI = epiblast, ICM = inner cell mass, PE = primitive endoderm, VE = visceral endoderm, PS = primitive streak.

**Figure 4.4. OCT4-eGFP is expressed in reprogramming somatic cells fated for different identities.**



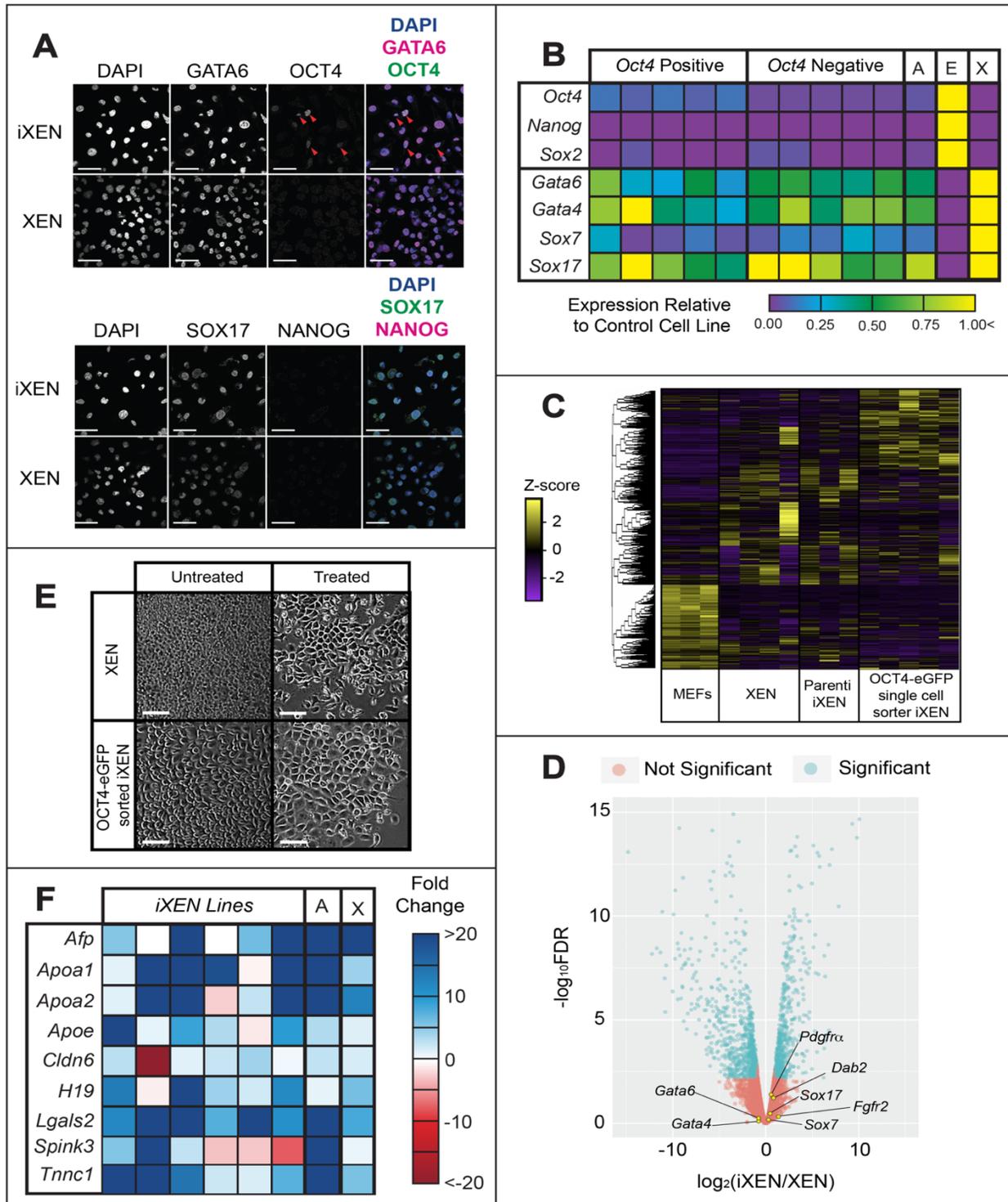
**Figure 4.4. (cont'd).** A) OCT4-eGFP positive colonies do not agree with iPSC colony morphology counts throughout reprogramming suggesting that OCT4 is not expressed in only iPSC colonies. n= 3. B) OCT4-eGFP expressing colonies display iPSC, iXEN or Mixed (colonies with both iPSC and iXEN characteristics) morphology colonies. n = 3. C and D) OCT4-eGFP positive reprogramming cells were single cell sorted using FACs and expanded to create stable cell lines. Scale bar = 100  $\mu$ m. Cell lines express heterogenous cell types with some lines containing a higher proportion of iXEN morphology than others. E) qPCR reveals that some of established cell lines express primitive endoderm/XEN markers. E = ESC, X = XEN, n = 12, gene expression is relative to positive controls.

**Figure 4.5. Single cell sorting reveals that OCT4-eGFP expressing reprogramming cells are fated for non-pluripotent cell types.**



**Figure 4.5. (cont'd).** A) Formation of a typical iXEN colony after single cell sorting shows the formation of a large colony proceeded by the expansion of outgrowths with the appearance of XEN morphology. B) Of the cell lines that were created, a majority displayed a homogenous morphology indicating the presence of only one cell type. C and D) Several morphologies were observed in stably created cell lines. Of the morphologies present a majority displayed iXEN morphology and a minority displayed iPS morphology. E and F) By end of stable cell line formation, low to medium OCT4-eGFP could be observed. Scale bar = 100  $\mu\text{m}$ .

**Figure 4.6. OCT4-eGFP single cell sorted iXEN lines express the same extraembryonic endoderm markers and differentiation potential as embryo-derived XEN.**



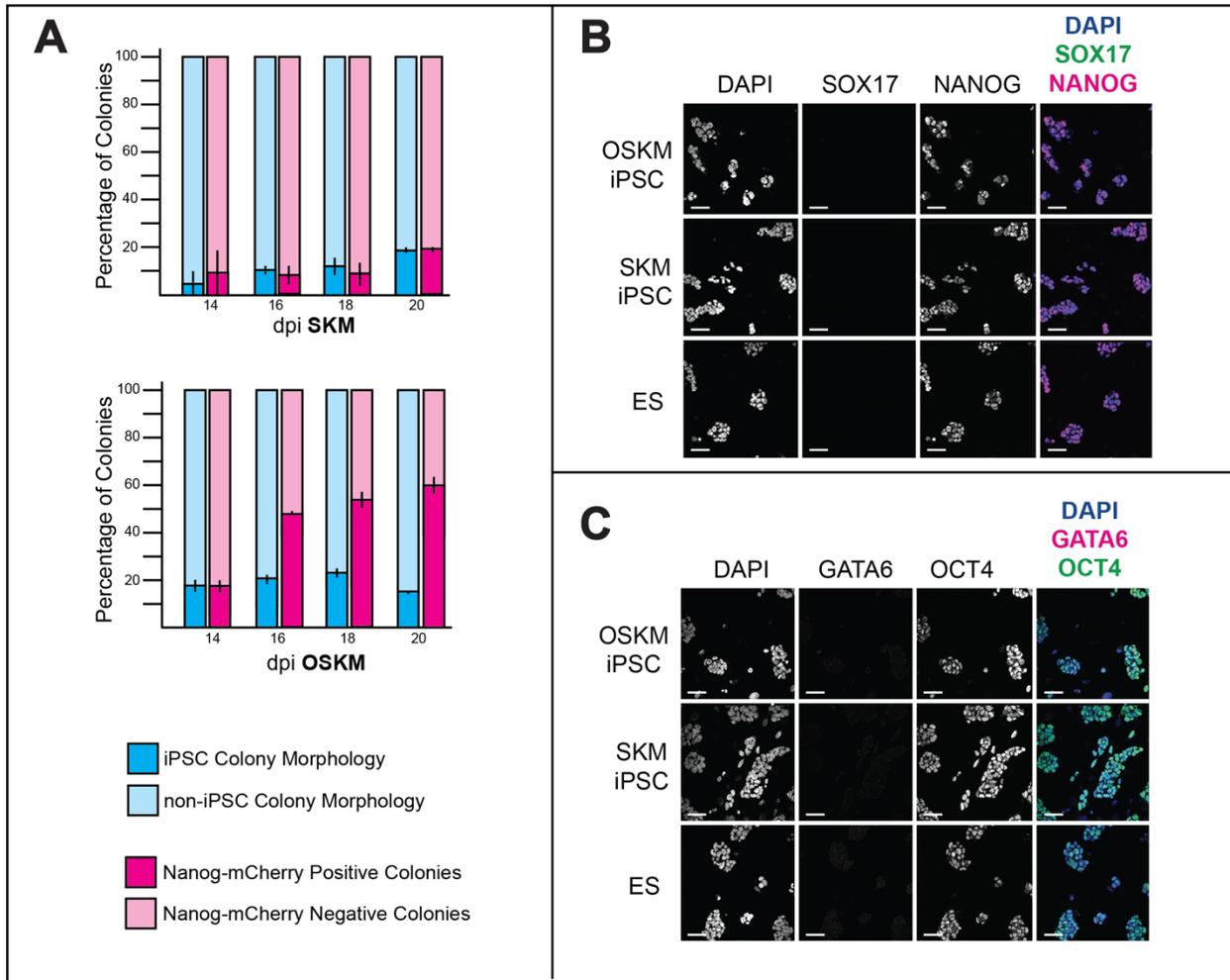
**Figure 4.6. (cont'd).** A) Confocal imaging of putative iXEN lines looking at the spatial distribution of primitive endoderm and pluripotent markers. Red arrows show OCT4 positive cells, images taken at 60x, Scale bar = 50 $\mu$ m. B) qPCR data reveal expression of primitive endoderm in putative iXEN cell lines with some expressing low levels of OCT4. Potential iXEN lines were grown in Reprogramming Medium 1, p>12, A = average of iXEN lines, E = Embryonic Stem Cells (ESCs) and X = XEN. C) XEN and iXEN lines do not cluster with starting MEFs however, OCT4-eGFP single cell sorted iXEN lines more closely resemble themselves than Parenti et al 2016 iXEN or embryonic derived XEN. Heatmap generated from RNA-seq. D) Volcano plot comparing differentially expressed genes between OCT4-eGFP single cell sorted iXEN lines and embryonic-derived XEN. E and F) Differentiation of OCT4-eGFP single cell sorted iXEN lines and XEN with exposure to BMP4 for 9 days reveals the expression of Visceral Endoderm morphology and gene expression. A = Average of OCT4-eGFP single cell sorted iXEN lines, X = XEN control, scale bar = 100  $\mu$ m.

**Table 4.1. Top GO term pathways upregulated in OCT4-eGFP iXEN and XEN**

| <b>Top 10 GO Terms in OCT4-eGFP iXEN</b>                               | <b>P-value</b> | <b>Query Match</b> |
|--|----------------|--------------------|
| RNA binding  | 2.06E-27       | 6.86%              |
| translational initiation   | 4.14E-28       | 21.61%             |
| nuclear-transcribed mRNA catabolic process,<br>nonsense-mediated decay | 1.91E-25       | 27.50%             |
| mRNA metabolic process   | 9.14E-25       | 9.06%              |
| SRP-dependent cotranslational protein targeting to<br>membrane         | 8.85E-24       | 28.57%             |
| cotranslational protein targeting to membrane                          | 4.02E-23       | 27.27%             |
| ribonucleoprotein complex  | 1.19E-29       | 10.68%             |
| cytosolic ribosome   | 5.79E-27       | 27.56%             |
| ribosomal subunit  | 4.57E-26       | 20.00%             |
| ribosome   | 2.17E-25       | 17.10%             |

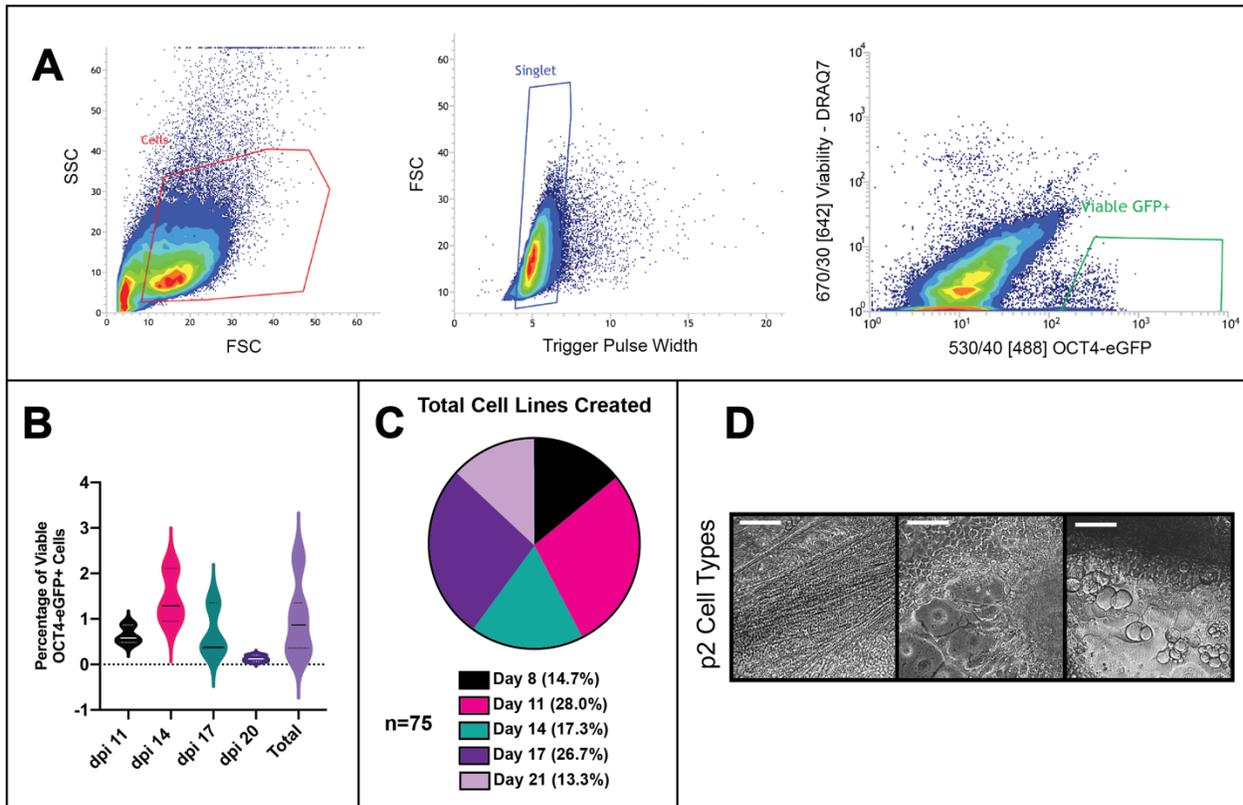
| <b>Top 10 GO Terms in XEN</b>                 | <b>P-value</b> | <b>Query Match</b> |
|---|----------------|--------------------|
| extracellular matrix structural constituent   | 2.68E-24       | 25.41%             |
| extracellular matrix organization             | 1.65E-21       | 15.31%             |
| extracellular structure organization          | 1.87E-21       | 15.28%             |
| external encapsulating structure organization | 2.42E-21       | 15.21%             |
| collagen fibril organization                  | 1.10E-15       | 34.92%             |
| carbohydrate derivative metabolism            | 1.08E-14       | 8.47%              |
| extracellular matrix                          | 8.72E-32       | 15.48%             |
| external encapsulating structure              | 1.13E-31       | 15.43%             |
| collagen-containing extracellular matrix      | 2.94E-29       | 16.67%             |
| lysosomal lumen                               | 3.69E-15       | 27.08%             |

**Figure 4.7. Expression of NANOG-mCherry and iPSC Markers in SKM and OSKM Reprogramming.**



Supplemental information related to Figure 4.1. A) NANOG-mCherry expression was assessed throughout SKM and OSKM reprogramming in relation to iPSC morphology. Endogenous Nanog-mCherry was expressed by day 14 in both groups. However, the number of NANOG-mCherry expressing colonies better aligns with iPSC morphology numbers in SKM reprogramming. B and C) Both SKM and OSKM reprogramming produces iPSC cell lines that uniformly express pluripotency markers similar to ES cells. 40x, scale bar = 50  $\mu$ m.

**Figure 4.8. Single cell sorting gating strategy, cells counted and cell growth outcomes.**



Supplemental information related to Figure 4.4. A) Single cell sorting gating strategy in which cells were selected based on size to eliminated cell debris and doublets. Cells of interest were sorted based on the absence of cell death marker, DAPI, and presence of eGFP. B) The average OCT4-eGFP live populations present in each reprogramming line collected for bulk collection were around 1% which is consistent with the reported reprogramming efficiency. n = 6 cell lines. C) 75 cells were able to expand to a passage of 12 or greater and were evenly represented throughout the single cell collection process. D) Early differentiation of suspected iPSC cells that only proliferated to passage 2. Scale bar = 100  $\mu$ m.

**Table 4.2. qPCR primers for detecting endogenous transcripts**

| <b>Gene Target</b> | <b>Forward Sequence (5' to 3')</b> | <b>Reverse Sequence (5' to 3')</b> |
|--------------------|------------------------------------|------------------------------------|
| <i>Oct4</i>        | GTTGGAGAAGGTGGAACCAA               | CCAAGGTGATCCTCTTCTGC               |
| <i>Nanog</i>       | ATGCCTGCAGTTTTTCATCC               | GAGGCAGGTCTTCAGAGGAA               |
| <i>Sox2</i>        | GCGGAGTGGAAACTTTTGTCC              | CGGGAAGCGTGTACTTATCCTT             |
| <i>Gata6</i>       | ATGCTTGCGGGCTCTATATG               | GGTTTTCGTTTCCTGGTTTG               |
| <i>Gata4</i>       | CTGGAAGACACCCCAATCTC               | ACAGCGTGGTGGTGGTAGT                |
| <i>Sox7</i>        | GGCCAAGGATGAGAGGAAAC               | TCTGCCTCATCCACATAGGG               |
| <i>Sox17</i>       | CTTTATGGTGTGGGCCAAAG               | GCTTCTCTGCCAAGGTCAAC               |
| <i>ActinB</i>      | CTGAACCCTAAGGCCAACC                | CCAGAGGCATACAGGGACAG               |

## CHAPTER 5.

Evaluating the ability of exogenous *Sall4* to replace *Oct4* in somatic cell reprogramming

Moauero A. and Ralston A.

A. Moauero wrote the chapter, assembled the figures, and performed the experiments.

A. Ralston edited the chapter.

## **Abstract**

Somatic cell reprogramming using the overexpression of transcription factors *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*), was first discovered as a pathway to form induced pluripotent stem cells (iPSCs). In addition to the production of iPSCs, *OSKM* reprogramming yields a distinct, non-pluripotent stem cell type, termed induced extraembryonic endoderm stem (iXEN) cells. It is not understood how the same four factors produce non-pluripotent stem cells alongside pluripotent stem cells. Our lab has shown that endogenous OCT4 is expressed in both induced stem cell types and may be a critical factor in the establishment of both iPSC and iXEN. Interestingly, there is another transcription factor, *SALL4*, that works upstream of OCT4 in the embryo and has been shown to play a role in pluripotent and extraembryonic endoderm formation. To better understand how iPSC and iXEN cells form in parallel, we sought to evaluate the role that *SALL4* may be playing in establishing both cell types in reprogramming. We accomplished this by replacing *Oct4* with *Sall4* alone or in combination with *Nanog* in the reprogramming cocktail. We were able to observe colonies that resembled iPSC and iXEN morphology but at a reduced efficiency. Further evaluation of these colonies revealed a lack of fluorescent marker expression associated with successful reprogramming. Interestingly, *Sall4* and *Nanog* reprogramming was sufficient to turn on key extraembryonic endoderm markers.

## Introduction

Overexpression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) have been shown to convert somatic cells into induced pluripotent stem cells (iPSCs) and induced extraembryonic endoderm (iXEN) cells (Nishimura et al., 2017; Parenti et al., 2016; Takahashi & Yamanaka, 2006). It has yet to be established how the same four factors give rise to two different embryonic-like stem cells. As explored in Chapter 4, one possibility is that OCT4 may be playing a critical role that allows for the formation of both iPSC and iXEN cells. In particular, it has been shown that during OSKM reprogramming, endogenous *Oct4* is activated in cells fated for iXEN formation and that OCT4-positive cells express XEN markers.

Although OCT4 is often considered a pluripotency factor, it is not surprising that OCT4 could play a role in iXEN formation based on our knowledge of embryo development. In the embryo, the epiblast and primitive endoderm serve as an embryonic counterpart to iPSC and iXEN, respectively. In epiblast and primitive endoderm formation, OCT4 is essential in forming both lineages. *Oct4* remains active in the epiblast but eventually becomes repressed in the primitive endoderm (Aksoy et al., 2013; Frum et al., 2013; Le Bin et al., 2014; Niakan et al., 2010; Wicklow et al., 2014). Given that OCT4 is an important factor in epiblast and primitive endoderm development, it is important to evaluate what other factors are turned on in a similar manner during embryo development to help elucidate the dual fate of reprogramming cells.

Interestingly, there is a transcription factor, SALL4 which is expressed in a similar pattern as OCT4 in embryo development. SALL4 is a C2H2-type zinc-finger transcription factor belonging to the *Spalt*-like gene family. This transcription factor has been shown to regulate development in many organ systems, including the formation of the early embryo, nervous system, heart, and limbs (Rao et al., 2010; J. Zhang et al., 2006). In early embryo development, SALL4, like OCT4, is detected at the two-cell stage as a result of maternal RNA contribution (Elling et al., 2006). Knockout studies in embryos show impaired development of the epiblast and primitive endoderm, with eventual embryonic demise after implantation at E5.5 (Elling et al., 2006). It is suspected that SALL4 functions upstream of *Oct4* to regulate the transcription of *Oct4* (J. Zhang et al., 2006). Using embryonic-derived stem cell lines known as embryonic stem (ES) cells and extraembryonic endoderm stem (XEN) cells to model epiblast and primitive endoderm respectively, studies have further investigated the role of SALL4. Knocking down *Sall4* RNA in these lines resulted in decreased pluripotent and extraembryonic endoderm marker expression (Lim et al., 2008; J. Zhang et al., 2006) further signifying the importance of this transcription factor in both cell types.

Given the importance of SALL4 in embryonic development, several labs have investigated the expression of *Sall4* during *OSKM* reprogramming. It has been shown that during reprogramming *Sall4* is upregulated early in the reprogramming process between days 2-6 (Buganim, Faddah, Cheng, Itskovich, Markoulaki, Ganz, Klemm, Van Oudenaarden, et al., 2012; Velychko et al., 2019). Additionally, knock down studies of *Sall4* RNA during reprogramming resulted in decreased putative iPSC colonies

(Tsubooka et al., 2009; Velychko et al., 2019). Together, this indicates that SALL4 is not only expressed during the reprogramming process but also participates in establishing colonies. Although *Sall4* plays a role in reprogramming and the formation of epiblast and primitive endoderm lineages in the embryo, it has yet to be determined to what extent exogenous *Sall4* can replace exogenous *Oct4* in reprogramming. Studying the effects of exogenous *Sall4* on the establishment of iPSC and iXEN cells will help inform us if this factor preferentially favors the establishment of one lineage over another. Based on the work that has been previously shown, we hypothesized that *Sall4* overexpression is sufficient to replace *Oct4* and produce both iPSC and iXEN cells at an equal rate.

## **Materials & Methods**

### ***Mouse Strains***

The following alleles were maintained in a CD-1 background; *Gata4*<sup>H2B-eGFP</sup> (Simon et al., 2018), *Gata6*<sup>tm1Hadj/J</sup> (Freyer et al., 2015), *Nanog*<sup>mCherry</sup> (Chapter 4), and *Pou5f1*<sup>tm2Jae</sup> (Lengner et al., 2007). All animal work conformed to the guidelines and regulatory standards of Michigan State University Institutional Animal Care and Use Committee.

### ***Mouse Embryonic Fibroblast (MEF) Preparation***

MEF lines were established from E13.5 embryos. After harvesting, head and viscera were removed and DNA was extracted and used for genotyping. Individual embryos were dissociated and plated on gelatin in MEF Medium [DMEM, 10% Fetal Bovine Serum

(Hyclone), Pen/Strep (10,000 units each)] and grown at 37°C with 5% CO<sub>2</sub> until confluent. Once confluent, MEF line were stored in liquid nitrogen.

### ***Reprogramming***

All MMLV-derived retrovirus was produced by transfecting 293T cells with pCL-ECO and pMXs plasmids. pMXs plasmids contained either *Oct4*, *Klf4*, *Sox2*, *cMyc*, *Nanog* or *Sall4* cDNAs (Addgene). Transfected 293T cell supernatant was harvested 48 hours later. mCherry virus was made in conjunction with all viral preps and used to infect CD-1 MEFs to determine viral titers. Viral preps were stored at -80 °C. For retroviral reprogramming (Takahashi and Yamanaka, 2006) MEFs were plated the day before at a density of 50-100 cells/mm<sup>2</sup>. Virus is then added at a MOI of 1 with Polybrene and incubated for 24 hrs. The following day, the cell medium was replaced with MEF medium, followed by Reprogramming Medium 1 [DMEM (Invitrogen), 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Fetal bovine serum (FBS; Hyclone), 10 ng/mL Leukemia Inhibitory Factor (LIF)] on days 2 and 4. Media was then replaced with Reprogramming Medium 2 [DMEM (Invitrogen), 0.1 mM Beta-mercaptoethanol, 2 mM Glutamax, 1X Non-essential amino acids, 100 U/mL Penicillin/streptomycin, 15% Knockout Serum Replacement (KOSR; Invitrogen), 10 ng/mL Leukemia Inhibitory Factor (LIF)] on day 6 and every other day there after until the end of the experiment.

### ***Colony Counting and Lab Images***

At indicated time points until day 20 of reprogramming, the number of iPSC and iXEN colonies were counted based on morphology. In addition, the presence of fluorescent markers was detected using a Lumen Prior 200 connected to an inverted Leica microscope at a 10X magnification.

### ***RNA Isolation and qPCR***

RNA was harvested using 1:6 chloroform to Trizol (Invitrogen). 1 µg RNA was reverse transcribed to create cDNA using QuantiTect Reverse Transcription Kit (Qiagen), following the manufacturer instructions. For qPCR, cDNA was amplified using a Lightcycler 480 (Roche) according to manufacturer's guidelines. The amplification efficiency of each primer pair (see Primers & Oligos) was measured by generating a standard curve from appropriate cDNA libraries using extraembryonic endoderm (XEN) cells and embryonic stem (ES) cells. All reactions were completed in quadruplicate.

## **Results**

### ***Exogenous *Sall4* is not sufficient to replace *Oct4* in the formation of iPSC or iXEN cells***

In order to test if exogenous *Sall4* can replace exogenous *Oct4* in OSKM reprogramming, we substituted *Oct4* MMLV-derived retrovirus for *Sall4* MMLV-derived retrovirus. We chose to use a MMLV-derived retrovirus rather than lentivirus as previous work has shown that SKM lentivirus is sufficient to produce iPSC colonies (Velychko et al., 2019). To determine if *Sall4* is a true substitute for *Oct4*, we chose the MMLV-derived retroviral

system which is dependent on *Oct4* to complete successful reprogramming. If *Sall4* is a good substitute for *Oct4* we would expect to see genuine iPSC and iXEN colonies forms at a rate similar to *OSKM* reprogramming.

Previous work had shown that *Sall4+SKM* was not sufficient to produce iPSC colonies (Heng et al., 2010) but did not evaluate for the formation of iXEN. It is possible that SALL4 is sufficient to produce an iXEN state without producing a pluripotent state. During *Sall4+SKM* reprogramming, colonies were observed. The total number of colonies formed was reduced by 74% relative to *OSKM* reprogramming (Fig. 5.1A). The colonies that were observed had a similar morphology to either iPSC or iXEN colonies (Fig. 5.1D). iPSC colonies are dome shaped with distinct borders where iXEN colonies appear flatter and more spread out with less-defined borders (Parenti et al., 2016).

Despite a resemblance in morphology, very little to no fluorescent markers associated with successful reprogramming were observed. Throughout reprogramming, there was no endogenous expression of OCT4-eGFP and only a few colonies expressing NANOG-mCherry by day 20 of the *Sall4+SKM* treatment, despite strong expression of both markers in *OSKM* treated reprogramming (Fig. 5.1B and C). During the 20-day process, *Sall4+SKM* reprogramming was not sufficient to induce the expression of core pluripotency factors. These results are in line with the reported literature that *Sall4+SKM* is not sufficient to produce iPSC colonies (Heng et al., 2010). In addition, OCT4-eGFP has been shown to not be a reliable marker of pluripotency and can be seen in cells fated

for an iXEN fate (Chapters 3 and 4). The lack of OCT4-eGFP would also indicate that *Sall4+SKM* is not sufficient to produce true iXEN colonies.

Alternatively, the absence of marker expression could be explained by *Sall4+SKM* reprogramming requiring a longer time to induce the expression of pluripotent and extraembryonic endoderm genes. To investigate this possibility, we picked several colonies to determine if they could maintain an iPSC or iXEN morphology and begin to express appropriate markers. In general, most cell lines failed to reach passage 12 and many stopped proliferating at around passage 5. Some of these cell lines displayed a morphology similar to established iPSC and iXEN cells (Fig. 5.1E), but with several key differences. Line A resembled iPSC cells with the formation of compacted cell clusters, however there was a single cell layer behind the clusters that is not present in typical iPSC lines. Additionally, Line B resembled iXEN morphology with a mix of round and geometric cells. However, Line B also had several cells with less defined borders and large nuclei which are not observed in typical iXEN lines. When all the lines were analyzed for the expression of pluripotent and extraembryonic endoderm markers using quantitative polymerase chain reaction (qPCR) (Fig. 5.1F), all but *Sox2* was not expressed. It is possible that *Sall4+SKM* may be inducing endogenous *Sox2*. Alternatively, *Sox2* expression may be residual expression of exogenous *Sox2* from the *Sox2* MMLV-derived retrovirus. Further evaluation into exogenous or endogenous *Sox2* is needed. Together, these data suggests that *Sall4+SKM* does induce changes in MEFs but that the colonies produced do not express the appropriate morphology or markers of genuine iPSC or iXEN cells.

### ***Exogenous Nanog plus Sall4 are not sufficient to replace Oct4 in somatic cell reprogramming***

Although it does not appear that exogenous *Sall4* alone is sufficient to replace *Oct4* in the *OSKM* reprogramming cocktail, it is possible that *Sall4+SKM* in addition to other factors may be sufficient. Buganim et al., 2012 sought to find other reprogramming cocktails that could induce an iPSC state. They demonstrated that exogenous *Nanog+Sall4* was sufficient to replace the use of exogenous *Oct4* to achieve a pluripotent state. However, this group did not report whether they observed iXEN colonies relative to *OSKM* reprogramming. We sought to investigate if *Nanog+Sall4+SKM* was sufficient to produce iXEN colonies alongside iPSC colonies.

Reprogramming with *Nanog+Sall4+SKM* produced colonies. The total number of colonies formed was reduced by 96% relative to *OSKM* reprogramming (Fig. 5.2A). Interestingly, the number of colonies declined throughout the reprogramming process which is suspected to be from cell death or reversion back to a fibroblast state. In addition, instead of reprogramming for 20 days, we continued to grow the samples to 25 days. This was to explore the possibility that reprogramming with *Nanog+Sall4+SKM* was delayed relative to *OSKM*. By day 25 there were no colonies remaining for picking and passaging. The colonies that were observed early in reprogramming had a similar morphology to genuine iPSC and iXEN colonies from *OSKM* reprogramming (Fig. 5.2D).

Despite the formation of colonies with an iPSC and iXEN morphology, there was no expression of endogenous OCT4-eGFP (Fig. 5.2B). In addition, GATA6-H2B-Venus

reporter lines were used to assess the presence of putative iXEN formation. GATA6-H2B-Venus positive colonies started to form as early as day 8 of the reprogramming process (Fig. 5.2C), although the number of positive colonies was significantly reduced from *OSKM* reprogramming. The positive colonies were scarce, making them difficult to pick and expand. Regarding GATA6 as a marker, fibroblasts often express low levels of GATA6 (Dittrich et al., 2021; Molkentin, 2000). It is common to observe occasional GATA6-H2B-Venus positive fibroblasts. However, during *Nanog+Sall4+SKM* reprogramming, there was a higher number of GATA6-H2B-Venus fluorescence cells not associated with colonies which were presumably fibroblasts (Fig 5.2E). This higher expression of GATA6-H2B-Venus in non-colony associated cells indicates that *Nanog+Sall4+SKM* is inducing the expression of GATA6 above background fibroblast levels.

***Nanog+Sall4, in conjunction with SKM, induces the expression of extraembryonic endoderm markers***

Although iXEN and iPSC lines could not be established from *Sall4+SKM* and *Nanog+Sall4+SKM* reprogramming, it is possible that these factors are close to establishing an iPSC or iXEN state. To determine if iPSC or iXEN markers are expressed, qPCR was performed on entire reprogrammed wells at day 21 (Fig. 5.3A). *OSKM* was the only reprogramming cocktail that expresses *Oct4*, which is in agreement with OCT4-eGFP reporter data (Fig. 5.1B and Fig. 5.2B). *Nanog+Sall4+SKM* was the only cocktail that showed high *Nanog* levels. This could be due to *Nanog+Sall4+SKM* being able to activate endogenous *Nanog* or could be due to the continued exogenous expression of

*Nanog* from the MMLV-derived retrovirus. In the *OSKM* treatment, *Nanog* is only expressed at 2.5% relative to the ES cell control. This low level of expression is expected as reprogramming efficiency is only around 1% (S. Yamanaka, 2009). In addition, all three cocktails contain exogenous *Sox2*. If exogenous *Sox2* is still active, this could explain why all treatment groups are expressing *Sox2*. Future studies will need to be performed to address what amount of transcripts are from exogenous or endogenous *Oct4*, *Nanog* and *Sox2*.

Although all three core pluripotency factors are not expressed in *Sall4+SKM* and *Nanog+Sall4+SKM*, this does not limit the ability for each cocktail to contribute to iXEN formation. Interestingly, *Sall4+SKM* showed very little to no expression of extraembryonic endoderm markers. However, *Nanog+Sall4+SKM* treated samples had upregulation of four core extraembryonic endoderm markers (GATA6, GATA4, SOX7 and SOX17). This lends evidence that *Nanog+Sall4* may be helping to induce an extraembryonic endoderm state.

## **Discussion**

Despite *Sall4* being expressed at similar embryonic stages as *Oct4* in embryo development and being required for the formation of the epiblast and primitive endoderm, exogenous *Sall4* is not sufficient to substitute exogenous *Oct4* in reprogramming. However, this does not exclude the possibility that other factors in combination with *Sall4* can replace the use of exogenous *Oct4*. Based on Buganim et al., 2012, the use of *Nanog+Sall4* is sufficient to replace *Oct4* and produce iPSCs. We chose to repeat this

experiment and evaluate for the formation of iXEN alongside iPSCs. To our surprise, *Nanog+Sall4+SKM* failed to establish a pluripotent state but instead may be working toward producing an extraembryonic endoderm state. We suspect this discrepancy was from the use of SKM lentivirus in Buganim et al., 2012 instead of MMLV-derived retrovirus. This small detail is crucial as *SKM* lentivirus alone is sufficient to produce iXEN and iPSC colonies (Velychko et al., 2019; Chapter 3) making it challenging to parse out the effects of *Nanog+Sall4+SKM* compared to *SKM* alone.

Although *Nanog+Sall4+SKM* was not sufficient to produce iPSC or iXEN lines, this treatment appears to be on a successful pathway to establishing a reprogramming state relative to *Sall4+SKM*. However, more studies need to be completed to confirm these findings. First, *Nanog+Sall4+SKM* reprogramming should be completed on a larger scale in hopes of collecting colonies that can be picked and expanded. If colonies can be expanded, they can then be analyzed for gene expression and morphology. Additionally, RNA-seq should be performed to compare the transcriptomes of *Nanog+Sall4+SKM* to *OSKM* to determine if any genes are differentially expressed. This differential expression will help us understand what pathways and targets are not activated by *Nanog+Sall4* but are activated by *Oct4*.

Since *Nanog+Sall4* was not sufficient to replace exogenous *Oct4*, it is important to evaluate other reprogramming cocktails that have been reported to achieve a pluripotent state to better understand how pluripotency is achieved and determine if these cocktails can establish other cell fates such as iXEN (Buganim, Faddah, Cheng, Itskovich,

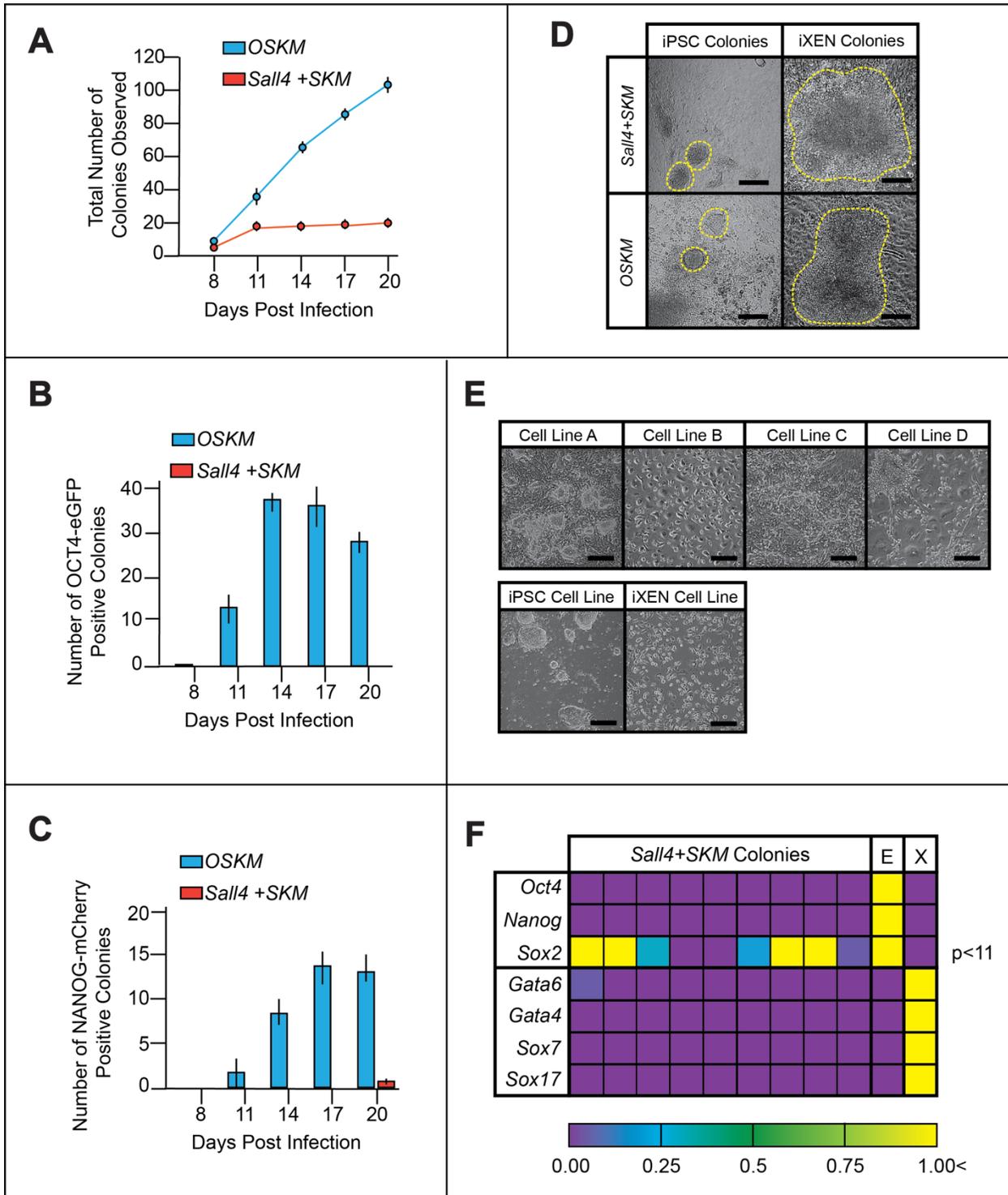
Markoulaki, Ganz, Klemm, Van Oudenaarden, et al., 2012; Shu et al., 2015b; Wang et al., 2019). Based on previously published cocktails, future targets to explore include *Gata6*, *Lin28*, *Essrb*, and *Dppa2* in various combinations with *Nanog*, *Sall4* and *SKM*.

### **Acknowledgements**

We thank Anna-Katerina Hadjantonakis for providing us *Gata4<sup>H2B-eGFP</sup>* and *Gata6<sup>tm1Hadj/J</sup>* mice. In addition, we would like to thank Shinya Yamanaka for his development of the Addgene plasmids.

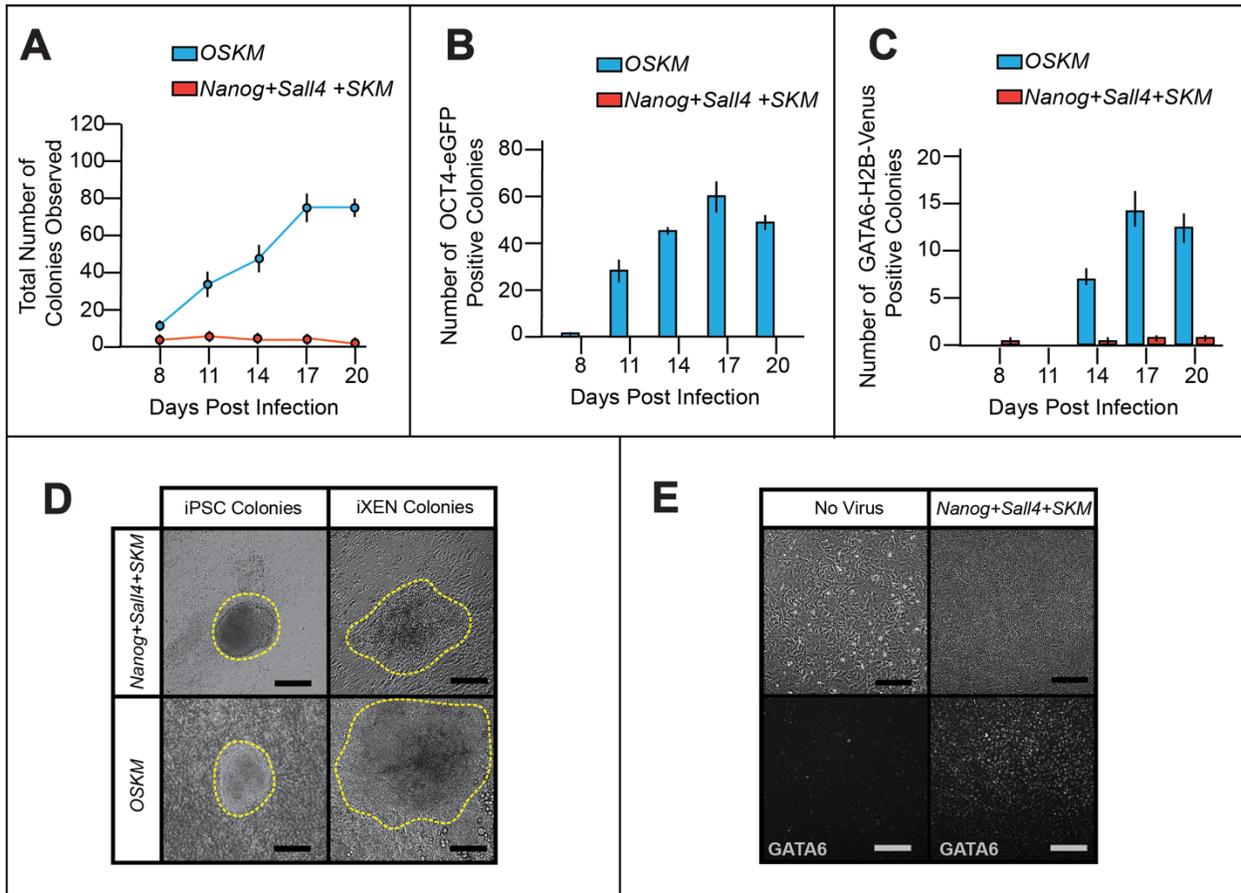
## APPENDIX

**Figure 5.1. *Sall4*+SKM somatic cell reprogramming failed to produce iPSC or iXEN colonies.**



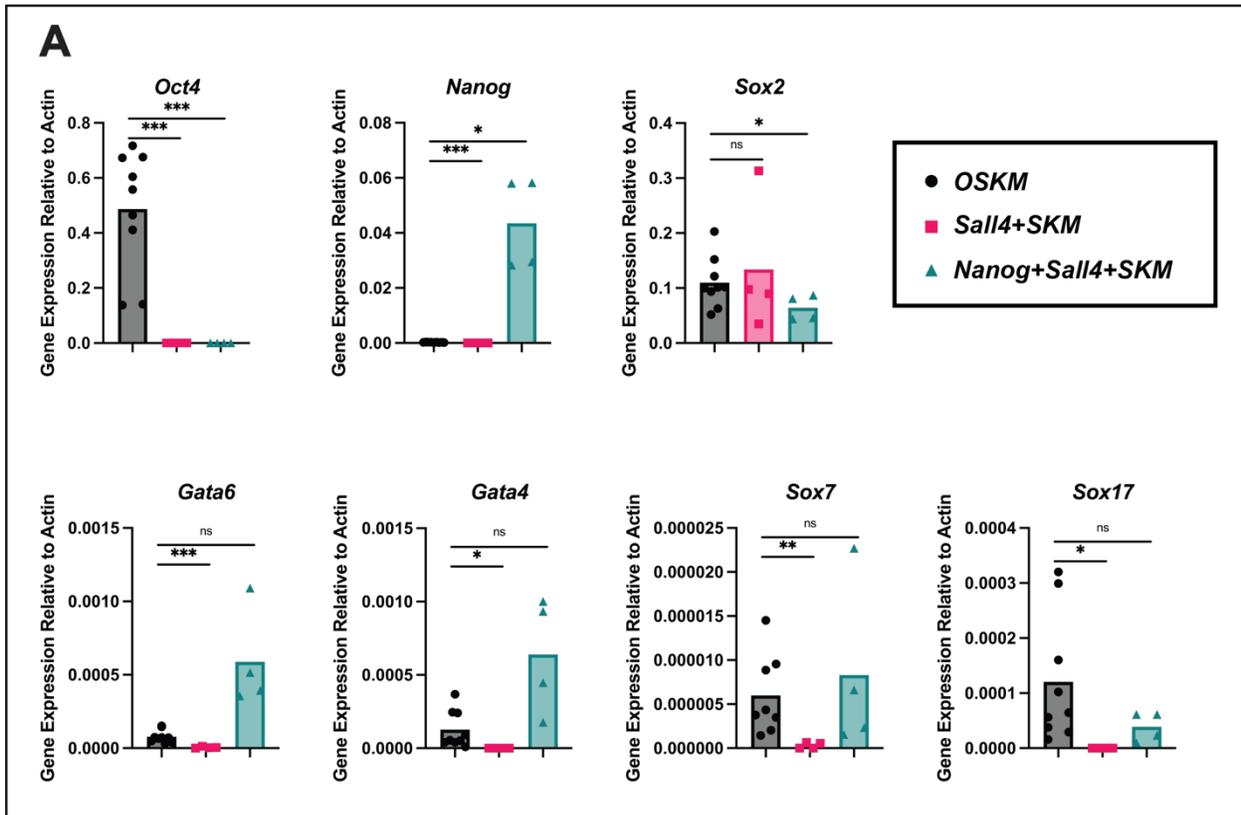
**Figure 5.1. (cont'd).** A) *Sall4*+*SKM* reprogramming produced colonies at a lower efficiency than *OSKM* (n = 9 per treatment). B and C) *Sall4*+*SKM* colonies did not express endogenous OCT4, and few colonies expressed endogenous NANOG (n = 3 per treatment). D) The colonies formed by day 20 of *Sall4*+*SKM* reprogramming had a similar iPSC and iXEN morphology. E and F) Expansion of *Sall4*+*SKM* colonies resulted in a loss of iPSC and iXEN morphology and no iPSC or iXEN marker expression when analyzed using qPCR. Expression is relative to ES cells (E) or XEN cells (X). Error bars were calculated using standard error. Scale bars = 200  $\mu$ m. qPCR samples were run in quadruplicates.

**Figure 5.2. *Nanog+Sall4+SKM* somatic cell reprogramming failed to produce iPSC or iXEN colonies.**



A) *Nanog+Sall4+SKM* reprogramming produced colonies at a lower efficiency than OSKM (n = 6 per treatment). B and C) *Nanog+Sall4+SKM* colonies failed to express endogenous OCT4. However, some colonies expressed endogenous GATA6 (n = 3 per treatment). D) The colonies formed in *Nanog+Sall4+SKM* reprogramming displayed a similar iPSC and iXEN morphology. E) Treated cells not associated with a colony displayed endogenous *Gata6* expression relative to non-treated fibroblasts. Error bars were calculated using standard error. Scale bars = 200  $\mu$ m.

**Figure 5.3. *Nanog*+*Sall4*+SKM somatic cell reprogramming leads to the expression of extraembryonic endoderm markers.**



A) *Sall4*+SKM did not express *Oct4*, *Nanog* or extraembryonic endoderm markers. The addition of *Nanog* to *Sall4*+SKM lead to the expression of extraembryonic endoderm markers but still did not activate *Oct4* expression. qPCR samples were run in quadruplicates. Welch's test was used to determine significance of gene expression between samples. \* p-value < 0.05, \*\* p-value < 0.005, \*\*\* p-value < 0.0005, ns = not significant. n = 4< for each treatment group.

**Table 5.1. qPCR primers for detecting endogenous transcripts**

| <b>Gene Target</b> | <b>Forward Sequence (5' to 3')</b> | <b>Reverse Sequence (5' to 3')</b> |
|--------------------|------------------------------------|------------------------------------|
| <i>Oct4</i>        | GTTGGAGAAGGTGGAACCAA               | CCAAGGTGATCCTCTTCTGC               |
| <i>Nanog</i>       | ATGCCTGCAGTTTTTCATCC               | GAGGCAGGTCTTCAGAGGAA               |
| <i>Sox2</i>        | GCGGAGTGGAAACTTTTTGTCC             | CGGGAAGCGTGTACTTATCCTT             |
| <i>Gata6</i>       | ATGCTTGCGGGCTCTATATG               | GGTTTTCGTTTCCTGGTTTG               |
| <i>Gata4</i>       | CTGGAAGACACCCCAATCTC               | ACAGCGTGGTGGTGGTAGT                |
| <i>Sox7</i>        | GGCCAAGGATGAGAGGAAAC               | TCTGCCTCATCCACATAGGG               |
| <i>Sox17</i>       | CTTTATGGTGTGGGCCAAAG               | GCTTCTCTGCCAAGGTCAAC               |
| <i>ActinB</i>      | CTGAACCCTAAGGCCAACC                | CCAGAGGCATACAGGGACAG               |

**Table 5.2. Genotyping primers**

| <b>Gene Target</b>                | <b>Forward Sequence (5' to 3')</b> | <b>Reverse Sequence (5' to 3')</b> |
|-----------------------------------|------------------------------------|------------------------------------|
| <i>Nanog</i> <sup>mCherry</sup>   | CCACTAGGGAAAGCCATGCGC<br>ATTT      | GGAAGAAGGAAGGAACCTGG<br>CTTTGC     |
| <i>Oct4</i> <sup>eGFP</sup>       | CCAAAAGACGGCAATATGGT               | CAAGGCAAGGGAGGTAGACA               |
| <i>Oct4 Wild Type</i>             | TGCCAGACAATGGCTATGAG               | CAAGGCAAGGGAGGTAGACA               |
| <i>Gata6</i> <sup>H2B-Venus</sup> | CCAGGGAGCTCTGAGAAAAG               | CCTTAGTCACCGCCTTCTTG               |
| <i>Gata6 Wild Type</i>            | CCAGGGAGCTCTGAGAAAAG               | GTCAGTGAAGAGCAACAGGT               |

## CHAPTER 6.

Where to explore next in the complicated landscape of reprogramming?

Moauro A. and Ralston A.

A. Moauro assembled the figures and wrote the chapter.

A. Ralston edited the chapter.

## **Abstract**

Somatic cell reprogramming using the overexpression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*) produces induced pluripotent stem cells (iPSCs). It was not until more recently that *OSKM* reprogramming was discovered to produce additional induced stem cell types alongside iPSCs, termed induced extraembryonic endoderm (iXEN) and induced trophoblast stem cells (iTSCs). This thesis has focused on understanding how iPSC and iXEN cells form in parallel. To accomplish this task, studies have been completed to investigate the use of fluorescent markers for the identification of early colonies and evaluate the use of different transcription factor cocktails in iXEN production with a focus on OCT4's expression during iXEN formation. Like most work, this thesis has only begun to scrape the surface of evaluating the formation of iXEN, leaving many studies left to be completed. This final chapter provides several proposals on where to explore next in the complicated landscape of reprogramming.

## Introduction

Somatic cell reprogramming to produce induced pluripotent stem cells (iPSCs) using the overexpression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (*OSKM*) was a revolutionary discovery (Takahashi & Yamanaka, 2006). Since the inception of *OSKM* reprogramming in 2006, this method continues to be studied. One reason is that induced stem cells provide an ethical and abundant source of stem cells which holds great promise as research models and therapies. However, *OSKM* reprogramming takes a month to complete and is inefficient (S. Yamanaka, 2009). Many researchers are continuing to develop new cocktails and methods to improve this process. In addition to improving the process of reprogramming, *OSKM* overexpression can be studied to better understand cell fate decisions and the induction of new cell identities. My thesis has largely focused on this later effort by understanding if key transcription factor OCT4 is playing a role in the cell fate decision between iPSC and iXEN formation.

My work has established methods and standards for evaluating the formation of iPSCs and iXEN colonies as they arise during reprogramming (Chapter 2). These methods can be easily modified to investigate other potentially interesting cell populations that arise by changing the markers observed. I have then gone on to evaluate the potential for blastocyst lineage markers to be used in evaluating the formation of iPSCs and iXEN cells with a focus on OCT4 expression (Chapters 3 and 4). Lastly, I began to explore how other published cocktails for iPSC formation affect the formation of other cell fates such as iXEN (Chapter 5).

This work has made several key discoveries that will hopefully guide future reprogramming efforts. However, a lot remains to be elucidated in this system. Over the last 16 years, the field has undoubtedly revealed that *OSKM* reprogramming is a complex and dynamic processes filled with changes in chromatin state, transcription factor binding and gene express (Cacchiarelli et al., 2015; Chronis et al., 2017; Knaupp et al., 2017; D. Li et al., 2017; Raab et al., 2017; Soufi et al., 2012, 2015; Takahashi et al., 2014). To make better sense of this labyrinthine process, I propose that the field of reprogramming evaluate four main topics which include the continued study of cell fate decisions using single-cell and longitudinal studies, better application of embryonic development knowledge to reprogramming, reanalysis of published reprogramming papers to evaluate for non-iPSC fates and further examination into the potential of iXEN cells.

### **Continued analysis of cell fate decisions using single-cell and longitudinal studies**

Since reprogramming is a complicated and inefficient process, this produces a lot of unwanted events and noise. There is a major need to reduce this noise in order to find meaningful information (Cevallos et al., 2020). When analyzing reprogramming cells in bulk, small populations of interesting cells are often reduced to background noise. Luckily as new single-cell methods become more accessible, it should advance our understanding of what is occurring in these rare populations. It is important that researchers continue to use methods including but not limited to scRNA-seq, flow cytometry, fluorescently-activated cell sorting and confocal imaging to obtain single-cell resolution.

Unfortunately, many of these widely available single-cell methods require the fixation or destruction of cells to analyze cell states. This means that we lose out on temporal studies and the ability to track the changes occurring within a cell over time. It is important that the field begins to better incorporate temporal studies with the use of tools like pseudotime, cell barcoding and live cell tracking.

Not only should single-cell techniques with temporal ability be applied to reprogramming in general, but there are two areas of reprogramming that have been understudied in which these tools could be most fruitful. These areas include the formation of mixed morphology colonies and the study of cell fate after colonies have been picked for expansion. To start, in Chapter 3, I demonstrate that mixed colonies are in abundance at a similar level to iPSC colonies. These colonies exhibit characteristics resembling both iPSC and iXEN cells. I propose that cells within these colonies should be evaluated for the propensity to form iPSC or iXEN fates. If both cell types exist within mixed colonies this provides another unique source for studying cell fate decisions and determine how and why both cell states are created from presumably a single reprogramming cell.

Additionally, in Chapter 3, I highlight that several fluorescent markers are not reliable for identifying putative iPSC or iXEN colonies until colonies have been picked and expanded to create stable cell lines. Typically reprogramming is thought to be complete after 3 weeks. However, the fluorescent data would suggest that the pluripotent and extraembryonic endoderm network are still changing after the 21-day process which is evident by the late expression of fluorescent reports. To date, there have been no detailed

studies evaluating the changes in gene expression and chromatin state after the end of reprogramming. Studying this later phase in which colonies cement their cell identity could provide valuable information into late cell fate decisions.

### **Using the second cell fate decision of embryo development to guide reprogramming studies**

*OSKM* reprogramming has been shown to produce three induced stem cell types: iPSCs, iXEN and iTSCs (Castel et al., 2020; X. Liu et al., 2020; Parenti et al., 2016; Takahashi & Yamanaka, 2006). Interestingly, early embryo development produces these same embryonic lineages termed epiblast, primitive endoderm and trophoctoderm, respectively (Fig 6.1.A). A large field of research has been and continues to be dedicated to understanding how the three lineages of the early embryo are formed. It is possible that the same mechanisms that guide embryo development also guide cell reprogramming. My thesis work mainly focuses on understanding the formation of iPSC and iXEN cells. To better understand this decision, we can focus on the second cell fate decision in embryo development (Fig 6.1.B).

During the second cell fate decision, the cells of the inner cell mass must commit to either the epiblast or primitive endoderm lineages. Since iPSC and iXEN cells are present in reprogramming, it is possible that they are formed in a manner similar to the epiblast and primitive endoderm, in which cells intended for either fate must pass through an inner cell mass-like state. Exploration of this state could be evaluated by looking for the concurrent expression of inner cell mass genes such as *Gata6*, *Nanog*, *Sox2* and *Oct4* (G. Guo et

al., 2010; Plusa et al., 2008) (Fig 6.2.A). As explored in Chapter 4, this state could help explain why *Oct4* is present in iPSCs and iXEN cells and why inner cell mass genes are expressed in reprogramming cells at day 17.

In addition to exploring the possibility of an inner cell mass-like state, the second cell fate decision is dependent on RTK-ERK and PI3K signaling. Cells of the epiblast express and secrete FGF4 to transiently and sporadically activate their own FGFR1/ERK signaling pathway. Cells of the primitive endoderm express FGFR1 receptors to receive a constant FGF4 signal to activate ERK while in parallel activating RTK and PI3K signaling through FGFR2 and PDGFRFA receptors (Azami et al., 2019; Bessonard et al., 2019; Chazaud et al., 2006; G. Guo et al., 2010; Kang et al., 2013, 2017; Molotkov et al., 2017; Simon et al., 2021; Y. Yamanaka et al., 2010). It is possible that just like in the embryo, cell fates can be pushed to a pluripotent or extraembryonic endoderm fate by disrupting the FGF4 and FGFR signaling (Frum et al., 2013; Frum & Ralston, 2020). If this is the case, growing reprogramming cells in the presence of FGF4 or FGFR inhibitor could produce more or less iXEN colonies.

### **Reanalyzing reprogramming studies to include analysis of iXEN and iTSCs**

Since the discovery of iPSCs using *OSKM* overexpression, several studies have been conducted to investigate changes in chromatin state and gene expression throughout the 21-day process. The goal of these studies was to uncover the specific pathways that lead to successful pluripotency formation (Cacchiarelli et al., 2015; Chronis et al., 2017; Knaupp et al., 2017; D. Li et al., 2017; Raab et al., 2017; Soufi et al., 2012, 2015;

Takahashi et al., 2014). Unfortunately, many of these studies were completed before iTSCs and iXEN cells were discovered. It would be fruitful to reevaluate these large, published data sets to look for evidence of iXEN and iTSC fate formation. This would help to make sense of published ectopic gene expression observed during reprogramming (Cacchiarelli et al., 2015; González & Huangfu, 2016; Meissner et al., 2007; Mikkelsen et al., 2008; Takahashi et al., 2014; Xing et al., 2020). This reanalysis would also help answer the following questions: are there DNA binding sites that *OSKM* resides on that are specific to iPSCs, iXEN, iTSCs or inner cell mass-like states? Does *OSKM* or a combination of these factors favor gene expression important for one cell fate over another? Alternatively, does *OSKM* promiscuously bind to DNA creating large sections of euchromatin that are not specific to any one fate?

In addition to the reanalysis of large data sets with a more widened focus on iTSCs and iXEN cell formation, there are several reprogramming studies that can be reconducted. Many published studies change the transcription factors used to induce pluripotency to create the highest quality iPSCs, improve reprogramming efficiency and discover new pathways to pluripotency formation. However, these studies have yet to evaluate for the presence of iTSCs or iXEN cells using these new cocktails. Looking for the existence of these cell types will provide better information on which transcription factors are specific for one cell fate over another. For example, in Chapter 4, the idea that iXEN could be produced without exogenous *Oct4* originated from Velychko et al., 2019 which only analyzed for the presence of iPSC formation and no other cell type. Additionally, papers like Buganim et al., 2012; Maekawa et al., 2011; and Wang et al., 2019 have tried several

cocktails to induce pluripotency networks but overlooked the possibility of alternative cell fates such as iTSCs and iXEN. Including all types of colonies in analysis can tell us how specific or nonspecific these factors truly are at inducing cell fates.

### **Evaluating the potential of iXEN cells**

Although it is known that iXEN cells can be created from reprogramming fibroblasts using *OSKM* overexpression, it is still not known what potential these cells hold for human studies and therapy development. To address these points, it is important to investigate whether human *OSKM* reprogramming produces iXEN cells and what promise iXEN cells holds for stem cell therapies.

To start, in human and mouse embryo development, the primitive endoderm cells are in contact with the epiblast cells (Stern & Downs, 2012) (Fig 6.3.A). This allows the primitive endoderm to transmit key signals to the epiblast that permit the formation of critical components such as the anterior-posterior axis, blood islands and the anterior neural plate in mice (Belaoussoff et al., 1998; Stuckey et al., 2011; Thomas & Beddington, 1996). To date there have been no reported establishment of human XEN cell lines derived from human embryos (Rossant, 2014). Instead, insights into the fundamental processes of human development have been mainly studied using the mouse embryo and mouse embryo-derived XEN cell lines (Moerkamp et al., 2013). This has limited our ability to study human XEN. This knowledge gap could be addressed by investigating if human fibroblasts induced with *OSKM* can produce iXEN cells. The establishment of a human

iXEN line will enable future studies that explore environmental or genetic influences on human XEN and how this might impact development (Linneberg-Agerholm et al., 2019).

In addition to studying human iXEN to better understand development, the establishment of human iXEN may provide great therapeutic promise. In mice, the primitive endoderm lineage has been shown to functionally contribute to definitive endoderm lineages by incorporating into the gut tube (Kwon et al., 2008). This demonstrates that the primitive endoderm is plastic and not strictly extraembryonic. Thus, XEN could be used to engineer human tissues of endodermal origin. In addition, it has been shown that canine iXEN cells can be easily transdifferentiated into hepatocyte-like cells, which demonstrates the medical relevance of iXEN cells (Nishimura et al., 2017). It is possible that iXEN cells may serve as an easier starting cell to generate definitive endoderm due to their similarities in gene expression and chromatin state (Nowotschin et al., 2019; Pijuan-Sala et al., 2019). To progress the use of iXEN, it is important to look at the transdifferentiation potential of iXEN cells and determine which transdifferentiated cell types may be useful for therapy development.

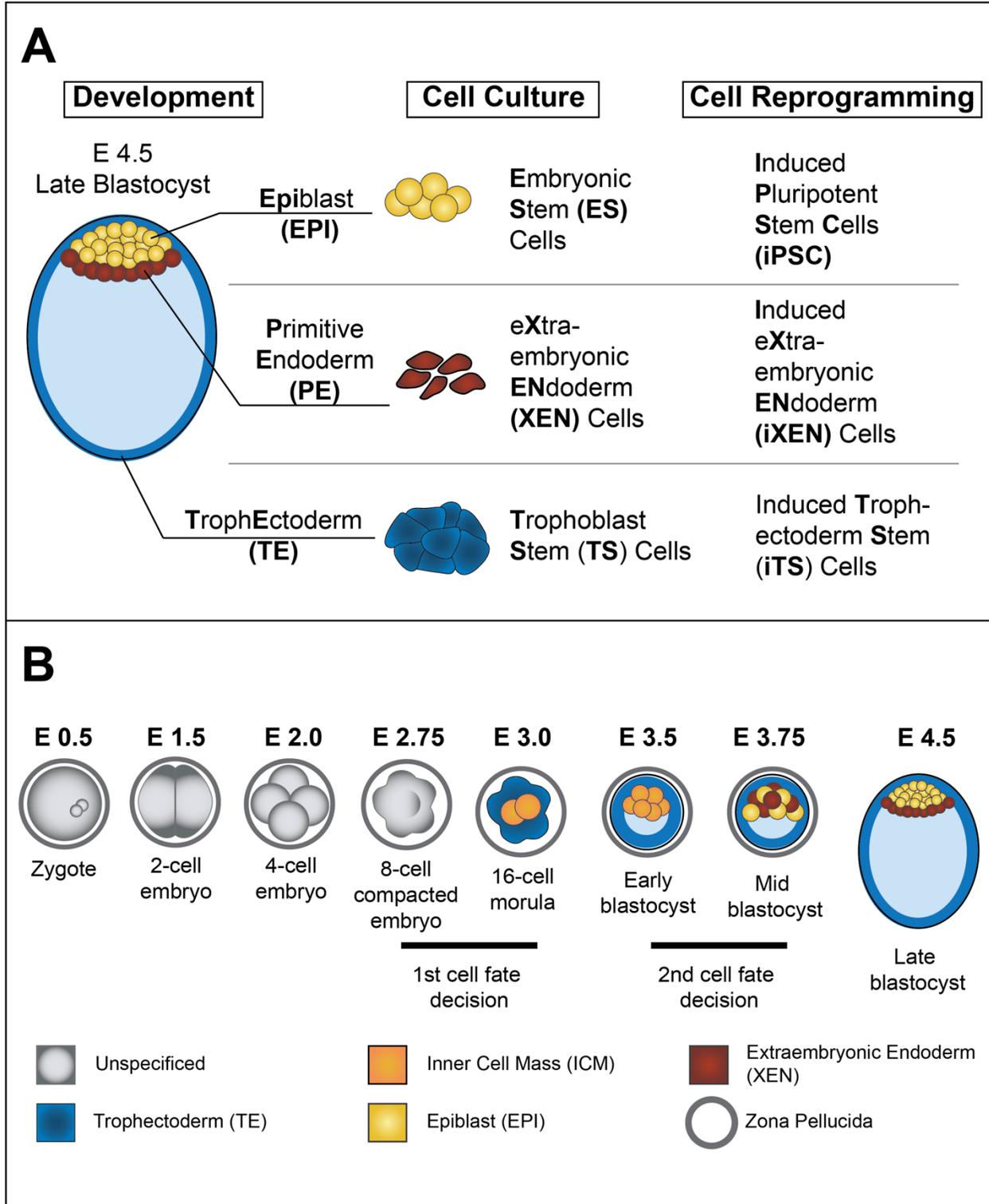
## **Conclusions**

As new information and knowledge arises, it is easy to look back at previous work and observe where pitfalls have occurred. Receiving training in an embryo development lab makes it easy to appreciate the parallels between reprogramming and development and wonder why more studies do not use the embryo as a teacher to understand reprogramming. However, the beauty of reprogramming is within its complexity which

allows for the continued discovery of new information. Pioneers in this field have taken on large challenges such as understanding the changes in chromatin structure and gene transcription during *OSKM* reprogramming, thus providing large data sets that are accessible to review and reanalyze. Even the prior efforts to create new pluripotency pathways using different transcription factor cocktails lays out several studies that can be easily reconducted and analyzed for new cell types. These studies are all low hanging fruit waiting to be picked. However, if we want to understand the impact of iXEN for research and medicine, it is important to take the steps that no one has dared to venture and explore what iXEN formation truly means for humans. I hope that moving forward these ideas will not be lost and that this research may continue.

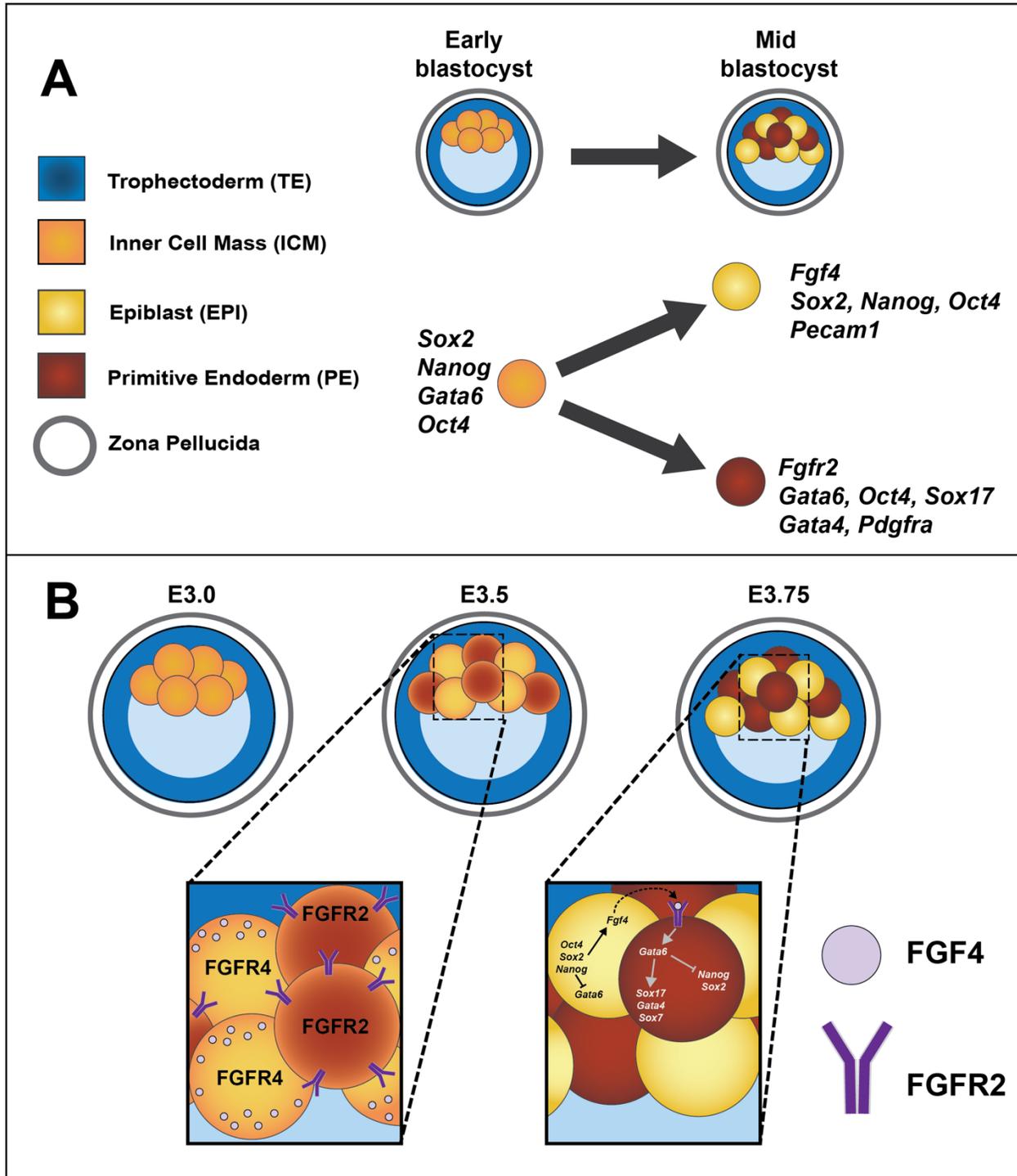
## APPENDIX

Figure 6.1. Embryo development and the formation of embryonic stem cells.



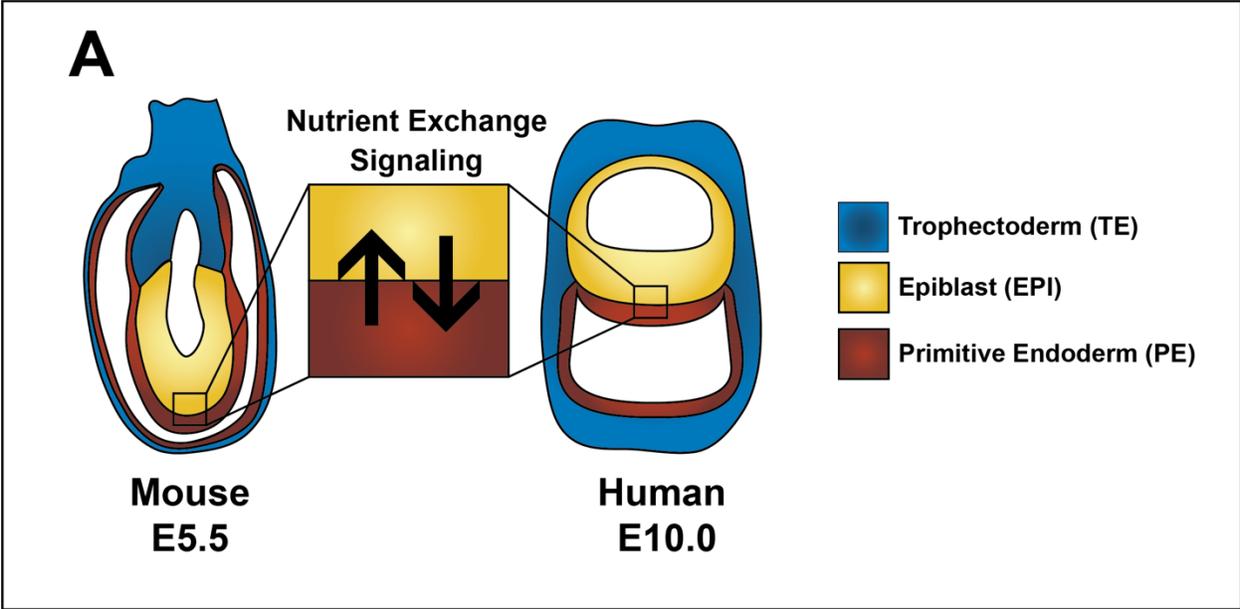
**Figure 6.1. (cont'd).** A) Embryonic stem cells can be harvested from the embryo to represent the different cell lineages in the late blastocyst. These embryonic-derived stem cells also have induced stem cell counter parts which are produced in somatic cell reprogramming. B) Early embryo development provides a unique model for understanding the formation of stem cells and studying stem cells as they journey from a totipotent cell (zygote) into pluripotent (EPI) and multipotent cells (PE and TE).

Figure 6.2. Second cell fate decision in mouse embryo development.



**Figure 6.2. (cont'd).** A) As cells of the ICM commit to the EPI or PE lineage, there are specific changes in gene transcription. In the ICM, *Oct4*, *Sox2*, *Nanog* and *Gata6* are co-expressed. As cells commit to the EPI, *Gata6* becomes repressed where in the PE *Gata6* remains expressed and *Nanog* and *Sox2* become repressed followed later by *Oct4*. B) As cells of the ICM choose the EPI vs PE lineage, ICM cells begin to express FGF4 or FGFR2. This creates a signaling mechanism in which the FGFR/ERK pathway helps to support *Gata6* expression and endodermal genes.

Figure 6.3. Human and mouse early embryo development.



A) In early embryo development human and mouse have morphological differences. However, in both species, the EPI remains in close contact with the PE which allows for critical nutrient exchange and signaling.

## CURRICULUM VITAE

Alexandra (Alex) Marie Moauro, Ph.D.  
moauroal@msu.edu  
Biochemistry Building, Rm 419  
Michigan State University  
East Lansing, MI  
May 20<sup>th</sup>, 2022

## Educational and Professional History

### Education

#### Undergraduate Education

| <u>Year</u> | <u>Degree</u>        | <u>Institution</u> |
|-------------|----------------------|--------------------|
| 2011-15     | B.S. in Biochemistry | Regis University   |
| 2011-15     | B.S. in Biology      | Regis University   |

#### Graduate Education (Current)

| <u>Year</u> | <u>Degree</u>       | <u>Institution</u>        |
|-------------|---------------------|---------------------------|
| 2016-2022   | Ph.D. in Physiology | Michigan State University |
| 2017-2024   | D.O. (OMS-III)      | Michigan State University |

Thesis research: “Expression and roles of blastocyst lineage-determining genes during somatic cell reprogramming”

Comprehensive Exam: Passed; May 27<sup>th</sup>, 2020

Dissertation: Passed; May 2<sup>nd</sup>, 2022

USMLE Step 1: Passed (228); June 17<sup>th</sup>, 2019

COMLEX Level 1: Passed (569); July 1<sup>st</sup>, 2019

#### Certification

TeamSTEPPS (Strategies Tools to Enhance Performance and Patient Safety)

ACLS (Advanced Cardiac Life Support)

BLS (Basic Life Support)

### Professional and Academic Positions

#### Academic

| <u>Title of Position</u>    | <u>Year</u> | <u>Institution</u> |
|-----------------------------|-------------|--------------------|
| TA for Nursing Biochemistry | 2014-15     | Regis University   |
| Chemistry Tutor             | 2013-15     | Regis University   |

## Professional

| <u>Title of Position</u> | <u>Year</u> | <u>Company</u>                     |
|--------------------------|-------------|------------------------------------|
| Medical Scribe           | 2015-16     | Denver Health Emergency Department |
| Medical Scribe           | 2015-16     | Chatfield Family Medicine          |
| Gymnastics Coach         | 2005- 14    | Gymnastika                         |

## Honors and Awards

| <u>Year</u> | <u>Honor</u>  |
|-------------|---|
| 2022        | Michigan Osteopathic Association Scientific Research Exhibit Best Basic Science Oral Presentation |
| 2022        | Walter S. Strodes, D.O. Memorial Scholarship from the Colorado Springs Osteopathic Foundation     |
| 2022        | RDSP Annual Research Day Best Poster Presentation   |
| 2022        | Kudos Award for Upholding MSU COM Core Values   |
| 2021        | Dissertation Completion Fellowship  |
| 2017-19     | Top Quintile, Preclerkship  |
| 2015        | Summa Cum Laude   |
| 2015        | American Institute of Chemistry Award   |
| 2014        | A.W. Forstall, S.J. Award for Excellence in Analytical Chemistry                                  |
| 2013        | William T. Miller, SJ Scholarship for Excellence in Organic Chemistry                             |
| 2011-14     | Dean's List   |

## Research

### Projects

| <u>Year</u> | <u>Project</u>   |
|-------------|--|
| 2016-2022   | Expression and roles of blastocyst lineage-determining genes during somatic cell reprogramming with Amy Ralston, PhD at Michigan State University. Stem cell Research. East Lansing, MI. |
| 2014-15     | Role of aspartic acid 101 in <i>E. coli</i> alkaline phosphatase architectural activity and stability with Stacy Chamberlin, PhD at Regis University. Biochemistry Research. Denver, CO  |
| 2013        | Tuberculosis drug synthesis to inhibit aldolase II with Kateri Ahrendt, PhD at Regis University. Organic Synthesis Research. Denver, CO.   |

## ***Future Goals and Research Interests***

- Medical Training Goals
  - Secure a residency that will provide me with the skill set that I need to provide outstanding patient care and take on challenging cases
  - Secure a residency that will allow me to grow my interests in research through clinical trial work and expand Diversity, Equity, and Inclusion (DEI) efforts
- Physician Scientist Goals
  - Obtain a residency that facilitates my interest in regenerative medicine research
  - Help translate laboratory research into patient therapies by partaking in clinical trials and partnering with labs to improve prospective therapies
  - Become a leader in the field of regenerative medicine
- Humanitarian Goals
  - Work with underserved populations to bring essential care to needed areas

## **Scholarship**

### ***Papers***

| <u>Year</u> | <u>Work</u>  |
|-------------|--|
| 2022        | <b>Moauro A<sup>1</sup></b> , Parenti A <sup>2</sup> , Halbisen M <sup>2</sup> and Ralston A*. (2022). Oct4 is expressed in cells fated for induced extraembryonic endoderm formation in somatic cell reprogramming. Stem Cell Reports. In preparation. Stem Cell Reports.   |
| 2022        | <b>Moauro A<sup>1</sup></b> , O'Hagan D <sup>2</sup> , Robin Kruger <sup>2</sup> and Ralston A*. (2022). A closer examination of fluorescent reporters NANOG, OCT4, GATA6 and GATA4 during somatic cell reprogramming reveals unexpected expression in multiple colony types. Submitted. Cell Reprogramming.                   |
| 2022        | <b>Moauro A<sup>1</sup></b> and Ralston A*. (2022). Distinguishing between iXEN and iPSC in OSKM somatic cell reprogramming. Methods in Molecular Biology. Springer. 2022; 2429:41-55. doi: 10.1007/978-1-0716-1979-7_4. <a href="https://doi.org/10.1007/978-1-0716-1979-7_4">https://doi.org/10.1007/978-1-0716-1979-7_4</a> |
| 2018        | Watts J <sup>1</sup> , Lokken A <sup>1</sup> , <b>Moauro A<sup>2</sup></b> , and Ralston A*. (2018). Capturing and interconverting cell fates in a dish. Cell Fate in Mammalian Development, Current Topics in Developmental Biology. 128, 181-199.  |

## Posters

| <u>Year</u> | <u>Poster</u>  |
|-------------|--|
| 2022        | <b>Moauro A<sup>1</sup></b> , Hickey S <sup>2</sup> , Halbisen M <sup>2</sup> and Ralston A*. 2022.<br>Exogenous <i>Oct4/Pou5f1</i> is not required to produce induced extraembryonic endoderm stem cells during somatic cell reprogramming. MD-PhD Conference Sponsored by CU Medical School. Frisco, CO. In preparation.   |
| 2022        | <b>Moauro A<sup>1</sup></b> , Hickey S <sup>2</sup> , Halbisen M <sup>2</sup> and Ralston A*. 2022.<br>Exogenous <i>Oct4/Pou5f1</i> is not required to produce induced extraembryonic endoderm stem cells during somatic cell reprogramming. International Society for Stem Cell Research Annual Conference. San Francisco, CA. In preparation.  |
| 2022        | <b>Moauro A<sup>1</sup></b> , Halbisen M <sup>2</sup> and Ralston A*. 2022.<br>OCT4/POU5f1 is expressed in both pluripotent and non-pluripotent stem cells during somatic cell reprogramming. Michigan Osteopathic Association. Southfield, MI.  |
| 2022        | <b>Moauro A<sup>1</sup></b> , Halbisen M <sup>2</sup> and Ralston A*. 2022.<br>OCT4/POU5f1 is expressed in both pluripotent and non-pluripotent stem cells during somatic cell reprogramming. American Physician Scientist Association National Meeting. Chicago, IL.  |
| 2021        | <b>Moauro A<sup>1</sup></b> and Ralston A*. OCT4 is expressed in cells fated for induced extraembryonic endoderm (iXEN) formation during somatic cell reprogramming. Society for Developmental Biology National Meeting. Online.   |
| 2020        | <b>Moauro A<sup>1</sup></b> and Ralston A*. OCT4 labels two distinct stem cell types in somatic cell reprogramming. Society for Developmental Biology National Meeting. Online.  |
| 2019        | <b>Moauro A<sup>1</sup></b> , O'Hagan D <sup>2</sup> , Parenti A <sup>2</sup> and Ralston A*. Are induced pluripotent stem cells (iPS) and induced extraembryonic endoderm (iXEN) cells formed through a common progenitor cell during <i>Oct4</i> , <i>Sox2</i> , <i>Klf4</i> and <i>c-Myc</i> (OSKM) somatic cell reprogramming?. Society for Developmental Biology Regional Meeting. Cleveland, OH.                       |
| 2019        | <b>Moauro A<sup>1</sup></b> , O'Hagan D <sup>2</sup> , Parenti A <sup>2</sup> and Ralston A*. Are induced pluripotent stem cells (iPS) and induced extraembryonic endoderm (iXEN) cells formed through a common progenitor cell during <i>Oct4</i> , <i>Sox2</i> , <i>Klf4</i> and <i>c-Myc</i> (OSKM) somatic cell reprogramming?. Reproductive and Developmental Sciences Program Annual Research Day (ARD). Portland, MI. |

- 2018 **Moauro A<sup>1</sup>** and Ralston A. Are iPS and iXEN formed through a common progenitor cell during OSKM cell reprogramming?. 2018 Fourth Floor Poster Presentation. Michigan State University.
- 2015 **Moauro A<sup>1</sup>**, Brown J<sup>2</sup>, Wallerius K<sup>2</sup>, Chamberlin S\*. 2015. The Role of Aspartic Acid 101 in *E. coli* Alkaline Phosphatase Architectural Activity and Stability. ACS.
- 2014 **Moauro A<sup>1</sup>**, Brown J<sup>1</sup> and Chamberlin S\*. The Role of Aspartic Acid 101 in *E. coli* Alkaline Phosphatase Architectural Activity and Stability. 2015. Chemistry Department Poster Presentation. Regis University.

### **Presentations**

- | <u>Year</u> | <u>Presentation</u>   |
|-------------|---|
| 2022        | OCT4/POU5f1 is expressed in both pluripotent and non-pluripotent stem cells during somatic cell reprogramming. <b>Moauro A<sup>1</sup></b> , Hickey S <sup>2</sup> , Halbisen M <sup>2</sup> and Ralston A*. Michigan Osteopathic Association Spring Research Exhibit. Detroit, MI. |
| 2021        | Evaluating <i>Oct4</i> 's Role in the Formation of Induced Extraembryonic Endoderm (iXEN) Cells During Somatic Cell Reprogramming. <b>Moauro A<sup>1</sup></b> and Ralston A*. Molecular, Cellular and Integrated Physiology Research Forum. Michigan State University.             |
| 2020        | Evaluating <i>Oct4</i> 's Role in the Formation of Induced Stem Cells During Somatic Cell Reprogramming. <b>Moauro A<sup>1</sup></b> and Ralston A*. Cancer Research Network Seminar. Michigan State University.  |
| 2020        | Medical Detective: Physician Scientists at Work. <b>Moauro A<sup>1</sup></b> and Yoon S <sup>1</sup> . MSUCOM OsteoCHAMPS. Online.  |
| 2020        | Uncovering the parallels between early embryo development and reprogramming. <b>Moauro A<sup>1</sup></b> and Ralston A*. Mouse Research Development Day. Michigan State University.   |

### **Grants**

- | <u>Year</u> | <u>Grant</u>  |
|-------------|---|
| 2020        | "Evaluating the Role of Oct4 in the Formation of Induced Extraembryonic Endoderm Cells and Induced Pluripotent Stem Cells in OSKM Reprogramming". Ruth L. Kirschstein National Research Service Award (NRSA) Individual |

Fellowship. F30. Submitted to NIH, NICHD. Total requested: \$269,498.45. Role: PI. Status: Submitted.

**Service**

**Offices Held in Professional Organizations**

| <u>Year</u> | <u>Committee</u>   |
|-------------|--|
| 2022-23     | DO-PhD Student Advisory Committee, <i>Class Representative</i>             |
| 2021-22     | MSU COM Diversity, Equity, and Inclusion Committee, <i>Student Advisor</i> |
| 2021-22     | DO-PhD Student Advisory Committee, <i>Ex Officio</i>                       |
| 2020-21     | DO-PhD Student Advisory Committee, <i>Chair</i>                            |
| 2019-20     | DO-PhD Student Advisory Committee, <i>Class Representative</i>             |
| 2018-19     | MSUCOM American Physician Scientist Association, <i>President</i>          |

**Volunteering**

| <u>Year</u> | <u>Event</u>   |
|-------------|--|
| 2020-22     | Biomedical Sciences Gateway Program PhD Recruitment Weekend, <i>Moderator and Greeter</i>                        |
| 2020-22     | Creation of a written DEI section and resource compilation for the DO-PhD handbook, <i>Writer</i>                |
| 2021        | Michigan Physician Scientist Interest Day for Underrepresented Minorities, <i>Organizer and Graphic Designer</i> |
| 2021        | Underrepresented Minority Summer Research Experience, <i>Mentor</i>  |
| 2021-22     | Strategic Enrollment Management MSU COM, <i>Student Advisor</i>  |
| 2020        | MSU COM OsteoCHAMPS, <i>Speaker and Mentor</i>   |
| 2020        | Greater Lansing Community Art Project through High Caliber Karting, <i>Artist</i>                                |

**Mentorship**

| <u>Year</u> | <u>Students</u>  |
|-------------|--|
| 2021-22     | MSU senior completing an independent research project through the department of Biochemistry |
| 2021        | Rotation student in MSU BMS  |

|         |  |
|---------|--|
| 2021    | Underrepresented minority student completing a summer research experience through MSU's Reproduction and Developmental Science Program |
| 2018-19 | MSU senior gaining research experience before entering medical school  |

***Outreach***

| <u>Year</u> | <u>Event</u>  |
|-------------|---|
| 2022        | MSU COM 3+4 Program Medical Career Paths Student Panel                      |
| 2021        | Regis University AED Careers in Medicine Panel                              |
| 2021        | NIH Graduate & Professional School Fair                                     |
| 2021        | MSU COM Strategic Enrollment Management, DO-PhD representative              |
| 2021        | University of Detroit Mercy Physician Scientists Information Session        |
| 2020        | University of Moana Physician Scientist and Osteopathic Information Session |

## REFERENCES

## REFERENCES

- Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G. K., Teo, R., Leng Ng, C. K., Herath, W., Lili, S., Hutchins, A. P., Robson, P., Kolatkar, P. R., & Stanton, L. W. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO Journal*, *32*(7), 938–953. <https://doi.org/10.1038/emboj.2013.31>
- Andras Nagy, Marina Gertsenstein, K. V. and R. B. (2006). Preparing Feeder Cell Layers from STO or Mouse Embryo Fibroblast (MEF) Cells: Treatment with  $\gamma$ -Irradiation. *Cold Spring Harb Protoc.* <https://doi.org/doi:10.1101/pdb.prot4400>
- Apostolou, E., & Stadtfeld, M. (2018). Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Current Opinion in Genetics and Development*, *52*, 77–85. <https://doi.org/10.1016/j.gde.2018.06.002>
- Artus, J., Douvaras, P., Piliszek, A., Isern, J., Baron, M. H., & Hadjantonakis, A. K. (2012). BMP4 signaling directs primitive endoderm-derived XEN cells to an extraembryonic visceral endoderm identity. *Developmental Biology*, *361*(2), 245–262. <https://doi.org/10.1016/j.ydbio.2011.10.015>
- Artus, J., Piliszek, A., & Hadjantonakis, A. K. (2011). The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. *Developmental Biology*, *350*(2), 393–404. <https://doi.org/10.1016/j.ydbio.2010.12.007>
- Azami, T., Bassalert, C., Allègre, N., Estrella, L. V., Pouchin, P., Ema, M., & Chazaud, C. (2019). Regulation of the erk signalling pathway in the developing mouse blastocyst. *Development (Cambridge)*, *146*(14). <https://doi.org/10.1242/dev.177139>
- Banito, A., Rashid, S. T., Acosta, J. C., Li, S. De, Pereira, C. F., Geti, I., Pinho, S., Silva, J. C., Azuara, V., Walsh, M., Vallier, L., & Gil, J. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. *Genes and Development*, *23*(18), 2134–2139. <https://doi.org/10.1101/gad.1811609>
- Bassalert, C., Valverde-Estrella, L., & Chazaud, C. (2018). Primitive Endoderm Differentiation: From Specification to Epithelialization. *Current Topics in Developmental Biology*, *128*, 81–104. <https://doi.org/10.1016/bs.ctdb.2017.12.001>
- Behringer, R., Gertsenstein, M., Vintersten Nagy, K., & Nagy, A. (2014). Manipulating the Mouse Embryo: A Laboratory Manual. *CSH Press*, *4th*
- Belaoussoff, M., Farrington, S. M., & Baron, M. H. (1998). Hematopoietic induction and respecification of A-P identity by visceral endoderm signaling in the mouse embryo. *Development*, *125*(24), 5009–5018

- Bessonard, S., Vandormael-Pournin, S., Coqueran, S., Cohen-Tannoudji, M., & Artus, J. (2019). PDGF Signaling in Primitive Endoderm Cell Survival Is Mediated by PI3K-mTOR Through p53-Independent Mechanism. *Stem Cells*, 37(7), 888–898. <https://doi.org/10.1002/stem.3008>
- Bielinska, M., Narita, N., & Wilson, D. B. (1999). Distinct roles for visceral endoderm during embryonic mouse development. *International Journal of Developmental Biology*, 43(3), 183–205
- Bline, A. P., Goff, A. Le, & Allard, P. (2020). What is lost in the weismann barrier? *Journal of Developmental Biology*, 8(4), 1–12. <https://doi.org/10.3390/jdb8040035>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., & Jaenisch, R. (2009). *Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells*. 2(2), 151–159
- Brown, K., Doss, M. X., Legros, S., Artus, J., Hadjantonakis, A. K., & Foley, A. C. (2010). Extraembryonic endoderm (XEN) stem cells produce factors that activate heart formation. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0013446>
- Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., van Oudenaarden, A., & Jaenisch, R. (2012). Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. *Cell*, 150(6), 1209–1222. <https://doi.org/10.1016/j.cell.2012.08.023>
- Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., Van Oudenaarden, A., & Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150(6), 1209–1222. <https://doi.org/10.1016/j.cell.2012.08.023>
- Buganim, Y., Markoulaki, S., Van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B. J., Porubsky, D., Kulenkampff, E., Faddah, D. A., Shi, L., Gao, Q., Sarkar, S., Cohen, M., Goldmann, J., Nery, J. R., Schultz, M. D., ... Jaenisch, R. (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell Stem Cell*, 15(3), 295–309. <https://doi.org/10.1016/j.stem.2014.07.003>
- Cacchiarelli, D., Trapnell, C., Ziller, M. J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z. D., Ratanasirintrao, S., Zhang, X., Ho Sui, S. J., Wu, Z., Akopian, V., Gifford, C. A., Doench, J., Rinn, J. L., Daley, G. Q., Meissner, A., Lander, E. S., & Mikkelsen, T. S. (2015). Integrative Analyses of Human Reprogramming

- Reveal Dynamic Nature of Induced Pluripotency. *Cell*, 162(2), 412–424. <https://doi.org/10.1016/j.cell.2015.06.016>
- Carey, B., Markoulaki, S., Beard, C., Hanna, J., & Jaenisch, R. (2010). A single-gene transgenic mouse strain for reprogramming adult somatic cells. *Nat Methods*, 3(1), 9–14. <https://doi.org/10.1038/nmeth.1410.A>
- Castel, G., Meistermann, D., Bretin, B., Firmin, J., Blin, J., Loubersac, S., Bruneau, A., Chevolleau, S., Kilens, S., Chariau, C., Gaignerie, A., Francheteau, Q., Kagawa, H., Charpentier, E., Flippe, L., Francois-Campion, V., Haider, S., Dietrich, B., Knöfler, M., ... David, L. (2020). Generation of human induced trophoblast stem cells. *BioRxiv*, 2020.09.15.298257. <https://doi.org/10.1101/2020.09.15.298257>
- Cevallos, R. R., Edwards, Y. J. K., Parant, J. M., Yoder, B. K., & Hu, K. (2020). Human transcription factors responsive to initial reprogramming predominantly undergo legitimate reprogramming during fibroblast conversion to iPSCs. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-76705-y>
- Chan, E., Ratanasirintrao, S., Park, I., Manos, P., Loh, Y., Huo, H., Miller, J., Hartung, O., Rho, J., Ince, T., G., D., & Schlaege, T. (2009). Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotechnology*, 27(11), 1033–1038. <https://doi.org/10.1038/nbt.1580>
- 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. *Developmental Cell*, 10(5), 615–624. <https://doi.org/10.1016/j.devcel.2006.02.020>
- Chen, J., Chen, X., Li, M., Liu, X., Gao, Y., Kou, X., Zhao, Y., Zheng, W., Zhang, X., Huo, Y., Chen, C., Wu, Y., Wang, H., Jiang, C., & Gao, S. (2016). Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Reports*, 14(6), 1540–1554. <https://doi.org/10.1016/j.celrep.2016.01.013>
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., ... Ng, H. H. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*, 133(6), 1106–1117. <https://doi.org/10.1016/j.cell.2008.04.043>
- Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, 5, 1438. <https://doi.org/10.12688/f1000research.8987.1>
- Chien, K. R. (2008). Regenerative medicine and human models of human disease.

*Nature*, 453(7193), 302–305. <https://doi.org/10.1038/nature07037>

- Ching, T., Huang, S., & Garmire, L. X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. *Rna*, 20(11), 1684–1696. <https://doi.org/10.1261/rna.046011.114>
- Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., & Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*, 168(3), 442–459.e20. <https://doi.org/10.1016/j.cell.2016.12.016>
- Cowan, C. A., Atienza, J., Melton, D. A., & Eggan, K. (2005). Developmental Biology: Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science*, 309(5739), 1369–1373. <https://doi.org/10.1126/science.1116447>
- Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51(6), 987–1000. [https://doi.org/10.1016/0092-8674\(87\)90585-X](https://doi.org/10.1016/0092-8674(87)90585-X)
- Dittrich, G. M., Froese, N., Wang, X., Kroeger, H., Wang, H., Szaroszyk, M., Malek-Mohammadi, M., Cordero, J., Keles, M., Korf-Klingebiel, M., Wollert, K. C., Geffers, R., Mayr, M., Conway, S. J., Dobрева, G., Bauersachs, J., & Heineke, J. (2021). Fibroblast GATA-4 and GATA-6 promote myocardial adaptation to pressure overload by enhancing cardiac angiogenesis. *Basic Research in Cardiology*, 116(1), 1–19. <https://doi.org/10.1007/s00395-021-00862-y>
- Dos Santos, R. L., Tosti, L., Radzishewska, A., Caballero, I. M., Kaji, K., Hendrich, B., & Silva, J. C. R. (2014). Erratum: MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner (*Cell Stem Cell* (2014) 15 (102–110)). *Cell Stem Cell*, 15(3), 392. <https://doi.org/10.1016/j.stem.2014.08.005>
- Elling, U., Klasen, C., Eisenberger, T., Anlag, K., & Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(44), 16319–16324. <https://doi.org/10.1073/pnas.0607884103>
- Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819), 154–156. <https://doi.org/10.1038/292154a0>
- Feng, B., Ng, J. H., Heng, J. C. D., & Ng, H. H. (2009). Molecules that Promote or Enhance Reprogramming of Somatic Cells to Induced Pluripotent Stem Cells. *Cell Stem Cell*, 4(4), 301–312. <https://doi.org/10.1016/j.stem.2009.03.005>
- Fleming, T. P. (1987). A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. *Developmental Biology*, 119(2), 520–531. [https://doi.org/10.1016/0012-1606\(87\)90055-8](https://doi.org/10.1016/0012-1606(87)90055-8)

- Freyer, L., Schröter, C., Saiz, N., Schrode, N., Nowotschin, S., Martinez-Arias, A., & Hadjantonakis, A. K. (2015). A loss-of-function and H2B-Venus transcriptional reporter allele for Gata6 in mice Early development. *BMC Developmental Biology*, *15*(1), 1–18. <https://doi.org/10.1186/s12861-015-0086-5>
- Frum, T., Halbisen, M. A., Wang, C., Amiri, H., Robson, P., & Ralston, A. (2013). Oct4 Cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Developmental Cell*, *25*(6), 610–622. <https://doi.org/10.1016/j.devcel.2013.05.004>
- Frum, T., & Ralston, A. (2020). Culture conditions antagonize lineage-promoting signaling in the mouse blastocyst. *Reproduction*, *160*(1), V5–V7. <https://doi.org/10.1530/REP-20-0107>
- Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., & Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, *85*(8), 348–362. <https://doi.org/10.2183/pjab.85.348>
- Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., Yao, C., Liu, X., Lu, Z., Xu, Z., Kang, L., Chen, J., Wang, H., Cai, T., & Gao, S. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell*, *12*(4), 453–469. <https://doi.org/10.1016/j.stem.2013.02.005>
- Ghaleb, A. M., & Yang, V. W. (2017). Krüppel-like factor 4 (KLF4): What we currently know. *Gene*, *611*, 27–37. <https://doi.org/10.1016/j.gene.2017.02.025>
- González, F., & Huangfu, D. (2016). Mechanisms underlying the formation of induced pluripotent stem cells. *WIREs Developmental Biology*, *5*(1), 39–65. <https://doi.org/10.1002/wdev.206>
- Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D., & Robson, P. (2010). Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. *Developmental Cell*, *18*(4), 675–685. <https://doi.org/10.1016/j.devcel.2010.02.012>
- Guo, S., Zi, X., Schulz, V., Cheng, J., Zhong, M., Koochaki, S. H. J., Megyola, C. M., Pan, X., Heydari, K., Weissman, S. M., Gallagher, P. G., Krause, D. S., Fan, R., & Lu, J. (2015). *state*. *156*(4), 649–662. <https://doi.org/10.1016/j.cell.2014.01.020>. Non-stochastic
- Gurdon, J. B., ELSDALE, T. R., & FISCHBERG, M. (1958). Sexually Mature Individuals of *Xenopus laevis* from the Transplantation of Single Somatic Nuclei. *Nature*,

182(4627), 64–65. <https://doi.org/10.1038/182064a0>

- Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D., & Soriano, P. (2003). Evolutionary Divergence of Platelet-Derived Growth Factor Alpha Receptor Signaling Mechanisms. *Molecular and Cellular Biology*, 23(11), 4013–4025. <https://doi.org/10.1128/mcb.23.11.4013-4025.2003>
- Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B. S., Li, P., Lim, S. L., Cao, S., Tay, J., Orlov, Y. L., Lufkin, T., Ng, H.-H., Tam, W.-L., & Lim, B. (2010). Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature*, 463(7284), 1096–1100. <https://doi.org/10.1038/nature08735>
- Hansson, J., Rafiee, M. R., Reiland, S., Polo, J. M., Gehring, J., Okawa, S., Huber, W., Hochedlinger, K., & Krijgsveld, J. (2012). Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency. *Cell Reports*, 2(6), 1579–1592. <https://doi.org/10.1016/j.celrep.2012.10.014>
- He, X., Chi, G., Li, M., Xu, J., Zhang, L., Song, Y., Wang, L., & Li, Y. (2020). Characterisation of extraembryonic endoderm-like cells from mouse embryonic fibroblasts induced using chemicals alone. *Stem Cell Research and Therapy*, 11(1), 1–16. <https://doi.org/10.1186/s13287-020-01664-0>
- Hemberger, M., Hanna, C. W., & Dean, W. (2020). Mechanisms of early placental development in mouse and humans. *Nature Reviews Genetics*, 21(1), 27–43. <https://doi.org/10.1038/s41576-019-0169-4>
- Heng, J. C. D., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J. H., Orlov, Y. L., Huss, M., Yang, L., Lufkin, T., Lim, B., & Ng, H. H. (2010). The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells. *Cell Stem Cell*, 6(2), 167–174. <https://doi.org/10.1016/j.stem.2009.12.009>
- Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., Ge, J., Xu, J., Zhang, Q., Zhao, Y., & Deng, H. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science*, 341(6146), 651–654. <https://doi.org/10.1126/science.1239278>
- Huang, K., Zhang, X., Shi, J., Yao, M., Lin, J., Li, J., Liu, H., Li, H., Shi, G., Wang, Z., Zhang, B., Chen, J., Pan, G., Jiang, C., Pei, D., & Yao, H. (2015). Dynamically reorganized chromatin is the key for the reprogramming of somatic cells to pluripotent cells. *Scientific Reports*, 5(November), 1–14. <https://doi.org/10.1038/srep17691>
- Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A. E., & Melton, D. A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnology*, 26(7), 795–797. <https://doi.org/10.1038/nbt1418>

- Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., & Melton, D. A. (2008). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature Biotechnology*, 26(11), 1269–1275. <https://doi.org/10.1038/nbt.1502>
- Ichida, J. K., Blanchard, J., Lam, K., Son, E. Y., Chung, J. E., Egli, D., Loh, K. M., Carter, A. C., Di Giorgio, F. P., Koszka, K., Huangfu, D., Akutsu, H., Liu, D. R., Rubin, L. L., & Eggan, K. (2009). A Small-Molecule Inhibitor of Tgf- $\beta$  Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. *Cell Stem Cell*, 5(5), 491–503. <https://doi.org/10.1016/j.stem.2009.09.012>
- Jerabek, S., Merino, F., Schöler, H. R., & Cojocaru, V. (2014). OCT4: Dynamic DNA binding pioneers stem cell pluripotency. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. <https://doi.org/10.1016/j.bbagrm.2013.10.001>
- Johnson, M. H., & McConnell, J. M. L. (2004). Lineage allocation and cell polarity during mouse embryogenesis. *Seminars in Cell and Developmental Biology*, 15(5), 583–597. <https://doi.org/10.1016/j.semcd.2004.04.002>
- Judson, R. L., Babiarz, J. E., Venere, M., & Blueloch, R. (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature Biotechnology*, 27(5), 459–461. <https://doi.org/10.1038/nbt.1535>
- Kang, M., Garg, V., & Hadjantonakis, A.-K. (2017). Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2. *Developmental Cell*, 41(5), 496–510.e5. <https://doi.org/10.1016/j.devcel.2017.05.003>
- Kang, M., Piliszek, A., Artus, J., & Hadjantonakis, A. K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development (Cambridge)*, 140(2), 267–279. <https://doi.org/10.1242/dev.084996>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Knaupp, A. S., Buckberry, S., Pflueger, J., Lim, S. M., Ford, E., Larcombe, M. R., Rossello, F. J., de Mendoza, A., Alaei, S., Firas, J., Holmes, M. L., Nair, S. S., Clark, S. J., Nefzger, C. M., Lister, R., & Polo, J. M. (2017). Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell*, 21(6), 834–845.e6.

<https://doi.org/10.1016/j.stem.2017.11.007>

- Kolios, G., & Moodley, Y. (2012). Introduction to stem cells and regenerative medicine. *Respiration*, *85*(1), 3–10. <https://doi.org/10.1159/000345615>
- Kulesa, H., Programme, D., & Molecular, E. (1995). •• *Myeloblast*. 1250–1262. <https://doi.org/https://doi-org.proxy1.cl.msu.edu/10.1101/gad.9.10.1250>
- Kunath, T., Arnaud, D., Uy, G. D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R. L., Avner, P., & Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development*, *132*(7), 1649–1661. <https://doi.org/10.1242/dev.01715>
- Kwon, G. S., Viotti, M., & Hadjantonakis, A. K. (2008). The Endoderm of the Mouse Embryo Arises by Dynamic Widespread Intercalation of Embryonic and Extraembryonic Lineages. *Developmental Cell*, *15*(4), 509–520. <https://doi.org/10.1016/j.devcel.2008.07.017>
- Le Bin, G. C., Muñoz-Descalzo, S., Kurowski, A., Leitch, H., Lou, X., Mansfield, W., Etienne-Dumeau, C., Grabole, N., Mulas, C., Niwa, H., Hadjantonakis, A. K., & Nichols, J. (2014). Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development (Cambridge)*, *141*(5), 1001–1010. <https://doi.org/10.1242/dev.096875>
- Lengner, C. J., Camargo, F. D., Hochedlinger, K., Welstead, G. G., Zaidi, S., Gokhale, S., Scholer, H. R., Tomilin, A., & Jaenisch, R. (2007). Oct4 Expression Is Not Required for Mouse Somatic Stem Cell Self-Renewal. *Cell Stem Cell*, *1*(4), 403–415. <https://doi.org/10.1016/j.stem.2007.07.020>
- Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., Qin, Y., Guo, L., He, J., Yu, S., Liu, H., Wang, X., Wu, F., Kuang, J., Hutchins, A. P., Chen, J., & Pei, D. (2017). Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell*, *21*(6), 819–833.e6. <https://doi.org/10.1016/j.stem.2017.10.012>
- Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., Ge, J., Liu, C., Zhang, W., Zhang, X., Wu, Y., Li, H., Liu, K., Wu, C., Song, Z., Zhao, Y., Shi, Y., & Deng, H. (2011). Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Research*, *21*(1), 196–204. <https://doi.org/10.1038/cr.2010.142>
- 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), 543–554. <https://doi.org/10.1016/j.stem.2008.08.004>
- Linneberg-Agerholm, M., Wong, Y. F., Herrera, J. A. R., Monteiro, R. S., Anderson, K. G. V., & Brickman, J. M. (2019). Naïve human pluripotent stem cells respond to Wnt,

- Nodal, and LIF signalling to produce expandable naïve extra-embryonic endoderm. In *Development* (Issue November). <https://doi.org/10.1242/dev.180620>
- Liu, S., Bou, G., Sun, R., Guo, S., Xue, B., Wei, R., Cooney, A. J., & Liu, Z. (2015). Sox2 is the faithful marker for pluripotency in pig: Evidence from embryonic studies. *Developmental Dynamics*, 244(4), 619–627. <https://doi.org/10.1002/dvdy.24248>
- Liu, X., Ouyang, J. F., Rossello, F. J., Tan, J. P., Davidson, K. C., Valdes, D. S., Schröder, J., Sun, Y. B. Y., Chen, J., Knaupp, A. S., Sun, G., Chy, H. S., Huang, Z., Pflueger, J., Firas, J., Tano, V., Buckberry, S., Paynter, J. M., Larcombe, M. R., ... Polo, J. M. (2020). Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature*, 586(7827), 101–107. <https://doi.org/10.1038/s41586-020-2734-6>
- Lokken, A. A., & Ralston, A. (2016). The Genetic Regulation of Cell Fate During Preimplantation Mouse Development. In *Current Topics in Developmental Biology*. <https://doi.org/10.1016/bs.ctdb.2016.04.006>
- Macarthur, B. D., Sevilla, A., Lenz, M., Müller, F. J., Schuld, B. M., Schuppert, A. A., Ridden, S. J., Stumpf, P. S., Fidalgo, M., Ma'ayan, A., Wang, J., & Lemischka, I. R. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nature Cell Biology*, 14(11), 1139–1147. <https://doi.org/10.1038/ncb2603>
- Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., & Yamanaka, S. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature*, 474(7350), 225–228. <https://doi.org/10.1038/nature10106>
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K., & Hochedlinger, K. (2007). Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell*, 1(1), 55–70. <https://doi.org/10.1016/j.stem.2007.05.014>
- Malik, V., Glaser, L. V., Zimmer, D., Velychko, S., Weng, M., Holzner, M., Arend, M., Chen, Y., Srivastava, Y., Veerapandian, V., Shah, Z., Esteban, M. A., Wang, H., Chen, J., Schöler, H. R., Hutchins, A. P., Meijnsing, S. H., Pott, S., & Jauch, R. (2019). Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. *Nature Communications*, 10(1), 1–16. <https://doi.org/10.1038/s41467-019-11054-7>
- Manini, and M. C. M. P. (2008). Roles of Krüppel-like factor 4 in normal homeostasis, cancer and stem cells. *Acta Biochimica et Biophysica Sinica*, 40(7), 554–564. <https://doi.org/10.1111/j.1745-7270.2008.00439.x>
- Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured

- in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 78(12 II), 7634–7638. <https://doi.org/10.1073/pnas.78.12.7634>
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10), 4288–4297. <https://doi.org/10.1093/nar/gks042>
- Meissner, A., Wernig, M., & Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnology*, 25(10), 1177–1181. <https://doi.org/10.1038/nbt1335>
- Mikkelsen, T. S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B. E., Jaenisch, R., Lander, E. S., & Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature*, 454(7200), 49–55. <https://doi.org/10.1038/nature07056>
- Moerkamp, A. T., Paca, A., Goumans, M. J., Kunath, T., Kruithof, B. P. T., & Kruithof-de Julio, M. (2013). Extraembryonic Endoderm cells as a model of endoderm development. *Development Growth and Differentiation*, 55(3), 301–308. <https://doi.org/10.1111/dgd.12036>
- Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J. C., & Reik, W. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Reports*, 20(5), 1215–1228. <https://doi.org/10.1016/j.celrep.2017.07.009>
- Molkentin, J. D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6: Ubiquitously expressed regulators of tissue-specific gene expression. *Journal of Biological Chemistry*, 275(50), 38949–38952. <https://doi.org/10.1074/jbc.R000029200>
- Molotkov, A., Mazot, P., Brewer, J. R., Cinalli, R. M., & Soriano, P. (2017). Distinct Requirements for FGFR1 and FGFR2 in Primitive Endoderm Development and Exit from Pluripotency. *Developmental Cell*, 41(5), 511–526.e4. <https://doi.org/10.1016/j.devcel.2017.05.004>
- Morgani, S. M., & Brickman, J. M. (2015). LIF supports primitive endoderm expansion during pre-implantation development. *Development*, 142(20), 3488–3499. <https://doi.org/10.1242/dev.125021>
- Nag, A., Savova, V., Fung, H. L., Miron, A., Yuan, G. C., Zhang, K., & Gimelbrant, A. A. (2013). Chromatin signature of widespread monoallelic expression. *ELife*, 2013(2), 1–19. <https://doi.org/10.7554/eLife.01256>

- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., & Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnology*, 26(1), 101–106. <https://doi.org/10.1038/nbt1374>
- Neri, F., Zippo, A., Krepelova, A., Cherubini, A., Rocchigiani, M., & Oliviero, S. (2012). Myc Regulates the Transcription of the PRC2 Gene To Control the Expression of Developmental Genes in Embryonic Stem Cells. *Molecular and Cellular Biology*, 32(4), 840–851. <https://doi.org/10.1128/mcb.06148-11>
- Niakan, K. K., Ji, H., Maehr, R., Vokes, S. A., Rodolfa, K. T., Sherwood, R. I., Yamaki, M., Dimos, J. T., Chen, A. E., Melton, D. A., McMahon, A. P., & Eggan, K. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes and Development*. <https://doi.org/10.1101/gad.1833510>
- Niakan, K. K., Schrode, N., Cho, L. T. Y., & Hadjantonakis, A. K. (2013). Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nature Protocols*, 8(6), 1028–1041. <https://doi.org/10.1038/nprot.2013.049>
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., & Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95(3), 379–391. [https://doi.org/10.1016/S0092-8674\(00\)81769-9](https://doi.org/10.1016/S0092-8674(00)81769-9)
- Nishimura, T., Unezaki, N., Kanegi, R., Wijesekera, D. P. H., Hatoya, S., Sugiura, K., Kawate, N., Tamada, H., Imai, H., & Inaba, T. (2017). Generation of Canine Induced Extraembryonic Endoderm-Like Cell Line That Forms Both Extraembryonic and Embryonic Endoderm Derivatives. *Stem Cells and Development*, 26(15), 1111–1120. <https://doi.org/10.1089/scd.2017.0026>
- Niwa, H., Miyazaki, J. I., & Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24(4), 372–376. <https://doi.org/10.1038/74199>
- Niwa, H., Ogawa, K., Shimosato, D., & Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature*, 460(7251), 118–122. <https://doi.org/10.1038/nature08113>
- Nowotschin, S., & Hadjantonakis, A. K. (2020). Guts and gastrulation: Emergence and convergence of endoderm in the mouse embryo. *Current Topics in Developmental Biology*, 136, 429–454. <https://doi.org/10.1016/bs.ctdb.2019.11.012>
- Nowotschin, S., Setty, M., Kuo, Y. Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., Church, D. M., Hoodless, P. A., Hadjantonakis, A. K.,

- & Pe'er, D. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature*, *569*(7756), 361–367. <https://doi.org/10.1038/s41586-019-1127-1>
- Ohinata, Y., Endo, T. A., Sugishita, H., Watanabe, T., Iizuka, Y., Kawamoto, Y., Saraya, A., Kumon, M., Koseki, Y., Kondo, T., Ohara, O., & Koseki, H. (2022). Establishment of mouse stem cells that can recapitulate the developmental potential of primitive endoderm. *Science*, *375*(6580), 574–578. <https://doi.org/10.1126/science.aay3325>
- Orkin, S. H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T. W., Shen, X., & Levasseur, D. N. (2008). The transcriptional network controlling pluripotency in ES cells. *Cold Spring Harbor Symposia on Quantitative Biology*, *73*, 195–202. <https://doi.org/10.1101/sqb.2008.72.001>
- Paca, A., Séguin, C. A., Clements, M., Ryczko, M., Rossant, J., Rodriguez, T. A., & Kunath, T. (2012). BMP signaling induces visceral endoderm differentiation of XEN cells and parietal endoderm. *Developmental Biology*, *361*(1), 90–102. <https://doi.org/10.1016/j.ydbio.2011.10.013>
- Palmieri, S. L., Peter, W., Hess, H., & Schöler, 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. *Developmental Biology*, *166*(1), 259–267. <https://doi.org/10.1006/dbio.1994.1312>
- Parenti, A., Halbisen, M. A., Wang, K., Latham, K., & Ralston, A. (2016). OSKM Induce Extraembryonic Endoderm Stem Cells in Parallel to Induced Pluripotent Stem Cells. *Stem Cell Reports*, *6*(4), 447–455. <https://doi.org/10.1016/j.stemcr.2016.02.003>
- Patra, S. K. (2020). Roles of OCT4 in pathways of embryonic development and cancer progression. *Mechanisms of Ageing and Development*, *189*(December 2019), 111286. <https://doi.org/10.1016/j.mad.2020.111286>
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667. <https://doi.org/10.1038/nprot.2016.095>
- Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L., Reik, W., Srinivas, S., Simons, B. D., Nichols, J., Marioni, J. C., & Göttgens, B. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, *566*(7745), 490–495. <https://doi.org/10.1038/s41586-019-0933-9>
- Plusa, B., Frankenberg, S., Chalmers, A., Hadjantonakis, A. K., Moore, C. A., Papalopulu, N., Papaioannou, V. E., Glover, D. M., & Zernicka-Goetz, M. (2005). Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *Journal of Cell Science*, *118*(3), 505–515.

<https://doi.org/10.1242/jcs.01666>

- Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., & Hadjantonakis, A. K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development*, *135*(18), 3081–3091. <https://doi.org/10.1242/dev.021519>
- Polo, J. M., Liu, S., Figueroa, M. E., Kulalert, W., Eminli, S., Tan, K. Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., Natesan, S., Wagers, A. J., Melnick, A., Evans, T., & Hochedlinger, K. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnology*, *28*(8), 848–855. <https://doi.org/10.1038/nbt.1667>
- Pour, M., Pilzer, I., Rosner, R., Smith, Z. D., Meissner, A., & Nachman, I. (2014). Epigenetic predisposition to reprogramming fates in somatic cells. *EMBO Reports*, *16*, 370–378. <https://doi.org/10.15252/embr>
- Raab, S., Klingenstein, M., Möller, A., Illing, A., Tosic, J., Breunig, M., Kualess, G., Linta, L., Seufferlein, T., Arnold, S. J., Kleger, A., & Liebau, S. (2017). Reprogramming to pluripotency does not require transition through a primitive streak-like state. *Scientific Reports*, *7*(1), 1–10. <https://doi.org/10.1038/s41598-017-15187-x>
- Rao, S., Zhen, S., Roumiantsev, S., McDonald, L. T., Yuan, G.-C., & Orkin, S. H. (2010). Differential Roles of Sall4 Isoforms in Embryonic Stem Cell Pluripotency. *Molecular and Cellular Biology*, *30*(22), 5364–5380. <https://doi.org/10.1128/mcb.00419-10>
- Rizzino, A., & Wuebben, E. L. (2016). Sox2/Oct4: A delicately balanced partnership in pluripotent stem cells and embryogenesis. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1859*(6), 780–791. <https://doi.org/10.1016/j.bbagrm.2016.03.006>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rossant, J. (2014). Mouse and human blastocyst-derived stem cells: vive les differences. *Development*, *142*(1), 9–12. <https://doi.org/10.1242/dev.115451>
- Rowe, R. G., & Daley, G. Q. (2019). Induced Pluripotent Stem Cells in Disease Modelling and Regeneration. *Advances in Experimental Medicine and Biology*, *1144*(7), 91–99. [https://doi.org/10.1007/5584\\_2018\\_290](https://doi.org/10.1007/5584_2018_290)
- Rugg-Gunn, P. J., Cox, B. J., Ralston, A., & Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10783–10790. <https://doi.org/10.1073/pnas.0914507107>

- Schwarz, B. A., Cetinbas, M., Clement, K., Walsh, R. M., Cheloufi, S., Gu, H., Langkabel, J., Kamiya, A., Schorle, H., Meissner, A., Sadreyev, R. I., & Hochedlinger, K. (2018). Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming. *Cell Stem Cell*, 23(2), 289–305.e5. <https://doi.org/10.1016/j.stem.2018.06.013>
- Serrano, F., Calatayud, C. F., Blazquez, M., Torres, J., Castell, J. V., & Bort, R. (2013). Gata4 blocks somatic cell reprogramming by directly repressing Nanog. *Stem Cells*, 31(1), 71–82. <https://doi.org/10.1002/stem.1272>
- Shanak, S., & Helms, V. (2020). DNA methylation and the core pluripotency network. *Developmental Biology*, 464(2), 145–160. <https://doi.org/10.1016/j.ydbio.2020.06.001>
- Shi, Y., Despons, C., Do, J. T., Hahm, H. S., Schöler, H. R., & Ding, S. (2008). Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Oct4 and Klf4 with Small-Molecule Compounds. *Cell Stem Cell*, 3(5), 568–574. <https://doi.org/10.1016/j.stem.2008.10.004>
- Shu, J., Zhang, K., Zhang, M., Yao, A., Shao, S., Du, F., Yang, C., Chen, W., Wu, C., Yang, W., Sun, Y., & Deng, H. (2015a). GATA family members as inducers for cellular reprogramming to pluripotency. *Cell Research*, 25(2), 169–180. <https://doi.org/10.1038/cr.2015.6>
- Shu, J., Zhang, K., Zhang, M., Yao, A., Shao, S., Du, F., Yang, C., Chen, W., Wu, C., Yang, W., Sun, Y., & Deng, H. (2015b). GATA family members as inducers for cellular reprogramming to pluripotency. *Cell Research*, 25(2), 169–180. <https://doi.org/10.1038/cr.2015.6>
- Simon, C. S., Rahman, S., Raina, D., Schröter, C., & Hadjantonakis, A. (2021). reveals lineage specific signaling dynamics. 55(3), 341–353. <https://doi.org/10.1016/j.devcel.2020.09.030>.Live
- Simon, C. S., Zhang, L., Wu, T., Cai, W., Saiz, N., Nowotschin, S., Cai, C. L., & Hadjantonakis, A. K. (2018). A Gata4 nuclear GFP transcriptional reporter to study endoderm and cardiac development in the mouse. *Biology Open*, 7(12), 1–11. <https://doi.org/10.1242/bio.036517>
- Smith, K. N., Singh, A. M., & Dalton, S. (2010). Myc represses primitive endoderm differentiation in pluripotent stem cells. *Cell Stem Cell*, 7(3), 343–354. <https://doi.org/10.1016/j.stem.2010.06.023>
- Soufi, A., Donahue, G., & Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, 151(5), 994–1004. <https://doi.org/10.1016/j.cell.2012.09.045>

- Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, *161*(3), 555–568. <https://doi.org/10.1016/j.cell.2015.03.017>
- Spearman, C. (2010). The proof and measurement of association between two things. *International Journal of Epidemiology*, *39*(5), 1137–1150. <https://doi.org/10.1093/ije/dyq191>
- Stern, C. D., & Downs, K. M. (2012). The hypoblast (visceral endoderm): An evo-devo perspective. *Development*, *139*(6), 1059–1069. <https://doi.org/10.1242/dev.070730>
- Stirparo, G. G., Kurowski, A., Yanagida, A., Bates, L. E., Strawbridge, S. E., Hladkou, S., Stuart, H. T., Boroviak, T. E., Silva, J. C. R., & Nichols, J. (2021). OCT4 induces embryonic pluripotency via STAT3 signaling and metabolic mechanisms. *Proceedings of the National Academy of Sciences*, *118*(3). <https://doi.org/10.1073/pnas.2008890118>
- 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 trophoblast in the mouse blastocyst. *Development*, *132*(9), 2093–2102. <https://doi.org/10.1242/dev.01801>
- Stuckey, D. W., Di Gregorio, A., Clements, M., & Rodriguez, T. A. (2011). Correct patterning of the primitive streak requires the anterior visceral endoderm. *PLoS ONE*, *6*(3), 1–9. <https://doi.org/10.1371/journal.pone.0017620>
- Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., & Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Current Biology*, *11*(19), 1553–1558. [https://doi.org/10.1016/S0960-9822\(01\)00459-6](https://doi.org/10.1016/S0960-9822(01)00459-6)
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., & Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nature Communications*, *5*. <https://doi.org/10.1038/ncomms4678>
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A., & Rossant, J. (1998). Promotion to trophoblast stem cell proliferation by FGF4. *Science*, *282*(5396), 2072–2075. <https://doi.org/10.1126/science.282.5396.2072>
- Thomas, P., & Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Current Biology*, *6*(11), 1487–1496. [https://doi.org/10.1016/S0960-9822\(96\)00753-1](https://doi.org/10.1016/S0960-9822(96)00753-1)

- Tonge, P. D., Corso, A. J., Monetti, C., Hussein, S. M. I., Puri, M. C., Michael, I. P., Li, M., Lee, D. S., Mar, J. C., Cloonan, N., Wood, D. L., Gauthier, M. E., Korn, O., Clancy, J. L., Preiss, T., Grimmond, S. M., Shin, J. Y., Seo, J. S., Wells, C. A., ... Nagy, A. (2014). Divergent reprogramming routes lead to alternative stem-cell states. *Nature*, *516*(7530), 192–197. <https://doi.org/10.1038/nature14047>
- Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M., & Yamanaka, S. (2009). Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. *Genes to Cells*, *14*(6), 683–694. <https://doi.org/10.1111/j.1365-2443.2009.01301.x>
- Velychko, S., Adachi, K., Kim, K., Hou, Y., MacCarthy, C. M., Wu, G., & Schöler, H. R. (2019). Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs. *Cell Stem Cell*, 1–17. <https://doi.org/10.1016/j.stem.2019.10.002>
- Viswanathan, S., Benatar, T., Mileikovsky, M., Lauffenburger, D. A., Nagy, A., & Zandstra, P. W. (2003). Supplementation-Dependent Differences in the Rates of Embryonic Stem Cell Self-Renewal, Differentiation, and Apoptosis. *Biotechnology and Bioengineering*, *84*(5), 505–517. <https://doi.org/10.1002/bit.10799>
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature Publishing Group*, *150*(3811), 563–565. <https://www-nature-com.proxy1.cl.msu.edu/articles/150563a0>
- Wang, B., Wu, L., Li, D., Liu, Y., Guo, J., Li, C., Yao, Y., Wang, Y., Zhao, G., Wang, X., Fu, M., Liu, H., Cao, S., Wu, C., Yu, S., Zhou, C., Qin, Y., Kuang, J., Ming, J., ... Pei, D. (2019). Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Jdp2-Jhdm1b-Mkk6-Glis1-Nanog-Essrb-Sall4. *Cell Reports*, *27*(12), 3473–3485.e5. <https://doi.org/10.1016/j.celrep.2019.05.068>
- Wicklow, E., Blij, S., Frum, T., Hirate, Y., Lang, R. A., Sasaki, H., & Ralston, A. (2014). HIPPO Pathway Members Restrict SOX2 to the Inner Cell Mass Where It Promotes ICM Fates in the Mouse Blastocyst. *PLoS Genetics*, *10*(10). <https://doi.org/10.1371/journal.pgen.1004618>
- Xenopoulos, P., Kang, M., Puliafito, A., DiTalia, S., & Hadjantonakis, A. K. (2015). Heterogeneities in nanog expression drive stable commitment to pluripotency in the mouse blastocyst. *Cell Reports*, *10*(9), 1508–1520. <https://doi.org/10.1016/j.celrep.2015.02.010>
- Xiao, X., Li, N., Zhang, D., Yang, B., Guo, H., & Li, Y. (2016). Generation of Induced Pluripotent Stem Cells with Substitutes for Yamanaka's Four Transcription Factors. *Cellular Reprogramming*, *18*(5), 281–297. <https://doi.org/10.1089/cell.2016.0020>
- Xie, H., Ye, M., Feng, R., & Graf, T. (2004). Stepwise reprogramming of B cells into

- macrophages. *Cell*, 117(5), 663–676. [https://doi.org/10.1016/S0092-8674\(04\)00419-2](https://doi.org/10.1016/S0092-8674(04)00419-2)
- Xing, Q. R., El Farran, C. A., Gautam, P., Chuah, Y. S., Warriar, T., Toh, C. X. D., Kang, N. Y., Sugii, S., Chang, Y. T., Xu, J., Collins, J. J., Daley, G. Q., Li, H., Zhang, L. F., & Loh, Y. H. (2020). Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Science Advances*, 6(37), 1–18. <https://doi.org/10.1126/sciadv.aba1190>
- Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. *Nature*, 460(7251), 49–52. <https://doi.org/10.1038/nature08180>
- Yamanaka, S. (2020). Pluripotent Stem Cell-Based Cell Therapy—Promise and Challenges. *Cell Stem Cell*, 27(4), 523–531. <https://doi.org/10.1016/j.stem.2020.09.014>
- Yamanaka, Y., Lanner, F., & Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development*, 137(5), 715–724. <https://doi.org/10.1242/dev.043471>
- Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*, 154(6), 1370–1379. <https://doi.org/10.1016/j.cell.2013.08.022>
- Yoshida, G. J. (2018). Emerging roles of Myc in stem cell biology and novel tumor therapies. *Journal of Experimental and Clinical Cancer Research*, 37(1), 1–20. <https://doi.org/10.1186/s13046-018-0835-y>
- Zakrzewski, W., Dobrzyński, M., Szymonowicz, M., & Rybak, Z. (2019). Stem cells: Past, present, and future. *Stem Cell Research and Therapy*, 10(1), 1–22. <https://doi.org/10.1186/s13287-019-1165-5>
- 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. *Nature Cell Biology*, 8(10), 1114–1123. <https://doi.org/10.1038/ncb1481>
- Zhang, P., Andrianakos, R., Yang, Y., Liu, C., & Lu, W. (2010). Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *Journal of Biological Chemistry*, 285(12), 9180–9189. <https://doi.org/10.1074/jbc.M109.077958>
- Zhang, X., Hua, R., Wang, X., Huang, M., Gan, L., Wu, Z., Zhang, J., Wang, H., Cheng, Y., Li, J., & Guo, W. (2016). Identification of stem-like cells and clinical significance

of candidate stem cell markers in gastric cancer. *Oncotarget*, 7(9), 9815–9831. <https://doi.org/10.18632/oncotarget.6890>

Zhang, Y., Park, C., Bennett, C., Thornton, M., & Kim, D. (2021). Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Research*, 31(7), 1290–1295. <https://doi.org/10.1101/gr.275193.120>

Zhao, X. Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C. L., Ma, Q. W., Wang, L., Zeng, F., & Zhou, Q. (2009). IPS cells produce viable mice through tetraploid complementation. *Nature*, 461(7260), 86–90. <https://doi.org/10.1038/nature08267>

Zhao, Y., Zhao, T., Guan, J., Zhang, X., Fu, Y., Ye, J., Zhu, J., Meng, G., Ge, J., Yang, S., Cheng, L., Du, Y., Zhao, C., Wang, T., Su, L., Yang, W., & Deng, H. (2015). A XEN-like State Bridges Somatic Cells to Pluripotency during Chemical Reprogramming. *Cell*. <https://doi.org/10.1016/j.cell.2015.11.017>

Zhong, Y., Choi, T., Kim, M., Jung, K. H., Chai, Y. G., & Binas, B. (2018). Isolation of primitive mouse extraembryonic endoderm (pXEN) stem cell lines. *Stem Cell Research*, 30(March), 100–112. <https://doi.org/10.1016/j.scr.2018.05.008>

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasadhan, R., Lin, T., Kim, J., Zhang, K., & Ding, S. (2010). Reprogramming of Human Primary Somatic Cells by OCT4 and Chemical Compounds. *Cell Stem Cell*, 7(6), 651–655. <https://doi.org/10.1016/j.stem.2010.11.015>

Zviran, A., Mor, N., Rais, Y., Gingold, H., Peles, S., Chomsky, E., Manor, S., Krupalnik, V., Zerbib, M., Hezroni, H., Jaitin, D. A., Larastiaso, D., Gilad, S., Benjamin, S., Gafni, O., Mousa, A., Sheban, D., Bayerl, J., Castrejon, A. A., ... William, J. (2021). *Europe PMC Funders Group Deterministic Somatic Cell Reprogramming Involves Continuous Transcriptional Changes Governed by Myc and Epigenetic- Driven Modules*. 24(2), 328–341. <https://doi.org/10.1016/j.stem.2018.11.014>.Deterministic