Identifying Novel MicroRNA Enhancers of Somatic Cell Reprogramming

Andrew J. Corso

The Institute of Medical Science
University of Toronto
September 2013

A thesis submitted in conformity with the requirements for the degree of Master of Science
Institute of Medical Sciences
University of Toronto

© Copyright Andrew Corso 2013
Identifying Novel MicroRNA Enhancers of Somatic Cell Reprogramming

Andrew Corso

Master of Science

Institute of Medical Sciences
University of Toronto

2013

Abstract

In addition to the well-characterized Induced Pluripotent Stem cells (iPSCs) that closely resemble Embryonic Stem cells (ESCs), a recent study has proven the existence of a stable state, resembling partially reprogrammed cells, termed F-class iPSCs. To study these distinct iPSC states, a reprogramming dataset has been generated, featuring the parallel analysis of multiple molecular platforms. MicroRNAs (miRNAs) are small RNA regulators of gene expression whose critical role in reprogramming is now being realized. In the present study, small RNA deep sequencing data from this novel reprogramming dataset was used to identify miRNAs that are likely to enhance reprogramming by detecting significantly up-regulated miRNAs in ESC-like iPSCs versus F-class iPSCs. These candidate miRNAs were cloned and overexpressed in reprogramming mouse embryonic fibroblasts and their effect on reprogramming efficiency was measured. miR-214 was discovered to increase iPSC generation efficiency, marking the first reprogramming-related role for this microRNA.
Acknowledgements

Firstly, I would like to thank Dr. Andras Nagy for welcoming me into his laboratory and allowing me the opportunity to explore this exciting field of research, as well as for the guidance and support that he has provided me with throughout my project. You kept a light-hearted demeanor while helping me through the toughest times of my project. To Dr. Peter Tonge, thank you for overseeing my project throughout my Master’s program, for providing guidance and valuable insight for my research, and for the critical reading of the literature review section of this thesis. Your demonstration of patience while teaching me the basics of scientific research was nothing short of remarkable. To Mira Puri, thank you for giving me your expert advice on everything from experimental methodology to project direction. To Samer Hussein, thank you for being available for guidance whenever needed, and for always managing to help me out with my software related questions. To Rebecca Cowling, thank you for answering the laboratory questions I was too embarrassed to ask anyone else and for critically editing the literature review and discussion sections of this thesis, without holding any punches. To Claudio Monetti, thank you for answering my sometimes naïve questions around the lab, and for your critical analysis of multiple sections of this thesis. Thank you to everyone in the Nagy laboratory for your support throughout the past 2 years. To Jennifer Clancy, thank you for providing data used in this project and for helping me out with all of my microRNA-related questions. To my committee members, Dr. Jason Moffat and Dr. Quaid Morris, thank you for providing essential feedback for my project, as well as encouraging me throughout my time as a Master’s student. The bits of advice you gave me were amazingly helpful. To Dr. Ian Rogers, thank you for stepping in as a replacement committee member with minimal notice, and providing your instant feedback and encouragement, especially during the production of this thesis. Finally, to my friends and family, thank you for simply dealing with me and my mundane, irritating (and sometimes unconscious) laboratory ramblings. Without the help of all of you, I could not have completed this project.
# Table of Contents

Abstract .......................................................................................................................... ii  
Acknowledgements ......................................................................................................... iii  
List of Abbreviations ....................................................................................................... ix  

## Chapter 1: Literature Review ...................................................................................... 1  
1.1 Reprogramming ......................................................................................................... 1  
   1.1.1 Embryonic Stem Cells .......................................................................................... 2  
   1.1.2 The Path to Induced Pluripotent Stem Cells ......................................................... 3  
   1.1.3 Defining Pluripotency in iPSCs ............................................................................ 5  
   1.1.4 Generation of iPSCs ............................................................................................ 5  
      1.1.4.1 Reprogramming Factor Cocktails .................................................................... 6  
      1.1.4.2 Reprogramming Techniques .......................................................................... 6  
   1.1.5 Molecular Mechanisms of Reprogramming .......................................................... 9  
      1.1.5.1 Stoichiometry of the Reprogramming Factors ................................................ 10  
      1.1.5.2 Phases of Reprogramming ............................................................................ 10  
      1.1.5.3 Reprogramming Milestones .......................................................................... 12  
      1.1.5.4 Downstream Effects of Reprogramming Factors ............................................ 12  
      1.1.5.5 Epigenetics of Reprogramming ..................................................................... 13  
   1.1.6 Partially Reprogrammed Cells ............................................................................ 16  
   1.1.7 Inducible Systems for Studying Reprogramming ............................................... 17  
1.2 MicroRNA ................................................................................................................ 20  
   1.2.1 The Path to MicroRNA Maturity ......................................................................... 20  
   1.2.2 MicroRNA Targets ............................................................................................. 23  
   1.2.3 MicroRNAs and Pluripotency ............................................................................. 23  
   1.2.4 MicroRNAs that Affect Reprogramming Efficiency ........................................... 24
Chapter 2: Aims & Hypothesis ................................................. 29

Chapter 3: Materials & Methods ............................................. 33
3.1 Work Done Prior to Thesis ................................................. 33
   3.1.1 Primary and Secondary Mouse Embryonic Fibroblast (MEF) Derivation ................................................................. 33
   3.1.2 Secondary Reprogramming Cell Culture .......................... 34
   3.1.3 Multi-platform OMICS analysis .................................... 34
   3.1.4 MicroRNA Sequencing ............................................... 35
3.2 Work Done During Thesis ................................................. 37
   3.2.1 Transforming Chemically Competent Bacterial Cells .......... 37
   3.2.2 Plasmid Vector Construction ....................................... 37
   3.2.3 Plasmid Transfection and Cell Culture .......................... 38
   3.2.4 Alkaline Phosphatase (AP) Stain ................................ 39
   3.2.5 Three-factor iPSCs .................................................. 39
   3.2.6 Reverse Transcription of MicroRNA ............................. 40
   3.2.7 MicroRNA qPCR .................................................... 40
   3.2.8 Statistical Analysis .................................................. 40

Chapter 4: Results .................................................................. 45
4.1 Multiple iPSC States .......................................................... 45
4.2 Modeling iPSC States with a Secondary Reprogramming System ............................................. 46
4.3 Optimizing MicroRNA RT-qPCR Methods ............................. 53
   4.3.1 Reverse Transcription .............................................. 54
   4.3.2 qPCR Primer Design .............................................. 54
   4.3.3 Internal Controls .................................................... 55
4.4 Validating the Role of miR-302/367 Cluster in Reprogramming ...................................................... 60
   4.4.1 Cloning the 302-Cluster ........................................... 60
   4.4.2 Overexpression of the 302-Cluster in Reprogramming Cells .................................................. 61
4.5 Screening Small RNA Sequencing Data for Candidate MicroRNAs ..........62
  4.5.1 Identifying Candidate MicroRNAs ...........................................62
  4.3.4 Validation of Small RNA Deep Sequencing Data .........................63

4.6 Validation of Candidate MicroRNAs in the Primary Reprogramming System 71
4.7 Elucidating the Effects of Candidate MicroRNAs on Reprogramming .......71

Chapter 5: Discussion .................................................................77
5.1 Optimization of MicroRNA RT-qPCR ...........................................78
5.2 Identification of Candidate MicroRNAs .........................................80
5.3 Validation in a Primary System ....................................................81
5.4 Effects of Candidate MicroRNAs on Reprogramming ......................82
  5.4.1 miR-214 as a Novel Enhancer of Reprogramming ......................82
  5.4.2 The Reprogramming Effects of the 302-Cluster ..........................84
  5.4.3 Potential Roles of other Candidate MicroRNAs .........................84
5.5 Analysis of the Hypothesis ..........................................................85

Chapter 6: Conclusion & Future Directions .................................86

References ....................................................................................91
List of Tables

Table 1. MicroRNAs that Impact Reprogramming ..............................................28
Table 2. List of Primers Used for RT, qPCR and MicroRNA Cloning ...............41
Table 3. Contributions of Collaborators to the Results of the Thesis ..............76
List of Figures

Figure 1. The Secondary Reprogramming System.......................................................19
Figure 2. A Schematic Diagram of the Processing and Gene Regulation
  Mechanisms of MicroRNA ...........................................................................22
Figure 3. Small RNA Deep Sequencing Data Processing Pipeline......................36
Figure 4. Fibroblasts Reprogram to Predictable Stable States.............................48
Figure 5. Secondary Reprogramming Model of iPSC States...............................50
Figure 6. Quantitative RT-PCR Optimization for MicroRNAs..........................56
Figure 7. Internal Normalizers for MicroRNA qPCR.........................................58
Figure 8. Validation of Small RNA Deep Sequencing Data with MicroRNA
  qPCR ...........................................................................................................64
Figure 9. Cloning of MicroRNAs and MicroRNA clusters..................................66
Figure 10. The 302-Cluster and Reprogramming ............................................67
Figure 11. 302-Cluster Expression in the Secondary Reprogramming System...68
Figure 12. Deriving Candidate MicroRNAs from Small RNA Deep Sequencing
  Data ............................................................................................................69
Figure 13. Candidate MicroRNA Levels in Primary iPSCs.................................74
Figure 14. Overexpression of Candidate MicroRNAs in Reprogramming MEFs..75
## List of Abbreviations

<table>
<thead>
<tr>
<th>Description</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkaline Phosphatase</td>
<td>AP</td>
</tr>
<tr>
<td>5’-azacytidine</td>
<td>AZA</td>
</tr>
<tr>
<td>Base Pairs</td>
<td>bp</td>
</tr>
<tr>
<td>Bone morphogenetic protein</td>
<td>BMP</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>C. elegans</td>
</tr>
<tr>
<td>Complementary DNA</td>
<td>cDNA</td>
</tr>
<tr>
<td>Chromatin immunoprecipitation</td>
<td>ChIP</td>
</tr>
<tr>
<td><em>Drosophila Melanogaster</em></td>
<td>D. Melanogaster</td>
</tr>
<tr>
<td>DiGeorge syndrome critical region 8</td>
<td>Dcgr8</td>
</tr>
<tr>
<td>Double distilled H$_2$O</td>
<td>dd H$_2$O</td>
</tr>
<tr>
<td>Dulbecco's Modified Eagle Medium</td>
<td>DMEM</td>
</tr>
<tr>
<td>Differentially Methylated Region</td>
<td>DMR</td>
</tr>
<tr>
<td>Deoxyribonucleic acid</td>
<td>DNA</td>
</tr>
<tr>
<td>Doxycycline</td>
<td>dox</td>
</tr>
<tr>
<td>Days post coitum</td>
<td>dpc</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>E. Coli</td>
</tr>
<tr>
<td>Ethylenediaminetetraacetic (Edetic) acid</td>
<td>EDTA</td>
</tr>
<tr>
<td>Embryonic Stem cell</td>
<td>ESC</td>
</tr>
<tr>
<td>Embryonic Stem cell cycle-regulating</td>
<td>ESSC</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Fetal Bovine Serum</td>
<td>FBS</td>
</tr>
<tr>
<td>Green Fluorescent Protein</td>
<td>GFP</td>
</tr>
<tr>
<td>Guanosine triphosphate</td>
<td>GTP</td>
</tr>
<tr>
<td>Histone 3 Lysine 4 tri-methyl</td>
<td>H3K4me3</td>
</tr>
<tr>
<td>Histone Deacetylase</td>
<td>HDAC</td>
</tr>
<tr>
<td>Human Embryonic Stem cell</td>
<td>hESC</td>
</tr>
<tr>
<td>Human Leukocyte Antigen</td>
<td>HLA</td>
</tr>
<tr>
<td>Induced pluripotent stem cell</td>
<td>iPSC</td>
</tr>
<tr>
<td>Internal ribosomal entry site</td>
<td>IRES</td>
</tr>
<tr>
<td>Kilobase</td>
<td>kb</td>
</tr>
<tr>
<td>Liquid Broth</td>
<td>LB</td>
</tr>
<tr>
<td>Leukemia Inhibitory Factor</td>
<td>LIF</td>
</tr>
<tr>
<td>Mouse embryonic fibroblasts</td>
<td>MEFs</td>
</tr>
<tr>
<td>MicroRNA</td>
<td>miRNA; miR</td>
</tr>
<tr>
<td>Messenger RNA</td>
<td>mRNA</td>
</tr>
<tr>
<td>Nucleotides</td>
<td>nt</td>
</tr>
<tr>
<td>Refers to platforms of molecular analysis (transcriptomics, epigenomics, etc.)</td>
<td>OMICs</td>
</tr>
<tr>
<td>Polyadenylation sequence</td>
<td>pA</td>
</tr>
<tr>
<td>piggyBac (transposon)</td>
<td>PB</td>
</tr>
<tr>
<td>Phosphate Buffer Solution</td>
<td>PBS</td>
</tr>
<tr>
<td>Polymerase Chain Reaction</td>
<td>PCR</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Primary transcript miRNA</td>
<td>pri-miRNA</td>
</tr>
<tr>
<td>RNA-induced Silencing Complex</td>
<td>RISC</td>
</tr>
<tr>
<td>Ribonucleic acid</td>
<td>RNA</td>
</tr>
<tr>
<td>Reverse transcription</td>
<td>RT</td>
</tr>
<tr>
<td>Reverse transcription Polymerase Chain Reaction</td>
<td>RT-PCR</td>
</tr>
<tr>
<td>Reverse tetracycline transcriptional activator</td>
<td>rtTA</td>
</tr>
<tr>
<td>Sleeping Beauty (transposon)</td>
<td>SB</td>
</tr>
<tr>
<td>Somatic Cell Nuclear Transfer</td>
<td>SCNT</td>
</tr>
<tr>
<td>Small Nucleolar RNA</td>
<td>snoRNA</td>
</tr>
<tr>
<td>Tetracycline operator</td>
<td>tetO</td>
</tr>
<tr>
<td>Untranslated region</td>
<td>UTR</td>
</tr>
<tr>
<td>Valproic acid</td>
<td>VPA</td>
</tr>
</tbody>
</table>
Chapter 1:
Literature review

1.1 Reprogramming

Organ transplantation is to date the only method of treatment for millions of individuals who suffer from organ failure-related diseases such as liver cirrhosis, heart failure, and cystic fibrosis. Replacing a failing organ with a functional and compatible match can significantly increase the longevity and improve the quality of life for individuals with a large variety of different illnesses. However, donor registration is low and compatible organs are difficult to find. Thus, many individuals still succumb to their illnesses before a suitable donor is found. In order to combat this problem, the field of regenerative medicine has made numerous strides to provide these patients with an alternative treatment to organ transplantation.
1.1.1 Embryonic Stem Cells

One of the most promising areas of research within the field of regenerative medicine is Embryonic Stem Cell (ESC) biology. During development, pluripotency only exists in cells of the inner cell mass for a brief period, yet this pluripotent state can be maintained by culturing these cells in vitro. ESCs have the capacity, like the inner cell mass, to contribute to every tissue and cell type in the body. Evans and Kaufman were the first to isolate these mouse derived ESCs in 1981 (Evans and Kaufman, 1981), a study which led to Martin Evans sharing the Nobel Prize in Physiology or Medicine in 2007. Due to the difficulty of obtaining ESCs from humans, it wasn’t until 1998 that James Thomson published the world’s first human ESC line by manipulating pre-implantation blastocysts donated by couples receiving fertility clinic treatment (Thomson et al., 1998). These cells, like mouse ESCs, can be differentiated into all three germ layers, and have the ability to self renew, allowing for indefinite expansion (Reubinoff et al., 2000). Excitingly, hESCs have now entered stage 1 clinical trials to treat both spinal cord injuries (Chapman and Scala, 2012) (a trial which has since been discontinued due to changes in business strategy), and age-related macular degeneration (Schwartz et al., 2012). These trials take advantage of immune privileged sites in the central nervous system and the eye, respectively.

Theoretically, the therapeutic use of human ESCs could address the issue of organ availability. However, a major stumbling block continues to be immunogenic compatibility between the donor cells and the recipient. The human body recognizes foreign cells by detecting Human Leukocyte Antigen (HLA) markers on the cell surface. In order to avoid rejection of the donor organs/cells by the body, the recipient must be a close HLA match to the donor. The idea of setting up a human ESC bank with a full complement of HLA types has been proposed (Nakajima et al., 2007; Taylor et al., 2005), but financial and ethical
issues have prevented this from moving forward. Ideally, a patient’s own cells would be best to treat their affliction, as this would render any immunogenic issues obsolete whilst bypassing ethical concerns. In 2006, the discovery of somatic cell reprogramming made this possible, and revolutionized not just the field of stem cell research but regenerative medicine as a whole.

1.1.2 The Path to Induced Pluripotent Stem Cells

Half a century ago, the first reports describing cells that could switch their lineage by changing the microenvironment were published. These studies showed that in *D. melanogaster*, cells originally destined to contribute to the genital ridge could give rise to the wings, legs, or head when transplanted into the appropriate area of the pupae (Gehring, 1967; Hadorn, 1966). In higher vertebrates, cells extracted from the neural crest of quails were transplanted into chickens, and when traced, the tissue to which the cells contributed was only dependent on the environment in which they were placed (Le Lievre and Le Douarin, 1975). These seminal studies founded the hypothesis that cell states are plastic and subject to change, unlike the previous theory that committed cells were solely destined to their terminal state. This laid the foundation for the induced transition from a lineage-specific state to a pluripotent one.

Reprogramming the nucleus of a somatic cell to a pluripotent state has been the subject of many different studies. Three approaches were used to explore transition of a differentiated cell to a stem cell-like state: fusion of ESCs to somatic cells, somatic cell nuclear transfer (SCNT), and forced expression of transcription factors. Although the nature of cell fusion creates a tetraploid cell, incapable of *in vivo* study, it served as a useful stepping-stone towards exploring the mechanisms at play in a reprogramming cell (Blau et al., 1983). SCNT is the process by which the nucleus of a somatic cell is injected into an enucleated
oocyte. This experiment was first performed in *Xenopus* by John Gurdon’s group in 1958 (Gurdon et al., 1958). These findings reinforce the idea that the state of a differentiated cell is not terminal, and that differentiation during development could be a reversible transition. Furthermore, they prove that factors exist within the cytoplasm of the oocyte that possess the ability to reprogram the differentiated nucleus. Since SCNT was demonstrated in *Xenopus*, several organisms have been cloned using SCNT, such as sheep (Wilmut et al., 1997), mice (Wakayama et al., 1998), pigs (Polejaeva et al., 2000), cats (Shin et al., 2002), and cows (Forsberg et al., 2002). Despite this widespread success, SCNT is a technically difficult procedure, and it was not until recently that SCNT was successfully used to generate human ESCs (Tachibana et al., 2013). Yet, SCNT has served as a fantastic tool to study and develop the concepts of cell fate, identity, and plasticity.

The most recent method of inducing somatic cells to reprogram to a pluripotent state is through forced expression of transcription factors. This method was first demonstrated in Shinya Yamanaka’s laboratory in 2006 (Takahashi and Yamanaka, 2006) in a landmark experiment in which 24 candidate ESC-specific genes were systematically overexpressed in mouse embryonic fibroblasts (MEFs) to identify factors necessary for reversion to an ESC-like state. By forced expression of four transgenes (c-Myc, Klf4, Oct4, and Sox2, referred to as the Yamanaka factors), it was possible to induce fibroblast reprogramming to a cell state that resembled ESCs, which Yamanaka named induced pluripotent stem cells (iPSCs). It is the generation of these cells that has caused a great deal of excitement within the field of regenerative medicine. They offer a source for patient-specific pluripotent stem cells that have the potential to be used therapeutically and model the progression of human disease, as has been done in Machado-Joseph Disease (Koch et al., 2012), Spinal Muscular Atrophy (Chang et al., 2011; Ebert et al., 2008; Sareen et al., 2012), Familial Dysautonomia (Lee et al., 2010), and Huntington’s Disease (Camnasio et al.,...
2012; Consortium, 2012; Zhang et al., 2010), among others. It is for their contributions towards the reprogramming of somatic cells to pluripotency that John Gurdon and Shinya Yamanaka shared the 2012 Nobel Prize in Medicine or Physiology.

1.1.3 Defining Pluripotency in iPSCs

When studying iPSCs, it is important to have a stringent method of classifying the reprogrammed cells as truly pluripotent. The pluripotent capacity of murine iPSCs is demonstrated by the successful generation of live-born mice via contribution to a chimeric embryo or tetraploid aggregation, and the ability to contribute to the germline (Zhao et al., 2010; Okita et al., 2007; Maherali et al., 2007). While this is the currently accepted gold standard for demonstrating pluripotency, it is not ethically feasible in humans. Thus the authentication of human iPSCs is limited to ability to form teratomas when cells are injected into immunodeficient mice. For the iPSCs to be considered truly pluripotent, the teratomas must form complex tissue representing each of the three germ layers (endoderm, mesoderm, and ectoderm).

1.1.4 Generation of iPSCs

Since it was first published that c-Myc, Klf4, Oct4, and Sox2 could reprogram somatic cells, several groups have shown that the exact methodology used to generate iPSCs is subject to change and, in some cases, even refinement. The combination of factors as well as the method of delivery can all be changed while still producing iPSCs.
1.1.4.1 Reprogramming Factor Cocktails

Multiple combinations of reprogramming factors have been shown to induce reprogramming in various cell types. In mouse fibroblasts, the combination of Oct4, Sox2, and Esrrb induces reprogramming (Feng et al., 2009), as well as Klf4, Sox2, and Nr5a2, albeit at a lower efficiency (Heng et al., 2010). In humans, Lin28 and Nanog can replace Klf4 and c-Myc in the reprogramming of fibroblasts (Junying Yu et al., 2007). The discovery that c-Myc is expendable for reprogramming is important for the future therapeutic use of iPSCs, as c-Myc is a known oncogene, and its reactivation in vivo could result in the development of tumours (Nakagawa et al., 2007; Wernig et al., 2008b). In mice, the reactivation of c-Myc within four-factor iPSCs is believed to be the primary cause of tumour formation (Okita et al., 2007). However, dispensing c-Myc comes at a cost, as these three-factor iPSCs are generated at a lower efficiency (approximately 1 order of magnitude) than four-factor iPSCs. Furthermore, cell types that endogenously express genes required for reprogramming can be reprogrammed without the addition of that factor. For example, neural cells naturally express Sox2 at high levels, and thus the addition of just two factors, Oct4 and Klf4, can induce reprogramming (Kim et al., 2009b) (Kim et al., 2009c).

1.1.4.2 Reprogramming Techniques

Yamanaka’s original iPSCs were derived from fibroblasts that had been transfected with retroviral vectors. This transfection system has the advantage of being naturally silenced as the reprogramming cell achieves pluripotency (Ellis, 2005), allowing for exogenous expression-free progression to a pluripotent state. However, retroviral genes insert randomly into the genome, a characteristic that could cause insertional mutagenesis leading to gene deregulation. The integration of four independent factors increases the likelihood of insertional
mutagenesis. In an effort to lower the probability of disturbing gene expression or function, polycistronic transgene vectors have been used, which allow for delivery of all four reprogramming factors with minimal genomic insertion, separating the transgenes with either internal ribosomal entry sites (IRES) or viral 2A peptides (Sommer et al., 2009).

In an attempt to improve upon Yamanaka’s methodology, alternative delivery strategies have been investigated, each with their own benefits and pitfalls. Reprogramming has been achieved by multiple groups with lentiviral vectors (Bleloch et al., 2007; Yu et al., 2007). These vectors are generally more efficient than the retrovirus and can infect both dividing and non-dividing cells, yet lentiviral transgenes are less efficiently silenced in iPSCs than retroviral constructs (Shuyuan Yao et al., 2004). However, both of these delivery systems involve the permanent integration of the reprogramming factors into the host genome. c-Myc and Klf4 are oncogenic, and thus the removal of these transgenes once iPSCs have been formed would be ideal (Lu et al., 2013; Rowland et al., 2005). One method to extract transgenes after reprogramming is flanking the transcription factors with LoxP sites, allowing for Cre recombinase-mediated excision (Kaji et al., 2009; Soldner et al., 2009). However, this leaves a large genetic scar at the site of excision, which could cause gene deregulation. Transposon reprogramming systems, such as the piggyBac system, have the advantage of being able to be excised after reprogramming. By the addition of transposase, the transposons are remobilized and lost at a high frequency, without leaving a genetic footprint (Woltjen et al., 2009). However, transposon excision is not 100% efficient. (Carey et al., 2009).

Despite rapid advances since cellular reprogramming was first described, one plaguing problem still remains with retroviral-, lentiviral-, and transposon-based reprogramming: the integration of DNA elements into the genome. These insertions could prove detrimental to patients who receive these iPSCs as a therapy, as insertional mutagenesis or leaky expression of the transgenes could
lead to tumour formation. Thus, an effort to develop a transgene-free system of reprogramming has been the focus of many laboratories. One strategy is to transfect cells with non-integrating vectors. This method was first proven in MEFs, by transfecting plasmids (Okita et al., 2008), and in mouse hepatocytes, using non-integrating adenoviral vectors (Stadtfeld et al., 2008b). Since then, several other methods have been employed to lend support to the notion that non-integrating expression of the Yamanaka factors is sufficient to induce reprogramming. These techniques include the use of Sendai viral vectors (Fusaki et al., 2009), adenoviral vectors (Zhou and Freed, 2009), and transient episomal vectors (González et al., 2009; Jia et al., 2010; Okita et al., 2008) in human fibroblasts among others.

Further attempts to demonstrate integration-free reprogramming have avoided the use of DNA altogether. The direct insertion of the reprogramming factor proteins into the cell was made possible by attaching them to cell-permeable peptides (Pan et al., 2009). While this methodology did not work initially, the same general principles were met with success by fusing a polyarginine domain to the C-terminus of the reprogramming factors (Kim et al., 2009a; Zhou et al., 2009). The delivery of synthetic mRNAs has also proven effective at reprogramming human fibroblasts to pluripotency (Warren et al., 2010). This method involves the in vitro modification of mRNA transcripts to avoid being targeted by the anti-viral RNA machinery of the cell. Reprogramming via repeated protein transductions into fibroblasts proved to be possible, but very inefficient.

Overall, non-integrating systems tend to be much less efficient than integrating transposon- or viral-based methods, which is most likely due to the deficiency of sustained transgene expression. Despite this lack of efficiency, these methods demonstrate that incorporation into the host genome is not necessary for iPSC generation. Thus, the ultimate goal is to develop a small
molecule cocktail that would induce reprogramming in somatic cells when added to the media. Progress has been made towards the replacement of c-Myc (Mali et al., 2010; Nakagawa et al., 2007), Sox2 (Ichida et al., 2009; Li et al., 2009), Klf4 (Lyssiotis et al., 2009) and recently Oct4 (Shu et al., 2013). Until recently, Oct4 was the only of the Yamanaka factors for which a chemical substitute had not been discovered. However, by activating c-Myc, Klf4, and Sox2 in MEFs, Hou et al. were able to identify FSK as a substitute for Oct4 (Hou et al., 2013). When added to previously identified small molecule substitutes for the other factors, along with epigenetic modulators, they found that they could induce integration-free iPSC generation in fibroblasts with efficiencies comparable to four-factor reprogramming (up to 0.2%) (Hou et al., 2013). While this seems to be a breakthrough in iPSC technology, we will have to wait and see how robust and reproducible this technique is.

1.1.5 Molecular Mechanisms of Reprogramming

The process of reprogramming is the subject of ongoing international investigation in order to elucidate the complete compliment of molecular mechanisms at play. It is clear that during reprogramming, dramatic changes in both the transcriptome and the epigenome take place, transforming the somatic nucleus to that of a pluripotent cell (Maherali et al., 2007; Chin et al., 2009; Hawkins et al., 2010). While the mechanisms behind these transformations have not yet been fully characterized, significant advances in our understanding of reprogramming have been made in the past 7 years.

1.1.5.1 Stoichiometry of the Reprogramming Factors

The levels of ectopic reprogramming factor expression appear to be a crucial determinant of iPSC generation. Increased expression of all four
reprogramming factors has been shown to greatly increase reprogramming efficiency (Polo et al., 2012). Yet it is known that elevated Oct4 levels in ESCs can perturb their pluripotency (Niwa et al., 2000). Consistent with this observation, newly formed iPSCs repress retroviral reprogramming factors (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007), while continuous expression of the reprogramming factors in the latter stages can inhibit the formation of iPSC colonies (Golipour et al., 2012). Subsequent findings indicate that the balance of expression between the reprogramming and the endogenous transcription is crucial for stable and consistent colony formation. For instance, reprogramming is more efficient when cells exhibit low Sox2 levels and high Oct4 levels (Nagamatsu et al., 2012; Tiemann et al., 2011; Yamaguchi et al., 2011). While the overall level of transcription factors is important, the phase in which these factors are expressed must be considered as well.

1.1.5.2 Phases of Reprogramming

Reprogramming has been hypothesized to occur in three phases: initiation, maturation, and stabilization. Typically, the early initiation stages of reprogramming are characterized by rapid proliferation, as well as a mesenchymal to epithelial transition (Li et al., 2010; Samavarchi-Tehrani et al., 2010). This transition is characterized by the formation of tight cell colonies coupled with the prompt up-regulation of E-cadherin, occludin, and other epithelial junction components. During this phase, the BMP pathway activates various microRNAs (miR-205 and miR-200) which are key regulators of MET (Samavarchi-Tehrani et al., 2010). Additionally, several established pluripotency markers have been identified as early markers of reprogramming, such as SSEA1 and alkaline phosphatase (AP) (Brambrink et al., 2008). Other typical ESC markers have been reported to appear later during reprogramming, such as Oct4, Sox2, and Nanog (Stadtfeld et al., 2008a). The appearance of the endogenous transcription factors indicates important milestones, as the
reprogramming cells become independent from the transgenes, allowing for a stable iPSC state. This conversion to a stable iPSC state generally occurs only after at least 10-14 days of continuous reprogramming factor expression.

In the maturation and stabilization phases of reprogramming, it has been reported that silencing of the reprogramming factors is necessary for stable iPSC formation. Interestingly, during these phases, genes involved in the gamete lineage and cytoskeletal dynamics were up-regulated rather than pluripotency regulators (Golipour et al., 2012). Furthermore, genes that are up-regulated in iPSCs compared to fibroblasts tend to be involved in pluripotency maintenance, whereas those genes that only varied between the two cell states were more likely to be involved with regulating the transition between the maturation and stabilization phases (Golipour et al., 2012). This finding suggests that the set of genes that controls this stage-like transition is independent from the genes that control pluripotency maintenance. A separate study recently identified two main waves of gene expression change in cells poised to reprogram (Polo et al., 2012). During the first wave, genes implicated with cytoskeletal organization, proliferation, and metabolism were activated, while those genes associated with development were significantly down-regulated. This wave also saw the gradual rise of early pluripotency markers. The second wave included the up-regulation of genes involved with stem cell maintenance and embryonic development. In this view of reprogramming, the first wave correlates to initiation stage, whereas the second wave correlates to the maturation and stabilization stages (Polo et al., 2012).

1.1.5.3 Reprogramming Milestones

The initiation, maturation, and stabilization stages of iPSC generation describe the general transitions that a somatic cell goes through to become pluripotent. During this progression, a reprogramming cell passes important
functional milestones. The most widely acknowledged of these milestones is the so-called “commitment point”. This is the point at which a reprogramming cell begins to stabilize as a pluripotent cell, independent from exogenous transgene expression. Earlier on during the process of reprogramming, a second milestone exists, which has been referred to as the “point of no return”. This concept relies on the notion that cells exposed to reprogramming factor expression for a short duration would not have their phenotype permanently altered. Removal of exogenous transcription factors prior to the “point of no return” would result in the reversion of the reprogramming cell back to its original somatic state. The existence of this milestone has been demonstrated by inducing reprogramming in secondary MEFs for five days and then stopping the transgene expression, allowing the reversion of the cells back into their fibroblast state (Samavarchi-Tehrani et al., 2010). Separating the “point of no return” and the “commitment point” is a significant window, in which, following the removal of reprogramming factors, cells neither revert back to their original phenotype nor do they progress to a stable pluripotent state. This zone of non-commitment has been termed “Area 51” due to the unknown abilities of the cells in this region in terms of both their prior cell identity and their ability to commit to the iPS lineage (Nagy and Nagy, 2010).

1.1.5.4 Downstream Effects of Reprogramming Factors

Understanding the molecular cascades triggered by the reprogramming factors is of great importance when attempting to bring iPSC technology to a therapeutic reality. Oct4, Sox2, and Klf4 are critical because of the role they play in the reprogramming network of transcription factors (Boyer et al., 2005; Jiang et al., 2008; Loh et al., 2006; Sridharan et al., 2009). These, along with several other pluripotency transcription factors such as Esrrb, Zpf42, Nanog, and Sall4, are known to co-localize to lineage-specific enhancer regions in ESCs (Chen et al., 2008; Kim et al., 2008). Additionally, Oct4, Sox2 and Nanog positively
regulate their own expression as well as the expression of other transcription factors and pluripotent genes involved in the wnt and Tgf-β pathways (Boyer et al., 2005; Chen et al., 2008; Jiang et al., 2008; Kim et al., 2008). c-Myc, conversely, is not part of this network, yet is a key component of cell growth and differentiation. It is not present at the enhancers, as the other pluripotency factors are, and has even been suggested to not be a true reprogramming factor at all (Lin et al., 2012; Nie et al., 2012). The function of c-Myc during reprogramming seems to be to increase overall transcription in the cell, as it generally occupies the promoter regions of many active genes in iPSCs and ESCs (Rahl et al., 2010). This is consistent with the findings that c-Myc is expendable during reprogramming, yet the process happens more rapidly when it is included (Nakagawa et al., 2007). It is possible that the increased transcriptional rate of the cell caused by c-Myc expression is largely brought about by the epigenetic changes it produces. These epigenetic changes are of critical importance to the reprogramming process.

**1.1.5.5 Epigenetics of Reprogramming**

Cell states are defined and maintained by a signature gene expression profile. Transcription generally starts with transcription factors binding the distal promoters and enhancers of a gene, recruiting coactivators and general machinery to the core promoter, and forming the RNA polymerase-II-containing preinitiation complex (PIC) (Green, 2005). The binding of these elements relies heavily on the chromatin structure—namely, the packaging of the DNA into nucleosomes (Li et al., 2007). Epigenetic modifications, such as histone modifications and DNA methylation, play a role in allowing transcription elements to bind DNA. Thus, the expression patterns specific to various cell types are heavily influenced by the epigenetics of the genome. Therefore, the transition of a differentiated somatic cell nucleus to pluripotency involves dynamic epigenetic reconstruction, as previously repressed pluripotency-specific genes need to be
activated, and transcribed lineage-specific genes are silenced. The activation of genes via the removal of the H3K27me3 coupled with the addition of the activating mark H3K4me2 has been discovered in the early time-points of reprogramming (Doege et al., 2013). Furthermore, the use of histone deacetylase (HDAC) inhibitors, such as Valproic acid (VPA) (Huangfu et al., 2008), along with DNA methyltransferase inhibitors, such as 5'-azacytidine (AZA) (Mikkelsen et al., 2008), to enhance reprogramming efficiency demonstrates the important role that epigenetics, and specifically histone modifications, play during reprogramming. These small molecules help to open up the chromatin of the reprogramming cell, allowing for transcription factors to bind to their targets. This open chromatin state is typical of an ESC, allowing it to maintain pluripotency (Efroni et al., 2008).

Interestingly, as many as 70% of reprogramming factor binding events in iPSCs involving Oct4, Sox2, and Klf4 take place in areas of the chromatin devoid of any histone modifications and which do not display DNase hypersensitivity, suggesting that these regions are in a closed chromatin state (Soufi et al., 2012). Additionally, c-Myc cannot bind these genomic regions without the co-expression of the other reprogramming factors (Soufi et al., 2012). This finding lends support to the idea that c-Myc serves not to initiate the reprogramming process, but rather to potentiate the action of Oct4, Sox2, and Klf4.

Epigenetic marks can also act as barriers to reprogramming by delaying the activation of key genomic regions, diverting reprogramming cells from a true reprogramming path. The overall role of DNA methylation during reprogramming is recognized as inhibitory, since repressing Dnmt1 expression, which is responsible for methylation maintenance, improves reprogramming efficiency (Mikkelsen et al., 2008). The promoters of many key pluripotency genes, such as Nanog and Oct4, remain hypermethylated until late in reprogramming (Mikkelsen et al., 2008; Polo et al., 2012), preventing the endogenous expression of these genes and impeding the progression to self-sustained pluripotency. Additionally,
genes that gain active chromatin marks at the early time-points of reprogramming appear to be hypomethylated (Koche et al., 2011). Overall, these findings indicate that the demethylation of critical regions of the genome implicated with pluripotency maintenance is a critical step in the reprogramming of somatic cells.

Since the loss of methylation marks seems to be imperative to the progress of reprogramming, it follows that the retention of methylation of lineage-specific genes may impede the process and lower its overall efficiency. This phenomenon, termed epigenetic memory, has been observed in reprogramming cells by several groups (Kim et al., 2010; Polo et al., 2010). Epigenetic memory may hinder an iPSCs ability to differentiate to a desired cell type, as demonstrated in iPSCs derived from fibroblasts and neural progenitors that were unable to differentiate to hematopoietic cells (Kim et al., 2010). However, when treated with chromatin modifying compounds, the neural progenitor-derived iPSCs were able to form blood cells efficiently, due to the demethylation of the hematopoietic-specific genes (Kim et al., 2010). Further studies show that epigenetic memory is lost after numerous passages, indicating that gains and losses in the epigenome continue to take place after ectopic factor expression has been halted (Polo et al., 2010) This suggests that the starting cell type has an impact on the amount of time it takes to fully reset the methylation status to an ESC-like state (Polo et al., 2010). Furthermore, telomere length slowly increases as iPSCs are passaged (Marion et al., 2009) indicating that reprogramming continues well beyond transcription factor removal. These points need to be considered when selecting iPSCs for practical use, as the complete reprogramming process continues well past the initial colony’s formation.

1.1.6 Partially Reprogrammed Cells

Not all reprogramming cells proceed directly to an ESC-like state. Reprogramming cells can become trapped in a stable intermediate that lies between a somatic cell state and an ESC-like state. Cells in this state are
commonly referred to as partially or incompletely reprogrammed (Fussner et al., 2011; Mikkelsen et al., 2008; Sridharan et al., 2009). Despite displaying some of the markers typical of early reprogramming and being able to form teratomas (Mikkelsen et al., 2008; Okita et al., 2007; Sridharan et al., 2009), partial iPSCs fail to contribute to chimeric embryos. iPSC state achievement is characterized by the retroviral silencing of transgenes accompanied by the endogenous activation of pluripotency genes (Okita et al., 2007; Silva et al., 2008; Stadtfeld et al., 2008a). In partial iPSCs, retroviral transgenes fail to inactivate. Treating partial iPSCs with epigenetic inhibitors, such as AZA effectively converts them to a fully reprogrammed state (Mikkelsen et al., 2008), indicating that barrier between partial iPSCs and ESC-like iPSCs is an epigenetic one. This transition is marked by the reactivation of the X chromosome and the establishment of a genome-wide ESC-like histone modification pattern (Maherali et al., 2007). Interestingly, the generation of partial iPSCs is not observed in three-factor (OSK) reprogramming, implying that c-Myc is responsible for partial iPSC derivation. Without the pro-proliferation effects of the c-Myc transgene, OSK reprogramming is much less efficient. c-Myc serves to foster cell-cycle progression in reprogramming cells that is naturally promoted within the ESC network of genes. Thus, partially reprogrammed cells may be artifacts of artificially induced cell cycle promotion. However, partial iPSCs have not generated much interest in the research community, as they are seen as failed reprogramming events and thus largely discarded. A further examination of these cells may provide hints to some of the unanswered questions about reprogramming.

1.1.7 Inducible Systems for Studying Reprogramming

We must attain a complete understanding of the molecular mechanisms at play during reprogramming before iPSCs can become a therapeutic reality. While the examination of completely and even incompletely reprogrammed iPSCs is
possible with a primary reprogramming system, characterizing the early phases of reprogramming is difficult due to the low efficiency of the reprogramming process (~0.01%). Moreover, the lack of predictive cell surface markers makes it practically impossible to tease the nascent reprogramming cells apart from the non-reprogramming background. Because of this issue, most reports characterizing the process of reprogramming tend to focus mainly on the latter phases of iPSC generation, where cell types are physically discernable from each other, while making broad claims about the in-between time-points when many of the molecular changes are happening.

One effective way to study the complete reprogramming process is to use a secondary reprogramming system. In a secondary system, cells are derived from an iPS colony that has been expanded and differentiated in vitro or in vivo through the generation of chimeras (Wernig et al., 2008a; Woltjen et al., 2009). In order to make differentiation possible, the reprogramming factors are controlled by an inducible promoter such as tetracycline-inducible system. The tetO promoter in this system is activated by exposure to both doxycycline (dox), a more stable derivative of tetracycline antibiotic family, as well as the reverse tetracycline transcriptional activator (rtTA) protein within the cell (Figure 1a). rtTA can be controlled either by a ubiquitous promoter such as ROSA or a tissue specific promoter. If both elements are present, dox will facilitate the binding of rtTA to the tetO promoter, and transcription of the downstream genes will occur. In the absence of dox, rtTA will fail to bind to the tetO promoter and the system will be silent. This allows for both temporal and spatial control over the activation of the reprogramming factors in vitro and in vivo. This method ensures that all cells express high levels of the reprogramming factors in the same proportions as the parental iPSC. Because of this, secondary reprogramming allows for the population-based generation of iPSCs at an elevated efficiency that is unmatched by any method of primary reprogramming. (Samavarchi-Tehrani et al., 2010; Woltjen et al., 2009). This powerful tool can be used answer the big questions
surrounding reprogramming, such as the molecular cascades that take place during the initiation phase and the regulatory machinery that directs the cell to a pluripotent state. One aspect of particular interest is the role that microRNAs play during reprogramming.
Figure 1. The Secondary Reprogramming System. (a) Doxycycline (dox) acts as a switch to turn on transcription at the inducible tetO promoter. When dox is not present, rtTA is unable to bind to the tetO promoter, and the downstream transgene is not transcribed. However, when dox is present (green squares), rtTA is able to bind the promoter, and transcription takes place. (b) The four Yamanaka factors, each downstream of a tetO promoter, are transfected into primary MEFs and inefficient reprogramming takes place when exposed to dox. Primary iPSCs are used to derive a chimeric embryo, which gives rise to secondary MEFs. When dox is added, these secondary MEFs reprogram at an elevated efficiency.
1.2 MicroRNA

MicroRNAs (miRNAs; miRs) are short, non-coding RNA molecules that act as pre-translational gene expression modulators in the cell. There are over 1,000 characterized miRNAs in both the human and the mouse genome, generally ranging in size from 21-24 nucleotides (nt) in their mature form. miRNA genes are located throughout the genome, in intergenic spaces, introns, and exons. miRNAs were initially discovered by Victor Ambros’ group in 1993 (Lee et al., 1993). They reported that in *C. elegans*, Lin-14 is regulated by a short RNA product which complimentarily binds to the 3’ UTR of the Lin-14 mRNA. This complimentary binding proved to be both sufficient and necessary for repression of mRNA translation.

While first found in *C. elegans*, highly conserved miRNAs have since been reported in the majority of eukaryotic organisms, suggesting widespread evolutionary importance (Lee et al., 2007; Tanzer and Stadler, 2004). miRNAs are thought to be involved in the majority of biological processes in humans (Lim et al., 2005), including insulin secretion (Poy et al., 2004), hematopoetic differentiation (Chen, 2004), and cell metabolism (Wilfred et al., 2007). Their expression is highly discrete, such that the identification of specific cell types can be refined by the signature of miRNA expression (Aboobaker et al., 2005; Laurent et al., 2008; Monticelli et al., 2005), and can even be used to discern between closely related cancer types (Lu et al., 2005).

1.2.1 The Path to MicroRNA Maturity

The processing of miRNAs to their functional form involves several steps. Most miRNAs are transcribed either from their own gene promoters or from the
introns of larger genes, and can be found as either independent miRNAs or in larger miRNA clusters, composed of two or more miRNAs in close physical proximity. The miRNA gene is largely self-complementary, and thus once it is transcribed by RNA Polymerase II, the resulting single-stranded RNA folds back onto itself, forming a hairpin structure (Figure 2). This primary miRNA (pri-miRNA) is capped and polyadenylated (Cai et al., 2004). The hairpin secondary structures of the pri-miRNA are then recognized by the microprocessor-complex. This complex includes Dcgr8, a compound critical for microprocessor-complex function, and Drosha, an RNase type III endonuclease, which cleaves the RNA molecule, resulting in pre-miRNA (Denli et al., 2004; Gregory et al., 2004; Han, 2004; Landthaler et al., 2004). This ~70nt pre-miRNA is recognized by the Exportin-5/Ran-GTP complex, which actively transports it from the nucleus to the cytoplasm of the cell (Bohsack et al., 2004; Lund, 2004; Yi, 2003). Here, Dicer, a second RNase type III endonuclease, recognizes the overhang left by the microprocessor-complex and cleaves the pre-miRNA, creating a ~22nt miRNA-duplex, known as mature miRNA. The cut made by Dicer within the loop of the hairpin structure of the pre-miRNA also leaves an overhang on the mature miRNA (Hutvagner et al., 2001; Zhang et al., 2012). The two strands of the mature miRNA are named the 3p and the 5p strands. If it is known which of the two strands is more highly expressed, then the less abundant of the two strands is labeled with a star (⋆) and referred to as the star strand. In some cases, however, both miRNA strands are capable of mRNA regulation (Okamura et al., 2008). The mature miRNA remains bound to Dicer. At this point, one of the two strands of the miRNA-duplex is excluded from the RNA-Dicer complex and degraded by the cell’s machinery. The remaining miRNA strand, known now as the template strand, along with Dicer and several other enzymes make up the RNA-induced silencing complex (RISC). The RISC is then able to use the single-stranded miRNA as a template to target mRNAs for either repression or degradation (Jing et al., 2005). When the miRNA and the mRNA are
Figure 2. A Schematic Diagram of the Processing and Gene Regulation Mechanisms of MicroRNA. Transcription of the miRNA gene produces pri-miRNA. The hairpin structure of pri-miRNA is recognized by Drosha (part of the microprocessor-complex) and cleaved to form pre-miRNA. The Exportin-5/Ran-GTP complex transports the pre-miRNA from the nucleus to the cytoplasm, where it is further cleaved by Dicer, forming mature miRNA. One of the strands of the mature miRNA duplex is incorporated into the RNA-induced silencing complex (RISC), which mediates gene regulation by either repressing translation (if miRNA and mRNA are not completely complementary) or mRNA cleavage (if the two are completely complementary).
complementarily bound, Ago2, a component of the RISC, is able to cleave the mRNA, signaling it for degradation. If, however, the miRNA and the mRNA are not completely complementary, the RISC silences the gene by preventing translation (Lim et al., 2005).

1.2.2 MicroRNA Targets

Identifying the target mRNA of each miRNA is not a straightforward task, as each miRNA may have hundreds of different mRNA targets. Very few accepted rules exist to predict miRNA targets, yet one principle is widely agreed upon in the field: the importance of the seed sequence. The seed sequence is located on the 5' end of miRNAs, and usually spans from nucleotides 2 to 8 (Jinek and Doudna, 2009). This region is the most evolutionarily conserved segment of the miRNA because of its importance in mRNA targeting (Lewis et al., 2003; Lim et al., 2003). The miRNA seed generally binds to the 3' untranslated region (UTR) of the mRNA. The majority of reported miRNA-induced mRNA silencing involves the complementary binding of the seed sequence to the mRNA, and in many cases, a seed-mRNA match suffices to repress translation (Doench and Sharp, 2004; Lewis et al., 2005; Krek et al., 2005). Many miRNA families have identical or nearly identical seed sequences, suggesting that these miRNAs target the same set of genes.

1.2.3 MicroRNAs and Pluripotency

miRNAs play a large role in pluripotency maintenance. When one of the essential miRNA-processing enzymes, such as Dicer, is knocked out in ESCs, the cells are unable to differentiate properly (Kanellopoulou et al., 2005).
Furthermore, the homozygous disruption of Dicer or Dgcr8 leads to embryonic lethality in mice (Bernstein et al., 2003; Wang et al., 2007), indicating a global importance of miRNAs during development. miRNAs also assist in somatic cell fate conversion, or transdifferentiation. Overexpression of specific miRNAs facilitates the transition of human fibroblasts to both neurons (Ambasudhan et al., 2011; Yoo et al., 2011) and cardiomyocytes (Jayawardena et al., 2012).

Several miRNA families are upregulated in both human and mouse iPSCs and ESCs over differentiated cell types, including the miR-302, miR-290, miR-106, miR-17, miR-520, miR-195, and miR-200 families (Bar et al., 2008; Laurent et al., 2008; Morin et al., 2008; Suh et al., 2004). Conversely, the let-7 family of miRNAs, show the opposite expression pattern and are down-regulated in ESCs (Bar et al., 2008; Laurent et al., 2008). During differentiation, key pluripotency genes, such as Nanog, Oct4, and Sox2, are downregulated, resulting in the inhibition of self-renewal. This downregulation also represses expression of Lin28 as well as the pluripotency-specific miR-290 cluster (Marson et al., 2008). Furthermore, Lin28 has been shown to inhibit the let-7 miRNA family from maturing (Rybak et al., 2008; Viswanathan et al., 2008). As levels of Lin28 drop, the maturation of let-7 is no longer inhibited, and self-renewal is further inhibited.

1.2.4 MicroRNAs that Affect Reprogramming Efficiency

The importance of miRNAs during reprogramming has been established, as fibroblasts that lack all mature miRNAs are unable to form iPSCs (Kim et al., 2012). One of the most important miRNA families to both pluripotency maintenance and reprogramming is the miR-290 cluster is incredibly abundant in ESCs, constituting ~60% of all small RNA species in the cell (Marson et al., 2008). When transiently transfected over a period of six days in three-factor (Oct4, Sox2, and Klf4) reprogramming cells, three members of the 290-cluster,
miR-291-3p, -294, and -295, separately enhanced the generation of iPSCs (Judson et al., 2009). These miRNAs are also part of the ESCC (Embryonic Stem Cell-specific) miRNA group, defined by its ability to maintain the highly proliferative state of ESCs by accelerating the transition between G1 and S-phase through the targeting of p21 and other cell cycle regulators (Wang et al., 2008). Further investigation of this cluster indicated that the miR-290 cluster acts downstream of c-Myc, yet the overexpression of the miR-290 cluster did not cause the same rapid proliferatory effects as c-Myc does.

The miR-290 cluster is not the only one that enhances reprogramming. Many publications have reported that the overexpression of various miRNA families increases the efficiency of reprogramming along with either the four (c-Myc, Klf4, Oct4, and Sox2) or three (Klf4, Oct4, and Sox2) reprogramming factors. These families include the miR-17/92 cluster (Li et al., 2011b), the miR-106a/363 cluster (Li et al., 2011b; Liao et al., 2011), miR-106b/25 cluster (Li et al., 2011b), and the miR-302/367 cluster (Liao et al., 2011; Judson et al., 2009; Subramanyam et al., 2011). Likewise, the repression of certain miRNAs that inhibit reprogramming leads to a heightened reprogramming efficiency, such as miR-21 or miR-29a in the three-factor reprogramming system (Yang et al., 2011), let-7 in the four-factor system (Melton et al., 2010; Yang et al., 2011), or miR-34 in both systems (Choi et al., 2011). Three of the verified targets of the miR-34 family during reprogramming include Sox2, Nanog, and N-Myc (Choi et al., 2011). It has also been demonstrated that the miR-34 family is a direct target of p53 (He et al., 2007), a tumour suppressor gene that is a known repressor of reprogramming (Hong et al., 2009; Kawamura et al., 2009; Utikal et al., 2009). Likewise, overexpression of miR-138, a p53-regulating miRNA, increases reprogramming efficiency in three-factor reprogramming (Ye et al., 2012). Furthermore, it has been reported that iPSCs can be divided into two separate categories, distinguished based on the status of the p53 network within these cells (Neveu et al., 2010). Cells in one group could transition to the other by
overexpression p53-targeting miR-92 and miR-141. Hence, the overexpression and/or repression of these miRNAs have a large impact on the pluripotent status of iPSCs.

Of all the miRNAs that have been reported to promote or regulate iPSC formation, the miR-302/367 cluster (302-cluster) has the largest impact on the reprogramming process. This cluster, consisting of 5 miRNAs located entirely within a 1kb region on mouse chromosome 3, is present at very high levels in both ESCs and iPSCs (Card et al., 2008). The 302-cluster has been shown to increase reprogramming efficiency when overexpressed alongside the Yamanaka factors (Judson et al., 2009). In addition, it is the only miRNA cluster that has ever been reported to reprogram somatic cells to pluripotency completely independent of any other factors (Anokye-Danso et al., 2011; Lin et al., 2008; Miyoshi et al., 2011). When cultured with VPA, miR-302/367 cluster-transfected fibroblasts reportedly generated iPSCs at a rate 100-times greater than Yamanaka factor reprogramming (Anokye-Danso et al., 2011). While these cells were able to produce teratomas and chimeras, this method has not been repeated by other groups, making it a somewhat controversial finding (Lüningschrör et al., 2013).

Despite the lack of corroborating evidence that the miR-302/367 cluster is able to independently reprogram fibroblasts to pluripotency, there is no lack of support for the importance of this cluster in generating iPSCs. Using chromatin immunoprecipitation (ChIP)-sequencing assays, Nanog, Oct4 and Sox2 all bind to the promoter regions of the miR-302/367, miR-290, and miR-106a clusters in ESCs (Marson et al., 2008). Additionally, the miR-302/367 cluster is activated by Oct4 and Sox2 (Card et al., 2008). One of the few known roles that the 302-cluster has in ESCs is, along with Oct4, the down-regulation of NR2F2, a critical regulator of neuronal differentiation and antagonist to Oct4 (Rosa and Brivanlou, 2010). These findings suggest that these pluripotent-specific miRNAs work
together with the reprogramming factors to reprogram the somatic nucleus and promote pluripotency. The more understanding of the roles that miRNA plays during reprogramming grows, the more we appreciate just how important miRNAs are to iPSC generation. Yet thus far, the majority of examined miRNAs are those that are up-regulated in pluripotent cells when compared to their parental somatic cells. The examination of miRNAs that are differentially expressed between pluripotent cells and those that fail to properly reprogram may reveal other miRNAs integral to the reprogramming process.
Table 1. MicroRNAs that Impact Reprogramming. Summary of miRNAs that, when overexpressed in reprogramming cells, either enhance or repress the efficiency of iPSC generation.

<table>
<thead>
<tr>
<th>miRNA or miRNA family</th>
<th>Effect on Reprogramming</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-302/367 family</td>
<td>Increases reprogramming efficiency; Induces reprogramming</td>
<td>Liao et al., 2011; Judson et al., 2009; Subramanyam et al., 2011; Anokye-Danso et al., 2011</td>
</tr>
<tr>
<td>miR-290 family</td>
<td>Increases reprogramming efficiency</td>
<td>Judson et al., 2009</td>
</tr>
<tr>
<td>miR-17/92 cluster</td>
<td>Increases reprogramming efficiency</td>
<td>Li et al., 2011</td>
</tr>
<tr>
<td>miR-106a/363 cluster</td>
<td>Increases reprogramming efficiency</td>
<td>Li et al., 2011; Liao et al., 2011</td>
</tr>
<tr>
<td>miR-106b/25</td>
<td>Increases reprogramming efficiency</td>
<td>Li et al., 2011</td>
</tr>
<tr>
<td>miR-138</td>
<td>Increases reprogramming efficiency</td>
<td>Ye et al., 2012</td>
</tr>
<tr>
<td>miR-25</td>
<td>Increases reprogramming efficiency</td>
<td>Lu et al., 2012</td>
</tr>
<tr>
<td>miR-200 family</td>
<td>Increases reprogramming efficiency</td>
<td>Wang et al., 2013</td>
</tr>
<tr>
<td>miR-199a</td>
<td>Increases reprogramming efficiency</td>
<td>Wang et al., 2012</td>
</tr>
<tr>
<td>miR-21</td>
<td>Decreases reprogramming efficiency</td>
<td>Yang et al., 2011</td>
</tr>
<tr>
<td>miR-29a</td>
<td>Decreases reprogramming efficiency</td>
<td>Yang et al., 2011</td>
</tr>
<tr>
<td>miR-34</td>
<td>Decreases reprogramming efficiency</td>
<td>He et al., 2007</td>
</tr>
<tr>
<td>let-7</td>
<td>Decreases reprogramming efficiency</td>
<td>Yang et al., 2011; Melton et al., 2010</td>
</tr>
</tbody>
</table>
Chapter 2: Aims & Hypothesis

With the advent of reprogramming, there has been a clear movement to progress iPSC technology to the clinic as quickly as possible. However, there are still many questions that remain about the reprogramming process, specifically regarding the molecular mechanisms at play. The complete elucidation of the genetic, epigenetic, and proteomic components of reprogramming may open the doors for a method of deriving iPSCs that is more efficient, less disruptive to the cell, and safer for the patient than those methods that are currently being used. Thus far, the majority of studies addressing these molecular mechanisms lack the integration of the various molecular platforms (for example, the interplay between epigenetic marks, such as H3K4me3 or CpG methylation status, and miRNA expression). Therefore, a unique dataset was generated in our laboratory that allows for the in-depth molecular analysis of secondary reprogramming cells.
as they progress to different reprogramming states. Deep sequencing of small RNA and epigenetic histone modifications were among the many platforms analyzed. This integrative dataset will not only allow the investigation of the molecular dynamics of reprogramming, but also the analysis of the interaction between “OMICs” platforms.

Recently, a field of study has emerged characterizing the role that miRNAs play during reprogramming. miRNAs serve as excellent candidates for the improvement of reprogramming, as they have been implicated in pluripotency maintenance (Kanellopoulou et al., 2005; Kim et al., 2012), they regulate the expression of multiple networks of genes (Wang et al., 2008), and they have already been shown to have powerful effects on the efficiency of iPSC generation (Anokye-Danso et al., 2011; Judson et al., 2009). Yet, this relatively new area of iPSC biology has only characterized the function of a small fraction of known miRNAs during reprogramming. It is highly likely that there are more miRNAs that positively regulate reprogramming than have already been described.

Taking advantage of the aforementioned dataset, miRNAs important to the efficient reprogramming of somatic cells will be identified by comparing miRNA expression in divergent iPSC states. Thus, it is hypothesized that an overexpression of those individual miRNAs specific to an ESC-like iPSC state will increase the efficiency of reprogramming in fibroblasts.

We have set four specific aims of this project: (1) identification of candidate iPSC-promoting miRNAs; (2) miRNA qPCR optimization and validation of small RNA deep-sequencing data; (3) validation of these candidate miRNAs in a primary reprogramming system; and (4) elucidation of miRNA effects on reprogramming via forced overexpression of miRNAs in primary reprogramming fibroblasts.
(1) **miRNA qPCR optimization.** A separate quantification method is necessary to verify the miRNA expression levels measured by deep sequencing. Since miRNA is much smaller than mRNA, typical qPCR methods—specifically the reverse transcription (RT) step—will need to be modified and optimized to measure miRNA levels accurately. Optimization of qPCR and RT primer design, RT primer pooling, and internal normalization will all be performed to obtain a repeatable and efficacious miRNA quantification protocol.

(2) **Identification of candidate iPSC-promoting miRNAs and validation of small RNA deep sequencing data.** Using the available dataset, miRNA expression will be compared between two distinct iPSC states (described in detail in the results section). By setting filters on miRNA expression data, miRNAs will be identified that are significantly up-regulated in ESC-like iPSC states compared to iPSC states more reflective of incomplete reprogramming.

(3) **Validation of candidate miRNAs in a primary reprogramming system.** Since the integrative reprogramming dataset was generated using a secondary reprogramming system, it could be argued that any significant fluctuation in miRNA expression is simply an artifact of secondary reprogramming. To address these issues, miRNA expression levels will be measured in a number of primary iPSC lines.

(4) **Elucidation of miRNA effects on reprogramming via forced overexpression of miRNAs in primary reprogramming fibroblasts.** Each novel candidate miRNA, along with several positive control
miRNAs (previously identified iPSC-promoting miRNA), will be cloned into an inducible transposon plasmid. These miRNAs will then be co-transfected along with the reprogramming factors into naïve fibroblasts and allowed to reprogram. The effect that these miRNAs have on reprogramming will be measured by monitoring the number of colonies formed at periodic time-points after the induction of reprogramming.
Chapter 3:
Materials & Methods

3.1 Work Done Prior to Thesis

3.1.1 Primary and Secondary Mouse Embryonic Fibroblast Derivation

15.5 dpc ROSA26-rtTA-IRES-GFP mouse embryos were decapitated, eviscerated, dissociated with 0.25% trypsin, 0.1% EDTA, and plated in DMEM, 10% FBS, penicillin-streptomycin and Glutamax, as described previously (Woltjen et al., 2009). Secondary (2°) mouse embryonic fibroblasts (MEFs) were obtained by aggregating these primary 1B iPSCs (Woltjen et al., 2009) via tetraploid complementation and harvesting fibroblast cells from E13.5 embryos. This was performed by Dr. Knut Woltjen.
3.1.2 Secondary Reprogramming Cell Culture

After expansion for 3 passages in DMEM, 10% FBS, penicillin-streptomycin, Glutamax, β-mercaptoethanol, sodium-pyruvate, non-essential amino acids, and LIF (hereafter referred to as “ESC media”), reprogramming was induced by exposure to 1.5µg/mL of doxycycline (dox). Reprogramming cells were cultured as two parallel lineages passaged on alternating days (three day intervals) to ensure that collection of cell samples only occurred 48h after passaging. High-dox samples (D2_H, D5_H, D8_H D11_H, D16_H, and D18_H) were collected from cultures maintained at 1.5µg/mL dox throughout reprogramming. Low-dox samples (D16_L and D21_L) were collected from a separate culture maintained at 1.5µg/mL dox for 8 days, then dropped to 5ng/mL dox for the remainder of the time-course. Dox was also removed from this culture at day 14 and maintained in dox-free media, and cells were sampled at day 21 (D21_L0). 2° iPSCs were generated from the high-dox culture by removing dox at day 19 and culturing cells in dox-free media until day 30. Rosa26rtTA ESCs were collected as a control. This was performed by Dr. Peter Tonge, Dr. Mira Puri, and Dr. Claudio Monetti.

3.1.3 Multi-Platform OMICs Analysis

All OMICs data (proteomics, transcriptomics, methyolomics, etc.) was derived from the same biological samples, allowing for direct comparisons between all platforms of analysis. miRNA platform analysis will be discussed in further detail here. For complete OMICs analysis methods, see Tonge et al., 2013 (manuscript in preparation).
3.1.4 MicroRNA Sequencing

miRNA was extracted and purified using the miRvana miRNA isolation kit (Ambion, #1560). Prior to library preparation, miRNA samples were quality validated using a Bioanalyser (Bio-Rad). Small RNA libraries were prepared for SOLiD™ next generation sequencing, with libraries sequenced to a depth of 27,420,558-118,946,232 tags (average 55,816,766 tags; up to 35 nucleotides in length), yielding a total of 725,617,952 tags. These tags were then mapped to the mouse genome (mm37 assembly) and miRNA-mapped tags determined as those overlapping with known miRNA loci (miRbase v18). Thus, 347,190,702 tags were mapped across the thirteen libraries (47% of tags). The processing of this raw data is depicted in Figure 3. This was performed by Dr. Jennifer Clancy, Dr. Hardip Patel, and Dr. Brian Parker.
Figure 3. Small RNA Deep Sequencing Data Processing Pipeline. Sequential steps of small RNA deep sequencing data processing, beginning with obtaining the raw reads, and ending with a fully normalized dataset. Figure adapted from Tonge et al., 2013 (manuscript under review).
3.2 Work Done During Thesis

3.2.1 Transforming Chemically Competent Bacterial Cells

A 5µL aliquot of the desired plasmid was mixed with 50µL of thawed chemically competent DH5α (Invitrogen) Escherichia coli (E. coli) bacteria. E. coli/plasmid mixture was incubated on ice for 15min, heat shocked at 42°C for 40s, and again on ice for 2min. 500µL of liquid broth (LB) was added, and the mixture was incubated at 37°C for 1h shaking at 450rpm. Cells were spun down at 5000rpm using a tabletop micro-centrifuge for 3min. The supernatant was removed and the pellet was resuspended in 50µL of LB, spread-plated on ampicillin agar plates, and incubated at 37°C overnight.

3.2.2 Plasmid Vector Construction

Two unique restriction plasmids (PacI and Ascl) were added to SB-tetO2-mCherry-polyA (provided by P. Tonge) downstream of the mCherry coding region, prior to the polyA sequence by designing and annealing two 48-mer oligos together. miRNAs were cloned by designing primers flanking the genomic locus of the miRNA, leaving at least 50 nucleotides between the primer and the miRNA gene (Table 2). Primers were designed to have a melting temperature between 50 and 55°C and were annealed with the appropriate restriction site (either PacI or Ascl). PCR was performed with genomic DNA (gDNA) obtained from primary MEFs using DNeasy Blood & Tissue Kit (Qiagen, #69506). Standard PCR conditions were: 98°C for 10s, 55°C for 30s, 72°C for 30s; ×32 cycles. PCR product was run on a 2.0% agarose gel + SYBR Safe DNA Gel Stain (Invitrogen, #S33102). DNA from the electrophoresis band was excised and purified using QIAquick Gel Extraction Kit (Qiagen, #28706) and ligated into
pGemTeasy (Promega, #A1360) using Mighty Mix DNA ligation kit (Takara, #6023). Ligation product was amplified using chemically competent DH5α bacteria, as described above, and transfected cells were grown overnight on X-gal ampicillin agar plates. Non-blue bacterial colonies were picked for expansion in LB and incubated overnight at 37°C under 225rpm agitation. Plasmids were obtained from bacteria using QIAprep Spin Miniprep Kit (Qiagen, #27106). Plasmids were digested with Pael and Ascl at 37°C overnight, run on a 2.0% agarose gel + SYBR Safe, and the band was excised. SBTC-(Ascl-Pael) was digested and isolated in the same manner. SBTC-backbone and miRNA fragment were ligated, transformed into HD5α bacteria, and expanded. Plasmids were purified and sequenced at the Centre for Applied Genomics (Toronto).

3.2.3 Plasmid Transfection and Cell Culture

MEFs were plated in ESC media on 10cm dishes at a density of 1-2×10^5 cells per 10cm². After 24h in culture (5% CO₂, 37°C), cells were resuspended with 0.25% trypsin, 0.1% EDTA, in 50mL of media, counted, and transfected using the Neon Tranfection System (Life Technologies) protocol. For each transfection, 1×10^6 cells were transfected with up to 10µg of DNA by applying an electric pulse of 1400V for 20ms, ×2. MEFs were plated on 0.1% gelatin-coated 10cm or 6-well plates in dox-containing ESC media (1.5µg/mL dox) immediately after transfection. After 24h, and every 48h thereafter, the media was changed. When indicated, the media was supplemented with puromycin (1µg/mL) 24h post-transfection and removed 72h after transfection following multiple PBS washes.
3.2.4 Alkaline Phosphatase Stain

After transfection, cells were plated equally in four wells of separate 6-well plates and were incubated with regular media change with dox-containing ESC media. Prior to alkaline phosphatase (AP) staining, 100mmol Tris-HCl pH 8.3 was made, and AP stain 1mL of solution per well of a 6-well plate was prepared as the VECTOR Red Alkaline Phosphatase Substrate Kit (Vector Laboratories, #SK-5100) protocol dictates. Cells to be stained with AP were washed twice with PBS and 1mL of the AP stain solution was added. Cells were incubated overnight at room temperature in the dark. AP solution was then removed and the cells were washed and stored in PBS. Colonies were counted using Volocity Software (PerkinElmer).

3.2.5 Three-Factor iPSCs

ROSA26-rTairetro-IRES-GFP MEFs were transfected with PB-TET-Oct4, PB-TET-Sox2, and PB-TET-Klf4 (Woltjen et al., 2009), in a 1:1:1 molar ratio (~1\(\mu\)g each), along with 250ng of pCyL43 PB transposase via Neon transfection protocol, as described above. Cells were cultured in dox-containing ESC media without passage for 21 days. Colonies were picked when a size of ~200-1000 cells was reached. Picked colonies were expanded with mitotically-inactive feeders in 96-well plates and later passaged to 6-well plates. Cell samples were collected at least 4 days after passage, pelleted, and stored at -80°C for miRNA/mRNA extraction.
3.2.6 Reverse Transcription of MicroRNA

Reverse Transcription (RT) stem-loop primers were designed with a common 44nt stem-loop sequence and a unique 6nt 3’ terminal region, specific to the miRNA being amplified (Table 2). The stem-loop region of the primer was designed with two 14nt homologous (stem) regions separated by a 16nt non-homologous (loop) region. When the secondary structure of the stem-loop primer is formed, the 6nt unique unpaired region binds to the 6nt at the 3’ end of the mature miRNA, forming a DNA-RNA duplex, necessary for reverse transcription. RT was then carried out according to TaqMan miRNA Reverse Transcription Kit (Life Technologies, #4366596).

3.2.7 MicroRNA qPCR

Common reverse primer was designed to anneal to stem-loop construct, while unique forward primers were designed to anneal to the 5’ region of the miRNA that was not bound by the RT stem-loop primer. Multiple cytosine and guanine nucleotides were added to the 5’ end of the forward primers to raise the primers’ melting temperature to 54-57°C. 2μL of 10x diluted RT product was added to 10μL of LuminoCt SYBR Green qPCR ReadyMix (Sigma) along with 10pmol of each qPCR primer and ddH₂O up to 20μL. miRNAs were quantified using the CFX96 Real-Time Detection System (Bio-Rad), qPCR conditions were: 95°C for 10s, 58°C for 8s; ×40 cycles.

3.2.8 Statistical Analysis

Unless otherwise stated, all data presented are representative of a minimum of 3 independent experimental replicates. MiRNA expression and data screening was performed with R Statistical Software. Distance matrixes and
heirarchical clustering were performed with Multiexperiment Viewer. Statistical analysis was performed by Excel (Microsoft), Prism (Graphpad), or Multiexperiment Viewer.

### Table 2. List of Primers Used for RT, qPCR, and MicroRNA Cloning.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-10b-5p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACCACAAA</td>
</tr>
<tr>
<td>miR-129-1-3p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACATCTTT</td>
</tr>
<tr>
<td>miR-195-5p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACGCACAAAT</td>
</tr>
<tr>
<td>miR-203-3p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACCTAGTG</td>
</tr>
<tr>
<td>miR-205-5p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACCCAGACT</td>
</tr>
<tr>
<td>miR-21-5p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACTCAACA</td>
</tr>
<tr>
<td>miR-214-3p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACACTGCC</td>
</tr>
<tr>
<td>miR-301b-3p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACGCTTTTG</td>
</tr>
<tr>
<td>miR-302d-3p_RT</td>
<td>GTCGTATCCAGTGCAAGGTCCGGAGGTATTCGCACT GGATACGACACACTC</td>
</tr>
<tr>
<td>miRNA</td>
<td>RT Primer Sequence</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>miR-378a-3p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACCTTCT</td>
</tr>
<tr>
<td>miR-483-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACCTCCT</td>
</tr>
<tr>
<td>miR-485-3p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACGAGAGG</td>
</tr>
<tr>
<td>miR-141-3p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACCCATCT</td>
</tr>
<tr>
<td>miR-142-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACAGTAGT</td>
</tr>
<tr>
<td>miR-149-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACGGGAGT</td>
</tr>
<tr>
<td>miR-16-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACCGCCA</td>
</tr>
<tr>
<td>miR-191-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACCGAGCTG</td>
</tr>
<tr>
<td>miR-30e-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACCTTCCA</td>
</tr>
<tr>
<td>miR-484_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACATCGGG</td>
</tr>
<tr>
<td>miR-195-5p_RT</td>
<td>GTCGATCCAGTGCAAGGTCCGAGGTATTCGCACT</td>
</tr>
<tr>
<td></td>
<td>GGATACGACGCAAT</td>
</tr>
<tr>
<td>miR-10b-5p_Fwd</td>
<td>GGCTACCCTGTAGAAACGAA</td>
</tr>
<tr>
<td>miR-129-1-3p_Fwd</td>
<td>GGAAGCCCTTACCCCAA</td>
</tr>
<tr>
<td>miR-195-5p_Fwd</td>
<td>GGGCTAGCAGCAGAAAT</td>
</tr>
<tr>
<td>miRNA</td>
<td>Forward Sequence</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>miR-203-3p_Fwd</td>
<td>CCGCGGTGAAATGTTTAGGAC</td>
</tr>
<tr>
<td>miR-205-5p_Fwd</td>
<td>CGTCCTTCATTCCACCGG</td>
</tr>
<tr>
<td>miR-21-5p_Fwd</td>
<td>GCGCCTAGCTTTATCAGACTGA</td>
</tr>
<tr>
<td>miR-214-3p_Fwd</td>
<td>CACAGCAGGCACAGACA</td>
</tr>
<tr>
<td>miR-301b-3p_Fwd</td>
<td>CGCCAGTGCAATGGTCTTGT</td>
</tr>
<tr>
<td>miR-302d-3p_Fwd</td>
<td>CGGCTAAGTGCTTCCATGT</td>
</tr>
<tr>
<td>miR-378a-3p_Fwd</td>
<td>CGGGACTGGACTTGGAGTC</td>
</tr>
<tr>
<td>miR-483-5p_Fwd</td>
<td>CGGAAGACGGGAGAAGAGA</td>
</tr>
<tr>
<td>miR-485-3p_Fwd</td>
<td>CGAGTCATACACGGCTCT</td>
</tr>
<tr>
<td>miR-141-3p_Fwd</td>
<td>GCGCTAACAAGTCTCTGGTAA</td>
</tr>
<tr>
<td>miR-142-5p_Fwd</td>
<td>GCGGCCATAAAGTAGAAAGC</td>
</tr>
<tr>
<td>miR-149-5p_Fwd</td>
<td>GTCTGGCTCCGTGTCTTC</td>
</tr>
<tr>
<td>miR-16-5p_Fwd</td>
<td>GGCCTAGCAGCAGCTAAATA</td>
</tr>
<tr>
<td>miR-191-5p_Fwd</td>
<td>CGCAACGGAATCCCAAAAG</td>
</tr>
<tr>
<td>miR-30e-5p_Fwd</td>
<td>GGCCCTGTAACATCCTTGAC</td>
</tr>
<tr>
<td>miR-484_Fwd</td>
<td>GTCAGGCTCAGTCCCCCT</td>
</tr>
<tr>
<td>miR-195-5p_Fwd</td>
<td>GGGCTAGCAGCAGCAGAAAT</td>
</tr>
<tr>
<td>StemLoop Cmn Rev</td>
<td>GTGCAGGGTGCCAGG</td>
</tr>
<tr>
<td>miR-483_insert_F.2</td>
<td>TTAATTAACAAACTGCTCATGGGTAC</td>
</tr>
<tr>
<td>miR-483_insert_R.2</td>
<td>GGCCTAGCGTGGGAGGAGGA</td>
</tr>
<tr>
<td>290_clstr_insert_Fwd</td>
<td>TTAATTAACCAGGTTTCCTCCAGGTTGG</td>
</tr>
<tr>
<td>290_clstr_insert_Rev</td>
<td>GGCCTAGCGGTCCCAGGTTCCAGGAACCTTC</td>
</tr>
<tr>
<td>Insertion ID</td>
<td>Sequence</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>195_clstr_insert_Fwd</td>
<td>TAATTAATCACACCTCTGTCTCACTC</td>
</tr>
<tr>
<td>195_clstr_insert_R</td>
<td>GGCGCGCCCAAGATTCTCTCAACATGGTG</td>
</tr>
<tr>
<td>miR-21_insert_F</td>
<td>TTAATTAACCAGAGATGTTTGCTTTTGC</td>
</tr>
<tr>
<td>miR-21_insert_R</td>
<td>GGCGCGCCGGACCACAAATTTTTGACTGC</td>
</tr>
<tr>
<td>miR-129-1_insert_F</td>
<td>TTAATTAATCCCTGACCCCATCTGC</td>
</tr>
<tr>
<td>miR-129-1_insert_R</td>
<td>GGCGCGCCCCTGGATCCACATGGTA</td>
</tr>
<tr>
<td>miR-141_insert_F</td>
<td>TTAATTAACACAGAGGCCACAGTCA</td>
</tr>
<tr>
<td>miR-141_insert_R</td>
<td>GGCGCGCCGTTAAGGTCAAGGCTGTG</td>
</tr>
<tr>
<td>miR-214_insert_F</td>
<td>TTAATTAATCCCCCTAAATCAACAAA</td>
</tr>
<tr>
<td>miR-214_insert_R</td>
<td>GGCGCGCCCCTCAGTGTTGTAGCT</td>
</tr>
<tr>
<td>miR-142_insert_F</td>
<td>TTAATTAAGCAGGAGCTGGCAGGAA</td>
</tr>
<tr>
<td>miR-142_insert_R</td>
<td>GGCGCGCCTTTCTTGTGGTGAGTG</td>
</tr>
<tr>
<td>302_clstr_insert_F</td>
<td>TTAATTAATTACTTTCTGTTGTTT</td>
</tr>
<tr>
<td>302_clstr_insert_R</td>
<td>GGCGCGCCGCGAGGGAGTTATACCATT</td>
</tr>
</tbody>
</table>
Chapter 4: Results

4.1 Multiple iPSC States

In an ongoing study in our laboratory, Dr. Peter Tonge has characterized a unique state of reprogramming cells, distinct from those iPSCs that cluster closely to ESCs in terms of gene expression. MEFs were transfected with the four reprogramming factors via piggyBac transposition and allowed to reprogram by culturing in dox media (Figure 4a). Clonal cell cultures derived from independent reprogramming events converge to this cell type, which he defines as F-class (for fuzzy-colony forming) iPSCs, in terms of morphology and transgene expression (Figures 4b and 4c). F-class iPSCs were also characterized by their pluripotency characteristics, such as the ability to produce teratomas when injected into mice (data not shown). Furthermore, gene
expression of F-class iPSCs sampled at day 16 after reprogramming factor expression closely resembles that of the same clonal line at day 30, indicating the stability of this cell state (Figure 4d). The derivation of F-class iPSCs has been attributed to the prolonged elevated expression of the four reprogramming factors.

4.2 Modeling iPSC States with a Secondary Reprogramming System

In order to investigate the reprogramming of somatic cells to both a type-L and an ESC-like state, a secondary reprogramming cell line was used, which had previously been generated in the lab (Woltjen et al., 2009). Initially, ROSA26-rtTA (reverse tetracycline transcription activator) MEFs were transfected with the four reprogramming factors (c-Myc, Klf4, Oct4, and Sox2), each downstream of a tet-O promoter (ROSA26 is a constitutively active promoter). These factors stably integrated into the host genome via piggyBac transposition. In the presence of both rtTA and dox, the tet-O promoter activates the expression of each of the transgenes. After culturing the transfected MEFs in dox media, a primary iPS clone was obtained and used to generate a chimeric embryo, from which secondary MEFs were derived. These MEFs undergo reprogramming on a population scale following the addition of dox, as opposed to the primary MEFs, which generated iPS cell colonies much less efficiently. This secondary system provided two main advantages over primary reprogramming: the abilities to i) study the progression of reprogramming from a single clonal sample, and to ii) sample and analyze the very early time-points of reprogramming.

To study the underlying molecular mechanisms at work during iPS cell generation, reprogramming was induced in these secondary MEFs by adding dox (1500 ng/mL) and collected samples daily until day 21 (Figure 5a). The
concentration of dox was maintained at this high level throughout reprogramming in one cell culture, and decreased at day 8 by 300-fold (to 5 ng/mL) in a second culture in order to generate F-class and ESC-like iPSCs, respectively.

Several of the time-points (days 0, 2, 5, 8, 11, 16, and 18 at high dox concentration, and days 16 and 21 at both low dox concentrations) were selected, along with rtTA ES cells, primary iPS cells (which gave rise to the secondary MEFs), and secondary iPS cells (derived by stopping dox treatment at day 21). Next generation sequencing was performed to quantify miRNA expression levels at all timepoints, in addition to complete proteome, transcriptome, and epigenome analysis (Figure 5b). Small RNA reads were mapped to the genome and quantified. Based on gene expression beadarray, samples that were maintained at a low dox concentration after day 8 clustered closely to both primary and secondary iPS cell samples, compared to the cells that were kept at high dox concentrations, which clustered closely to F-class primary iPSCs (Figure 5c). Additionally, the low-dox cultures exhibited compact ES cell-like colony formation by day 16. Cultures exposed to high levels of dox throughout the experiment contained many rapidly proliferating cells that were contributing to loose colony formation, just as was observed in the primary system.

This secondary reprogramming system provided a unique dataset, incorporating various OMICs platforms (transcriptomics, epigenomics, proteomics, etc.) which can be cross-referenced to each other in order to address the key questions surrounding the molecular mechanisms at work during the transition of somatic cells to pluripotency. Integral to the focus of this project was the validation of the small-RNA deep sequencing data.
Figure 4. Fibroblasts Reprogram to Predictable Stable States. (a) Primary fibroblasts were transfected with the Yamanaka factors in four independent piggyBac transposons. Colonies were selected at day 10 regardless of morphology and expanded for gene expression analysis. (b) Unsupervised hierarchical clustering of gene expression (beadchip) reveals two distinct groups of iPSCs, termed F-class and ESC-like iPSCs. Representative images of F-class iPSCs (clone 1), ESC-like iPSCs (clone 23), and ESCs. (c) qPCR of the reprogramming factors (exogenous and endogenous) in both iPSC groups. Non-parametric t-test, * p<0.05, *** p<0.001. (d) Scatterplot comparison of F-class iPSC clone 1 gene expression at day 16 and day 30 of transgene expression. Blue cutoff lines represent 4-fold difference in expression.
Figure 5. Secondary Reprogramming Model of iPSC States. (a) A schematic representation of the secondary reprogramming system. Y-axis indicates doxycycline levels and circles represent cells sampled for OMICs analysis. (b) Illustration of the various molecular platforms analyzed for each cell sample. (c) Pearson correlation heat map representation of unsupervised hierarchical clustering of gene expression (beadchip) of primary iPSCs (L-class and ESC-like), and secondary system cell samples. Dotted red line indicates the group of samples that cluster closely with F-class iPSCs.
4.3 Optimizing MicroRNA RT-qPCR methods

In order to validate the small-RNA deep sequencing data, and to develop a more feasible method of measuring miRNA expression for future experiments, a separate quantification method needed to be utilized. While many studies rely on ordering commercial quantification kits (Eichelser et al., 2013; Fareh et al., 2011), designed to specifically measure one miRNA species, this method proved costly when measuring the levels of dozens of miRNAs. It was thus decided to develop and optimize a miRNA quantitative PCR (qPCR) system, which allowed the design of custom primers for both characterized miRNAs and uncharacterized small RNA species.

What makes the quantification of miRNA so much more difficult than that of mRNA is the size of the RNA strand. miRNA is typically only 21-23 nucleotides long, the standard size of a primer. This makes the reverse transcription (RT) step of qPCR more complicated than with longer RNA species’. In order to properly convert miRNA to complementary DNA (cDNA), a modified stem-loop primer must be used. The stem-loop primer is a DNA molecule with a self-complimentary region at the terminal 5’ end which binds to its complement near the 3’ terminus, folding to create a hairpin-like structure (Figure 6a). This structure is designed to leave a small miRNA-specific overhang, which then anneals to the 3’ end of the miRNA, forming a double-stranded nucleic acid primer needed for DNA synthesis via reverse transcriptase. In our stem-loop design, this overhang is 6 nucleotides in length, followed by a 14 nucleotide stem (complementary region) and a 16 nucleotide loop (Clancy et al., 2007). The reverse qPCR primer was designed to anneal to the unfolded stem-loop primer, while the forward primer was designed to bind to the region of the miRNA that was not bound by the stem-loop primer during the RT step. This creates a 60 base-pair qPCR product, which is quantifiable using standard qPCR reagents.
There are several steps in this protocol that required optimization.

**4.3.1 Reverse Transcription**

Standard commercial miRNA qPCR protocols call for the addition of only one stem-loop primer in each RT reaction. This method of performing a separate RT reaction for each miRNA being quantified does not allow for proper internal control, nor does it make efficient use of the reaction reagents, such as the reverse transcriptase and the extracted miRNA. Combining multiple stem-loop primers should theoretically help reduce the impact of these two caveats. So, in order to test whether it was possible to accurately quantify miRNA by pooling several stem-loop primers together, 16 miRNAs were quantified using pooled RT reaction groups of 4, 7, 10, and 16 total stem-loop primers (Figure 6b). None of the resulting Ct values varied significantly from the others in each miRNA, and thus it was concluded that up to 16 stem-loop primers can be pooled in a single RT reaction without any loss of quantification accuracy.

**4.3.2 qPCR Primer Design**

The design of the forward primer is restricted to the nucleotides of the miRNA that were not bound by the stem-loop primer. Thus, an A/T-rich miRNA would produce a forward primer that has a much lower melting temperature than that of a G/C-rich miRNA. Multiple guanine and cytosine nucleotides were added to the 5’ end of the forward primer on all primers with a low melting temperature in order to bring the melting temperature up to 54-57°C. By doing so, the quantification of A/T-rich miRNAs was improved (Figure 6c).
4.3.3 Internal Controls

The typical internal control used with miRNA qPCR is small nucleolar RNA (snoRNA), such as snoRNA-202 (Brattelid et al., 2011), as it is a small enough species to be isolated with small RNA kits and is also present at relatively high and consistent levels throughout different cell types. However, in our dataset, snoRNA-202, along with 4 other snoRNAs that were also used in previous studies, varied up to $2^5$-fold from sample to sample when the same amount of total miRNA was quantified (Figure 7a). In order to ensure consistent, robust miRNA expression reads from qPCR that could be compared between sample time-points, finding a new internal control was necessary. In an effort to find such internal controls, our collaborators in Thomas Preiss’s lab identified miRNAs with low variation and high expression across all 16 sequenced time-points. 5 miRNAs, miR-16, -30e, -149, -191, and -484, were selected (Figure 7b). These miRNAs, when averaged and used to normalize miRNA expression in qPCR, showed improvements in Pearson Correlation scores in 14 out of the 16 miRNAs when compared to normalizing to RNA concentration alone (Figures 7c and 7d).
Figure 6. Quantitative RT-PCR Optimization for MicroRNAs. (a) A schematic representation of the steps involved in miRNA Quantitative RT-PCR. (b) miRNA qPCR for pools of stem-loop primers used in the Reverse Transcription (RT) step. Colour indicates the number of unique stem-loop primers in each RT reaction in equal molar ratios. Error bars represent the standard deviation of the qPCR triplicate. (c) Two standard curves of the same miRNA, one (green line) using Fwd1 primer (MT=40°C), and the other (blue line) using Fwd2 primer (MT=57°C), both using the same reverse primer.
Figure 7. Internal Normalizers for MicroRNA qPCR. (a) qPCR of various snoRNAs commonly used as controls across a subset of time-points in the secondary reprogramming dataset. Quantification performed using designed primers (DP) or commercially available TaqMan quantification kit (TM). (b) Expression levels from small RNA deep sequencing dataset of 5 miRNAs used as internal normalizers. Red line indicates the geometric average of the miRNA expression levels. (c) A comparison of measured expression levels of miRNAs obtained by qPCR and small RNA sequencing with and without normalizing to the 5 miRNAs in panel b. (d) Pearson correlation (R2) values of qPCR and small RNA sequencing for various miRNAs with and without normalizing to the 5 miRNAs in panel b.
4.4 Validating the Role of miR-302/367 Cluster in Reprogramming

In previous studies, the overexpression of the miR-302/367 cluster (302 cluster), which consists of 5 miRNAs (miR-302a, -302b, -302c, -302d, and -367) in a single polycistronic transcript, has been shown to induce efficient reprogramming in MEFs independent from any of the four transgenic reprogramming factors, when the cells are cultured with valproic acid (VPA). In order to replicate these findings, the 302-cluster first needed to be cloned.

4.4.1 Cloning the 302-Cluster

We cloned the 302-cluster by designing primers annealed to unique restriction sites up- and downstream of the 302-cluster (Figure 8a), subcloning the PCR fragment, excising the cluster and ligating it into the SBTC plasmid. Upstream of the miRNA cluster is an inducible tet-O promoter followed by the reporter gene mCherry, allowing visualization of the promoters activation (Figure 8b). The cluster is followed by a poly-A signal, and all elements are encompassed within the inverted terminal repeats of a Sleeping Beauty transposon. To test if the miRNAs would be processed correctly and produce mature miRNAs, this SBTC-302 plasmid was transfected into MEFs, along with a plasmid encoding both Sleeping Beauty transposase (SBase) as well as puromycin resistance. Cells were cultured with and without dox, which was added immediately following transfection. A second transfection was performed with an empty SBTC plasmid (i.e. lacking a miRNA cluster, but with all other elements unchanged) in place of SBTC-302. 24 hours after transfection, mCherry fluorescence was visible in the transfected MEFs cultured with dox, but not in the MEFs cultured without dox (Figure 8c). After 24 hours, puromycin was added to
the media in order to select against those MEFs that did not receive the SBase plasmid. Puromycin was left on for 48 hours, after which time the MEFs were allowed to proliferate in the absence of antibiotic selection for 48 hours. The cells were then collected and miRNA was extracted and quantified. Levels of miR-302d were detected as being over 10-fold higher in SBTC-302 transfected cells cultured with dox than those without, verifying both the sensitivity of the tet-O promoter as well as the correct processing of pre-miRNAs into mature miRNAs (Figure 8c).

4.4.2 Overexpression of the 302-Cluster in Reprogramming Cells

To replicate the findings of Anokye-Danso et al., the SBTC-302 plasmid was transfected into MEFs and cultured it with and without VPA. Separate transfections were also performed with the four transgenic factors with and without SBTC-302 to serve as a positive control. After 28 days in culture, no colonies had been generated from the MEFs transfected with the 302-cluster, with or without VPA, while many colonies had arisen in 4F and 4F+302 cultures. Similar experiments were performed several times with the same results. Furthermore, there was a significantly higher amount of colonies produced in 4F+302 cultures than 4F alone. This was detectable as early as day 14 (Figure 9). It was also noted that the 302-cluster causes an increase in reprogramming efficiency in cells co-transfected with three factors (i.e. without c-Myc) as well. Thus, while it was not possible to replicate the previous findings stating that the 302 cluster is able to reprogram cells by itself, the importance of the 302 cluster to reprogramming was validated by showing that the cluster increases the reprogramming efficiency of MEFs in both the four-factor and three-factor systems.
4.5 Screening Small RNA Sequencing Data for Candidate MicroRNAs

4.5.1 Identifying Candidate MicroRNAs

The 302-cluster provides an excellent example of the impact that the overexpression of a single miRNA cluster can have on reprogramming. In order to identify other miRNAs that have a similar effect, the secondary reprogramming database was used. It was posited that those miRNAs which promote reprogramming to an ESC-like state would be up-regulated in ESC-like iPS samples and down-regulated in F-class cells, just as is seen with the 302-cluster (Figure 10). Therefore, in order to derive a candidate list of miRNAs that could be implicated with an increase in reprogramming efficiency, miRNA expression was compared between samples collected at day 16 under high dox concentration (D16_H) with those at low dox concentration (D16_L). miRNAs whose expression differed less than 8-fold between these two samples were eliminated from consideration. A second filter, eliminating those miRNAs with less than an 8-fold expression difference between D16_H and secondary iPS cells, was applied to further ensure that each miRNA is truly differentially regulated in ES cell-like reprogramming cells. Finally, miRNAs with less than $2^6$ reads per million were not considered, as even small variances in counts between samples would result in significant fold up- or down-regulation.

After these three filters were applied, 38 candidate miRNAs remained (Figures 11a and 11b), 3 of which were up-regulated in the samples maintained at a high dox concentration. Because the primary focus of this study was to identify miRNAs which increase reprogramming efficiency when over-expressed, the 35 miRNAs which were found to be up-regulated in the low dox samples were considered. Both clusters and individual miRNAs were identified in these 35 candidates. Of the 35 miRNAs, 9 belonged to the miR-302/367 cluster (all 3p and
5p miRNAs in the cluster except for miR-367-5p), 8 to the miR-106a cluster (both 3p and 5p strands of miR-106a, -18b, -20b, and -363), 1 to a cluster on chromosome 14 homologous to the 106a cluster (miR-17-3p), and 1 to the miR-290 cluster (miR-290-3p). 7 miRNAs (miR-127-3p, -136-5p, -341-3p, -376b-3p, -485-3p, and -541-5p) were located in a large imprinted region on chromosome 12 which has been shown to lose methylation during reprogramming (Li et al., 2011a; Stadtfeld et al., 2011). Furthermore, miR-21-5p, a miRNA which has been shown to repress reprogramming by suppressing Oct4 (Yang et al., 2011), was among the 35 identified miRNAs. These aforementioned miRNAs, constituting 26 of the 35 identified miRNAs, have all had their reprogramming functions previously characterized. The functions of the remaining 7 individual miRNAs (miR-141-3p, -214-5p, -129-1-3p, -483-5p and -3p, and -142-5p and -3p) and one miRNA cluster (miR-195-5p and -497-5p) during reprogramming have not yet been studied.

4.5.2 Validation of Small RNA Deep Sequencing Data

Quantification of miRNA was compared between qPCR and small RNA deep sequencing (Figure 12). RT was undergone using the same miRNA samples used to generate libraries for small RNA deep sequencing. After performing the qPCR, both qPCR and deep sequencing data were normalized to the average quantification value, in order to compare the two platforms. High R² values indicated a close correlation in miRNA expression value between the platforms, suggesting robust validation of the deep sequencing data.
Figure 8. Cloning of MicroRNAs and MicroRNA Clusters. (a) The SB-TetO-mCherry (SBTC) plasmid used to overexpress the miR-302/367 cluster and other miRNAs. miRNAs are inserted in between two unique restriction enzymes (Ascl and PacI). (b) Overexpression of miRNAs in MEFs using the SBTC plasmid cultured with and without dox. Fluorescent images taken 24h after transfection and dox addition. Bar graphs indicate qPCR of mature miRNA that was cloned into the SBTC plasmid in cells cultured with and without dox. All samples are compared to MEFs transfected with an empty SBTC plasmid.
miR-302/367 Cluster
miR-141
miR-142
miR-214
miR-21
miR-129-1
miR-10b

+ dox - dox
neg ctrl

miRNA Expression (% of dox+)

+ dox - dox
neg ctrl

+ dox - dox
neg ctrl

+ dox - dox
neg ctrl

+ dox - dox
neg ctrl

+ dox - dox
neg ctrl

+ dox - dox
neg ctrl

0%
50%
100%

miRNA Expression (% of dox+)
Figure 9. The 302-Cluster and Reprogramming. Primary MEFs were transfected with STBC-302 (302-cluster), the four reprogramming factors (4 factors), or both (4 factors + 302-cluster). Cells were plated equally in 3 separate wells and maintained in dox media. MEFs that were only transfected with SBTC-302 were cultured with and without VPA. Every 7 days, one well from each transfection group was AP stained and photographed.
Figure 10. 302-Cluster Expression in the Secondary Reprogramming System. Small RNA deep sequencing data of all 5 miRNAs of the 302-cluster (miR-302a, -302b, -302c, -302d, -367) at all time-points of the secondary reprogramming dataset.
Figure 11. Deriving Candidate MicroRNAs from Small RNA Deep Sequencing Data. (a) Scatterplot of miRNA expression in D16_H compared to D16_0. Filters applied to data are described in the legend, and appear as colour-coded points on the scatterplot. Black points represent miRNAs that pass all three filters. (b) 35 of the 38 miRNAs represented by black dots are upregulated in D16_0. They are made up of 4 characterized clusters, 1 uncharacterized cluster, 1 characterized miRNAs, and 5 uncharacterized miRNAs.
Figure 12. Validation of Small RNA Deep Sequencing Data with MicroRNA qPCR. Line graphs showing the correlation between small RNA deep seq data (orange line) and miRNA qPCR data (green line), measured from the same miRNA sample. qPCR data was normalized to the 5 miRNA internal controls from (Figure 7b).
4.6 Validation of Candidate MicroRNAs in the Primary Reprogramming System

While these identified miRNAs are over-expressed in ES-like iPSCs in the secondary system, it is important to measure their expression levels in a primary system. Consistency between the two reprogramming systems would indicate that the correlation between these miRNA and ESC proximity is biological in nature and not simply an artifact of the secondary system. In order to compare the two systems, miRNA from several primary iPSC cultures (5 F-class iPSC lines and 3 ESC-like iPSC lines) was extracted and sequenced. Furthermore, three-factor iPSCs were generated and subcultured, allowing for further comparison to ESC-like iPSCs. 3 of these three-factor clones were chosen for miRNA extraction and deep sequencing.

While the 302- and 290-clusters are more highly expressed in ESC-like iPSCs than F-class iPSCs, they both display even higher expression in three-factor iPSCs (Figure 13). This expression pattern is clearly visible in miR-141, miR-142, and miR-195. miR-483 and miR-129-1 expression is higher in the three-factor iPSCs than in the F-class iPSCs, while miR-21 and miR-214 do not display an increase in expression in three-factor or ESC-like iPSCs over F-class iPSCs.

4.7 Elucidating the Effects of Candidate MicroRNAs on Reprogramming

In order to test whether the novel miRNAs identified in the aforementioned data screen impact reprogramming when over-expressed, as the 302 cluster does, each miRNA was cloned into the SBTC plasmid. Just as the 302 cluster was cloned, primers with unique restriction sites annealed to their 5’ ends were
designed to flank each miRNA or miRNA cluster, leaving a minimum of 40nt in
between the miRNA and each primer to allow for proper miRNA hairpin
formation. Along with the 7 novel miRNAs and 1 novel cluster, the 1 miRNA (miR-
21) as well as 1 of the other miRNA clusters (the 290-cluster) that had been
previously implicated in reprogramming were also cloned into this construct.
Each of the SBTC-miRNA constructs were transfected into MEFs, puromycin
resistance-selected, and collected for miRNA extraction and quantification, just
like SBTC-302 (Figure 9c). The majority of SBTC-miRNA constructs displayed a
large up-regulation of the respective miRNAs when cultured with dox compared
to the controls. Two miRNAs, miR-21 and miR-214, did not produce as large of a
miRNA expression difference as the other cloned miRNAs. This is most likely due
to the high expression levels of these miRNAs in MEFs. Since the basal levels of
these miRNAs are already very high in MEFs, forcing more of the same miRNAs
to be expressed would still increase the amount of that miRNA in the cell, but it
would not impact the fold-difference in expression as much as the forced
expression of a less abundant miRNA would.

To quantify the effect that each miRNA had on reprogramming efficiency,
primary MEFs were transfected with the three reprogramming factors (Oct4,
Sox2, and Klf4) along with one of the miRNAs/miRNA clusters. MEFs were also
transfected with an empty SBTC plasmid as well as with no extra plasmid in
order to quantify for baseline colony formation (without the addition of miRNAs).
Transfected cells were plated and cultured on four separate dishes, each of
which was maintained with dox media. At days 7, 14, 21, and 28, cells were AP
stained and photographed, and the number of colonies in each dish was
quantified using feature-counting software (Volocity) (Figures 14a and 14b).
Those miRNAs that appeared to cause an increase in the number of colonies
over baseline were used again in repeats of this experiment (Figure 14c). The
largest impact on reprogramming efficiency was made by the 302 cluster, which
increased the number of colonies by up to 5-fold over baseline. Of all the
miRNAs tested, including those that had been previously defined to increase reprogramming efficiency, the only one to increase the number of colonies significantly over the control was miR-214.
Figure 13. Candidate MicroRNA Levels in Primary iPSCs. Small RNA deep sequencing data from F-class 4 factor (n=5), ESC-like 4 factor (n=3) and three-factor primary iPSC clones. Horizontal lines represent the mean, and vertical bars represent standard deviation.
Figure 14. Overexpression of Candidate MicroRNAs in Reprogramming MEFs. (a) MEFs were transfected at day 0 with the three reprogramming factors and one SBTC-miRNA plasmid. AP stains were performed days 7, 14, 21, and 28. (b) Plot of the number of AP+ colonies, counted using Volocity software (n=1). (c) miRNAs that generated more iPSCs than the control were repeated (n=3). * p<0.05, ** p<0.01 (2-tailed paired t-test).
<table>
<thead>
<tr>
<th>Results Section</th>
<th>Collaborators</th>
<th>Work Done by Collaborators</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Multiple iPSC States</td>
<td>Dr. Peter Tonge</td>
<td>• 100% of this section</td>
</tr>
<tr>
<td>4.2 Modeling iPSC States with a Secondary Reprogramming System</td>
<td>Dr. Peter Tonge, Dr. Mira Puri, Dr. Claudio Monetti</td>
<td>• 100% of this section</td>
</tr>
<tr>
<td>4.3 Optimizing miRNA qPCR Methods</td>
<td>Dr. Jennifer Clancy</td>
<td>• Provided RT-qPCR primer design technique • Identified miRNAs for internal normalization</td>
</tr>
<tr>
<td>4.4 Validating the Role of the 302/367-Cluster in Reprogramming</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>4.5 Screening Small RNA Sequencing Data for Candidate miRNAs</td>
<td>Dr. Jennifer Clancy, Dr. Hardip Patel, Dr. Brian Parker</td>
<td>• Sequenced and normalized small RNA deep sequencing data</td>
</tr>
<tr>
<td>4.6 Validation of Candidate miRNAs in the Primary Reprogramming System</td>
<td>Dr. Peter Tonge</td>
<td>• Provided four-factor Primary iPSCs</td>
</tr>
<tr>
<td>4.7 Elucidating the Effects of Candidate miRNAs on Reprogramming</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5: Discussion

The role that miRNAs play during reprogramming is only beginning to be fully understood. Studies prior to the discovery of reprogramming have shown that miRNAs play a necessary role in pluripotency maintenance (Kanellopoulou et al., 2005; Wang et al., 2007), and recent work has made it clear that miRNAs are essential to the reprogramming process (Kim et al., 2012). Indeed, multiple ESC miRNAs have been shown to promote iPSC generation by increasing reprogramming efficiency (Judson et al., 2009; Wang et al., 2013; Yang et al., 2011), and one miRNA cluster has even been suggested to induce the reprogramming process, completely devoid of any of the Yamanaka factors (Anokye-Danso et al., 2011). Yet it is likely that many miRNAs that are important for reprogramming have not been identified. This is possibly due to the innate difficulty of identifying those cells that will become ESC-like iPSCs and those that
will progress to a non-ESC-like state. A platform that can enrich for those cells that will become proper iPSCs is required to identify and study the early stages of reprogramming.

Recently, unpublished work by Dr. Peter Tonge has proven the existence of a stable reprogramming state, termed the F-class state, distinct from the ESC-like iPSC state that is commonly studied. Others have largely ignored this iPSC state, describing the cells as partially reprogrammed and considering it a failed reprogramming event. To fully characterize the progression of reprogramming fibroblasts to both the F-class and the ESC-like iPSC states, a massive experiment was designed to sample a secondary reprogramming cell culture at various time-points as it progressed to both iPSC states, and to analyze the transcriptomics, proteomics, epigenomics of each sample in parallel. In the current study, this unique reprogramming dataset was screened to identify miRNAs that could potentially promote iPSC generation. miRNAs were identified that are up-regulated in ESC-like iPSCs over the F-class iPSCs, and the effect that these miRNAs have on reprogramming efficiency was further investigated.

5.1 Optimization of MicroRNA RT-qPCR

The first aim of this project was to optimize miRNA qPCR in order to allow for reliable and efficient measurement of reprogramming, as well as the validation of the small RNA deep sequencing data. Optimization of this process was achieved by adjusting three separate parameters of the qPCR process: reverse transcription, qPCR primer design, and internal normalization.
Normally, reverse transcription utilizes poly-dT primers, but miRNA’s do not have polyA tails, and thus each miRNA species requires independent stem-loop primers in independent RT reactions. It was found pooling up to 16 stem-loop primers in a single RT reaction did not significantly impact the quantification detection. While stem-loop primer pooling is not a novel concept, it was necessary to validate a method that was new to the laboratory. Being able to pool stem-loop primers not only allowed for savings on cost and time preparing RT and qPCR reactions, but more importantly, it allowed for the use of one common stock for all the cDNA, reducing miRNA usage and making internal normalization more robust and reducing technical error.

The nature of miRNA qPCR does not allow flexibility in forward primer design as the majority of the miRNA will inevitably become the forward primer sequence. Because of this, the melting temperature of the forward primer can only be manipulated by annealing base pairs that will not bind the template cDNA. By adding purine nucleotides to the 5’ end of the forward primers, it was possible to raise the melting temperature to a range similar to the reverse primer (54-57°C). This was demonstrated by improved 5-point standard curves, demonstrating reliable quantification by the primers.

Internal normalization is critical to qPCR in order to make quantification data comparable across samples. The use of snoRNAs, the accepted norm for internal normalization of miRNAs, was not appropriate, due to large perturbations in overall snoRNA levels throughout the secondary reprogramming dataset. Suitable substitute for snoRNAs were found by identifying 5 miRNAs that had both high expression and low variance throughout all secondary reprogramming time-points. These internal normalizers, when quantified and averaged, improved the correlation between qPCR and deep sequencing in many of the examined miRNAs. While these 5 miRNAs greatly improved qPCR normalization over the
typical snoRNA normalizers, they may not be applicable for normalization outside of this system.

The combination of these optimization parameters allowed for the validation of the small RNA deep sequencing data. This was an important finding, since a great deal of information can be gained from this deep sequencing dataset, especially when combined with the other parallel OMICs platforms. Since the collection of the reprogramming time-points was not repeated, verification of the data within the same biological sample was necessary for any findings to be considered valid.

5.2 Identification of Candidate MicroRNAs

The strategy employed to identify candidate miRNAs was to find miRNAs that were up-regulated in ESC-like iPSC samples versus F-class iPSC samples. Two of the most studied miRNA clusters in reprogramming, the 302-cluster and the 290-cluster, displayed high expression in ESC-like iPSC samples and low expression in F-class samples. This, combined with the positive effect these clusters have on reprogramming (Judson et al., 2009; Subramanyam et al., 2011), sparked the idea that the presence of these miRNAs may be driving reprogramming cells to the ESC-like iPSC state. Thus, it was hypothesized that other miRNAs displaying this same pattern of expression may play a similar role in reprogramming.

The stringent filters used to screen miRNA deep sequencing data were chosen to produce a small number of reliable hits. The main filter—the difference between D16_0 (ESC-like) and D16_H (F-class)—was supplemented with a
second filter—the difference between 2° iPSCs and D16_H—in order to exclude samples whose high expression at D16_0 represented a statistical outlier and not a biological effect. A third filter—the expression threshold—was added to reduce the background noise from very lowly expressed miRNAs, in which even a small fluctuation in transcript number would have a large effect on fold-change in expression. A large subset of the total hits that resulted from this screen were miRNAs that had been previously identified as enhancers of reprogramming. This validated our rationale that enhancers of reprogramming would display similar expression patterns to the 302- and 290-clusters.

5.3 Validation in a Primary Reprogramming System

The next aim in this study was to investigate the expression levels of each candidate miRNA in a primary reprogramming system. This aim was especially important to the hypothesis because it addressed a central issue regarding the validity of using a secondary reprogramming system. The primary and secondary reprogramming systems have many fundamental differences, and different molecular mechanisms may be activated. Thus, any differences in miRNA expression could, theoretically be merely an artifact of secondary reprogramming and not a hallmark of reprogramming in general.

While the small RNA deep sequencing data from the primary iPSCs validated the differential expression of many of the candidate miRNAs, it revealed mixed results for others. Both of the controls (302- and 290-clusters) showed a modest increase in expression from F-class iPSCs to the ESC-like iPSCs (including three-factor iPSCs). While this same general pattern was observed in 5 of the 7 candidate miRNAs, the difference between the F-class and the three-
factor iPSCs was consistently more pronounced than between the two four-factor iPSC states. This lends support to the claim that three-factor iPSCs consistently reprogram to a state that closely resembles ESCs. The remaining two miRNAs did not display the same expression pattern in the primary system as the 302- and 290-clusters. Interestingly, both of these miRNAs were expressed at very high levels in all three groups of iPSCs (F-class, ESC-like and three-factor). While the findings of the primary reprogramming deep sequencing suggest that these miRNAs are inconsequential to reprogramming and that the expression difference in the secondary system was merely a coincidence, the decision was made not to eliminate these miRNAs from further examination.

5.4 Effects of Candidate MicroRNAs on Reprogramming

5.5.1 miR-214 as a Novel Enhancer of Reprogramming

miR-214 was the only miRNA that showed a statistically significant increase in reprogramming efficiency over the controls. The role that this miRNA plays during reprogramming has not yet been defined, but several studies have characterized miR-214 in cancer and other diseases. miR-214 has been shown to have anti-apoptotic effects and to promote cell survival (Yang et al., 2008). It has also been shown to promote muscle differentiation by regulating the Hedgehog signaling pathway (Flynt et al., 2007), further implicating it with cell differentiation. Interestingly, c-Myc has also been shown to regulate this pathway in Burkitt Lymphoma and Medulloblastoma (Rao et al., 2003; Yoon et al., 2013). Since miR-214 was found to up-regulate iPSC generation in three-factor reprogramming fibroblasts without the expression of c-Myc, it follows that the effect of this miRNA was to act in c-Myc’s place, promoting proliferation. It would
be interesting to see if miR-214 has the same iPSC-enhancing effects in four-factor reprogramming. Furthermore, miR-214 has been shown to target PTEN, a well-characterized tumour suppressor gene (Yang et al., 2008), which is also targeted by miR-21 (Meng et al., 2007). In addition to being one of the hits of the screen performed in the current study, miR-21 was one of the three miRNAs that increased the number of iPSC colonies generated from three-factor reprogramming, albeit without statistical significance. These findings, taken together with the aforementioned studies, imply that miR-214 acts in the same pathway as miR-21 in terms of regulating reprogramming. However, miR-21 has been defined as an antagonist to reprogramming, and has been shown to be repressed by c-Myc expression (Yang et al., 2011). This apparent contradiction warrants further investigation into the specific roles of miR-21 and -214 in reprogramming. It is possible that both of these miRNAs both possess dual roles that depend on the expression status of c-Myc.

The bump in efficiency experienced when miR-214 was overexpressed, however, did not compare to the massive increase in iPSCs formed when the three factors were co-transfected with the 302-cluster. miR-214 was shown to be expressed at high levels throughout reprogramming in the secondary system, and overexpression of this miRNA only increased expression by approximately 50%. Additionally, the 302-cluster gives rise to 5 mature miRNAs per transcript, whereas miR-214 is the only mature miRNA expressed at its locus. This points to the possibility that the lesser impact on reprogramming efficiency brought about by miR-214 is due to an overall lower number of mature miRNAs processed by the cell.
5.4.2 The Reprogramming Effects of the 302-Cluster

The large increase in reprogramming efficiency produced by the 302-cluster not only serves as a positive control for this experiment, but also replicates the findings, made in previous studies (Judson et al., 2009; Subramanyam et al., 2011), that the 302-cluster enhances reprogramming efficiency. Furthermore, this serves to validate the methodology of testing the impact of miRNA expression on iPSC generation, making the findings of the current study robust.

While overexpression of the 302-cluster in four- or three-factor reprogramming cells caused an increase in iPSC generation, it was not possible to replicate the findings of Anokye-Danso et al.—that reprogramming can be induced by the forced expression of the 302-cluster (Anokye-Danso et al., 2011). This may be due to factors that varied between experimental set-ups, such as vector delivery method. However, this is not likely, as when the same plasmid that was used in this study was transfected with the Yamanaka factors, there was an increase in reprogramming efficiency indicating that the 302-cluster was indeed being expressed and functional. Furthermore, as stated earlier, the ability of the 302-cluster to induce reprogramming without the addition of any of the four Yamanaka factors has not been reproduced by independent researchers, suggesting an error in the initial finding.

5.4.3 Potential Roles of Other Candidate MicroRNAs

Other promising candidate miRNAs identified did not prove to promote iPSC generation. Despite displaying very similar expression patterns to the 302- and 290-cluster in both the primary and secondary systems, it appears that neither miR-142 nor miR-141 have a significant effect on primary iPSC
generation when overexpressed. Both these miRNAs have been implicated with cell survival in cancers. miR-141 has been shown by numerous sources to be disregulated in renal carcinoma (Kong et al., 2013; Liu et al., 2010), with one study suggesting that by targeting CDC25B, a known cell division cycle regulator, it regulates cell survival (Kong et al., 2013). Interestingly, miR-142 has been classified as both a tumour suppressor in colon cancer (Shen et al., 2013), and an antagonist to a tumour suppressor in testicular cancer (Tanaka et al., 2013). Despite the contradiction in the literature, it is apparent that miR-142, along with miR-141, functions as a regulator of cell proliferation. The roles that these miRNAs play in cancer, combined with the present observations that they do not significantly impact reprogramming efficiency, suggest that these miRNAs may be involved in pluripotency maintenance rather than the progression of a reprogramming cell to an ESC-like iPSC state.

5.5 Analysis of the Hypothesis

It was hypothesized that an overexpression of those individual miRNAs specific to an ESC-like iPSC state will increase the efficiency of reprogramming in fibroblasts. Our study has shown that while it is possible to enhance reprogramming with the overexpression of one miRNA, as demonstrated with miR-214, this is not the case for all miRNAs significantly upregulated in ESC-like iPSCs. Additionally, the increase in reprogramming efficiency produced by miR-214 paled in comparison to the increase displayed by the 302-cluster, indicating that it might take several miRNAs in the same family to bring about a large effect on iPSC generation. Combining various candidate miRNAs may be an interesting path to take moving forward.
Chapter 6: Conclusion & Future Directions

Overall, this study has shown that while some miRNAs or miRNA clusters are able to increase the efficiency of reprogramming, this is not a characteristic of all miRNAs specific to an ESC-like iPSC state. While some miRNAs displayed similar expression patterns to iPSC-promoting miRNAs, they were not able to affect the generation of iPSCs \textit{de novo}. Conversely, the only miRNA discovered here to have a positive impact on reprogramming did not display the same pattern in the primary system as the 302- and 290-clusters. This suggests that the expression level differences of miRNAs between the iPSC states do not necessarily produce an effect on the reprogramming process. This could be explained by an alternate function possessed by the miRNA that is not involved in the progression of reprogramming cells to an ESC-like state. Alternatively, these miRNAs may indeed be involved with a pathway that promotes iPSC
generation, yet have no function or ability to reprogram when overexpressed alone. In this study, miRNAs were only co-expressed along with other miRNAs if they were expressed in a known miRNA cluster. miRNAs of the same family, that target the same mRNAs or that had similar seed sequences were not taken into consideration. These avenues could all prove to be fruitful if pursued.

In the screen of small RNA deep sequencing data performed to generate the candidate miRNA list, the three miRNAs that were down-regulated in ESC-like iPSCs compared to F-class iPSCs were ignored, as they represented the minority of miRNA expression patterns. It may be interesting to see the effect of knocking these miRNAs down by creating miRNA sponges, which competitively bind particular families of miRNAs, inhibiting their repression of their targets. While it is expected that the down-regulation of these miRNAs would enhance reprogramming, these miRNAs may need to be inhibited in combinations with other miRNAs. Additionally, synergistic effects might be uncovered by combining the forced expression of the candidate miRNAs with the down-regulation of these miRNAs. One way of tackling all of these questions at once would be to co-transfect the Yamanaka factors with individual inducible plasmids regulating all suspected enhancers of reprogramming as well as with miRNA sponges targeting all suspected reprogramming repressors, all in equal molar proportions. The combinations of miRNAs/sponges that increase reprogramming efficiency to the highest degree will produce a higher number of iPSC colonies. Thus, picking a large number of colonies and genotyping them to identify which miRNA/sponges were integrated into the genome would reveal the bias in factor combination, implying a synergistic effect.

Those miRNAs that are shown to enhance reprogramming may also be able to replace one or more of the reprogramming factors. Identifying such miRNAs would provide alternate reprogramming protocols, potentially increasing the efficiency of iPSC generation or eliminating the use of oncogenes. One way
to identify miRNAs with these qualities would be to design a screen in which the expression of candidate miRNAs is induced in MEFs that express c-Myc plus two of the three other Yamanaka factors. Such MEFs would not be able to reprogram unless a substitute for the missing reprogramming factor was expressed. Transfecting MEFs with three of the four reprogramming factors may be one way to perform this screen, yet this would be very inefficient, as the proper combination of transgenes and miRNA plasmids would all have to be transfected into the same cell. In order to increase the efficiency of this screen, a good strategy would be to begin by culturing primary iPSCs that had been derived using dox-inducible reprogramming plasmids and delivered via piggyBac transposition. These iPSCs would be characterized so that the exact number of each reprogramming factor integrated into the cell’s genome would be known. This can be achieved by performing a southern blot with probes specific to each reprogramming factor. Once a colony is identified with one insertion of each transgene, those iPSCs would be transfected with piggyBac transposase, randomly mobilizing the transposons, allowing for the reintegration or loss of reprogramming factors. iPSCs would then be cultured, subcloned, and screened for integration of the four reprogramming factors by gDNA PCR. This should produce several iPSC lines with various combinations of the Yamanaka factors, all under dox-inducible control. iPSCs with the desired combination of transgenes incorporated into their genome would then be used to derive a mouse embryo via tetraploid complementation, which would give rise to secondary MEFs, each with the same transgene insertions. These secondary MEFs could then be transfected with a broad array of plasmids, each driving the expression of a different miRNA. If the miRNAs that are used in this screen integrate into the genome, only one transfection is necessary, as the identity of the miRNA can be determined after iPSC formation. After transfection, the MEFs would be cultured in dox media, allowing for reprogramming factor expression. Untransfected secondary MEFs should be used as controls in case transgene excision failed and reprogramming was still possible without additional factors. After several weeks, only those
MEFs that were supplemented with reprogramming factor-replacing miRNAs would be able to generate iPSCs, and thus any colonies produced would represent a positive result. gDNA from these colonies would be collected and used to determine which miRNA replaced the function of the missing reprogramming factor.

The ultimate goal of this type of screen would be to find miRNAs that could replace all four of the reprogramming factors (or at least Oct4 Sox2, and Klf4), allowing for miRNA-induced reprogramming. This type of screen does not need to be performed in such a way that only one factor-replacing miRNA would be identified at a time. For example, secondary MEFs could be generated that only include two, or even one reprogramming factor and the same type of screen could be attempted. Any miRNAs, or combinations of miRNAs, would then replace multiple reprogramming factors. However, there may not exist a combination of miRNAs that can replace the effects of one of the reprogramming factors, and thus no positive hits would ever result from a screen that did not include that irreplaceable factor. If this is indeed the case, then screening for miRNA replacements of reprogramming factors one at a time may be the best strategy.

These factor-removed secondary MEFs could be used for other screens as well, such as small molecule or siRNA screens. The difference between the methodology behind these screens and the aforementioned screen is that small molecules and siRNA do not incorporate into the genome. This is advantageous to the desired result of these screens—to identify non-integrating reprogramming reagents—but it makes the screen more tedious. Secondary MEFs must be plated in 96-well plates and cultured with dox media. Cells would then be either transfected with siRNA or have their media supplemented with a small molecule. Cells in each well would then be AP stained and screened for colony formation.
This technique could generate a large number of possible replacements for the reprogramming factors.


Expression Signatures. Stem Cell 5, 111–123.


Fareh, M., Turchi, L., Virolle, V., Debruyne, D., Almairac, F., Divonne, S.D.-L.-F.,


Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y., Cai, J., Guo, X.,


Lyssiotis, C.A., Foreman, R.K., Judith Staerk, Garcia, M., Mathur, D., Markoulaki,


massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Research 18, 610–621.


Pan, C., Lu, B., Chen, H., and Bishop, C.E. (2009). Reprogramming human fibroblasts using HIV-1 TAT recombinant proteins OCT4, SOX2, KLF4 and c-


2025.


