2012-03-12

Dissecting Somatic Cell Reprogramming by MicroRNAs and Small Molecules: A Dissertation

Zhonghan Li
University of Massachusetts Medical School

DISSECTING SOMATIC CELL REPROGRAMMING BY MICRORNAS AND SMALL MOLECULES

A Dissertation Presented

By

Zhonghan Li

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MARCH 12th, 2012

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

DISSECTING SOMATIC CELL REPROGRAMMING BY MICRORNAS AND SMALL MOLECULES

A Dissertation Presented

By

ZHONGHAN LI

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

_____
Tariq M. Rana, Ph.D., Thesis Advisor

_____
Zuoshang Xu, M.D., Ph.D., Member of Committee

_____
Kendall L. Knight, Ph.D., Member of Committee

_____
Gary S. Stein, Ph.D., Member of Committee

_____
Danesh Moazed, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets
the requirements of the Dissertation Committee

_____
Charles G. Sagerstrom, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies
that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

Program
Program in Biochemistry and Molecular Pharmacology
March 12th, 2012

COPYRIGHT INFORMATION

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. Tariq M. Rana, for giving me such a wonderful chance to work in his lab. Tariq has been a great mentor with passion for science, who has guided me with unlimited support to overcome all the hurdles and pursue what inspires me the most. I enjoyed every day working in his lab. I also would like to thank the alumni of Rana lab (Katsu, Chia-ying, Chih-chung, Huricha, and Indrani) and all the current lab members (Chao-shun, Nianwei, Jingya, Veena, Kumi, Xinpeng, Tianxu, Ying, Gian, Nadia, Maloy, Jason, Baiyuan, Ahsan, Gati and Idrees) for their help and friendship during the past few years.

In addition, I want to acknowledge members of my Thesis Committee, Drs. Charles G. Sagerstrom, Kendall L. Knight, Gary S. Stein and Zuoshang Xu, for their insightful suggestions on my research and generous help on my future career. I would also like to thank Dr. Danesh Moazed for being the external member of my Thesis Committee and sharing his input on my research. My appreciation shall also be extended to all the people from both UMass Medical School and Sanford-Burnham Institute, for their help and support on my research, especially Christine Pruitte, Ling Wang, Lili Lacarra and Subu Govindarajan.

Finally and most importantly, I would like to thank my parents, for giving me their unconditioned love, and my wife for her constant love and unwavering support during all these years.

ABSTRACT

Somatic cells could be reprogrammed into an ES-like state called induced pluripotent stem cells (iPSCs) by expression of four transcriptional factors: Oct4, Sox2, Klf4 and cMyc. iPSCs have full potentials to generate cells of all lineages and have become a valuable tool to understand human development and disease pathogenesis. However, reprogramming process suffers from extremely low efficiency and the molecular mechanism remains poorly understood.

This dissertation is focused on studying the role of small non-coding RNAs (microRNAs) and kinases during the reprogramming process in order to understand how it is regulated and why only a small percentage of cells could achieve fully reprogrammed state. We demonstrate that loss of microRNA biogenesis pathway abolished the potential of mouse embryonic fibroblasts (MEFs) to be reprogrammed and revealed that several clusters of mES-specific microRNAs were highly induced by four factors during early stage of reprogramming. Among them, miR-93 and 106b were further confirmed to enhance iPSC generation by promoting mesenchymal-to-epithelial transition (MET) and targeting key p53 and TGFβ pathway components: p21 and Tgfbr2, which are important barrier genes to the process.

To expand our view of microRNAs function during reprogramming, a systematic approach was used to analyze microRNA expression profile in iPSC-enriched early cell population. From a list of candiate microRNAs, miR-135b was

found to be most highly induced and promoted reprogramming. Subsequent analysis revealed that it targeted an extracellular matrix network by directly modulating key regulator Wisp1. By regulating several downstream ECM genes including Tgfbi, Nov, Dkk2 and Igfbp5, Wisp1 coordinated IGF, TGFβ and Wnt signaling pathways, all of which were strongly involved in the reprogramming process. Therefore, we have identified a microRNA-regulated network that modulates somatic cell reprogramming, involving both intracellular and extracellular networks.

In addition to microRNAs, in order to identify new regulators and signaling pathways of reprogramming, we utilized small molecule kinase inhibitors. A collection of 244 kinase inhibitors were screened for both enhancers and inhibitors of the process. We identified that inhibition of several novel kinases including p38, IP3K and Aurora kinase could significantly enhance iPSC generation, the effects of which were also confirmed by RNAi of specific target genes. Further characterization revealed that inhibition of Aurora A kinase enhanced phosphorylation and inactivation of GSK3β, a process mediated by Akt kinase. All together, in this dissertation, we have identified novel role of both small non-coding RNAs and kinases in regulating the reprogramming of MEFs to iPSCs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

1.1 Introduction

Since the discovery in *Caenorhabditis elegans (C. elegans)* that double-stranded (ds) RNA could trigger a potent gene-specific silencing phenomenon, termed RNA interference (RNAi) (Fire et al, 1998), considerable effort has been made in many biological disciplines to address some of the fundamental questions surrounding RNAi. For example, is RNAi a general mechanism for gene regulation that is conserved across species? What are the physiological triggers of RNAi and how does it play a role in biological processes? Work designed to address such questions have led to the recognition that RNAi is a widespread natural phenomenon that is conserved across fungi, plants, and animals.

Long dsRNAs generate potent RNAi and silence target genes by inducing cleavage of their mRNA. However, in mammals, long dsRNA activates the innate immune response by inducing interferon pathways. Further mechanistic studies led to the discovery that mRNA cleavage induced by RNAi was guided by small ~21 nucleotide (nt) RNA fragments derived from long dsRNAs, which revealed that these small interfering RNAs (siRNAs) are the essential triggers for RNAi. Since these discoveries were made, great effort has been directed at identifying endogenous physiological triggers that have similar properties to siRNAs. Several endogenous small RNA species have been identified, including small non-coding RNAs (microRNAs), piwi-interacting RNAs (piRNAs), and endogenous siRNAs (endo-siRNAs).

MicroRNAs (miRNAs) are single-stranded RNAs ~21 nt in length that are involved in almost every area of biology, including developmental processes, disease pathogenesis, and host-pathogen interactions (Ambros, 2011; Kim et al, 2009a; Krol et al, 2010). The biogenesis of mature miRNAs relies mainly on digestion of the precursor RNA hairpin structure by two members of the RNase III family, Drosha and Dicer, while other miRNAs can be generated through splicing of miR-coding introns. MicroRNAs are loaded into a functional ribonucleoprotein assembly called the RNA-induced silencing complex (RISC), which serves as the catalytic engine for miRNA-mediated post-transcriptional regulation. Although some studies have suggested a potential role for miRNAs in translational activation, the more common mechanism of miRNA-mediated gene regulation involves repression. In general, miRNAs bind imperfectly to the 3‘ UTR of target mRNA and block their expression by directly inhibiting the translational steps and/or by enhancing mRNA destabilization (Bagga et al, 2005; Fabian et al, 2010; Guo et al, 2010).

Piwi-interacting RNAs, piRNAs, are germ cell-specific and larger than miRNAs, spanning ~24–29 nts in length. piRNAs were discovered in *Drosophila melanogaster* development studies and most of these RNAs matched to intergenic repetitive element sequences including retrotransposons. Distinct from miRNAs, piRNAs directly interact with Piwi proteins and have been shown to regulate transposon activities in *Drosophila* (Klattenhoff & Theurkauf, 2008). piRNAs associate with Piwi proteins and their biogenesis does not involve

Drosha or Dicer activities. Although not tested, it is possible that Piwi proteins provide nuclease function to generate piRNAs. piRNAs are encoded in clusters throughout the genome. Since Piwi proteins exhibit RNA cleavage activities, a unique amplification loop has been proposed for piRNA biogenesis, in which each piRNA-mediated cleavage creates the 5‘ end of a new piRNA (Brennecke et al, 2007).

A third class of small RNAs, the endo-siRNAs, was originally discovered in *Drosophila* (Czech et al, 2008; Ghildiyal et al, 2008) where they were shown to be expressed in both gonadal and somatic tissues, and bind mainly to the Ago2 protein (Czech et al, 2008). Endo-siRNAs can be generated from such distinct loci as transposon elements, natural antisense transcripts (NAT), and pseudogenes, as well as from other long hairpin mRNAs (Czech et al, 2008; Ghildiyal et al, 2008). However, the biogenesis of these small RNAs remains unclear and, at least in Drosophila, requires the involvement of the protein Loquacious (LOQS) (Czech et al, 2008). Classes of small RNAs in various species, and their origin and function are summarized in Table 1.1.

Although piRNAs and endo-siRNAs exhibit an interesting ability to regulate certain genomic loci elements, miRNAs are the most abundant species of small RNAs in mammalian cells. Despite their importance in biology and medicine, the molecular and cellular mechanisms of miRNA biogenesis and

Table 1.1. Small RNAs Involved in gene silencing

Several species of small RNAs have been identified over the years. In mammals, there are three main classes of small RNAs: miRNAs, endo-siRNAs and piRNAs. miRNAs and endo-siRNAs have similar lengths (~21 nt) while piRNAs are generally larger and are often 24–31 nt in length. Among the small RNAs, piRNAs and endo-siRNAs are reported to regulate transposon elements, imprinted gene expression, and germline cell development, while miRNAs are more broadly involved in many different biological processes, including developmental biology and cancer pathogenesis.

Class	Size (nt)	Function	Origin	Species
microRNAs	21-25	Translation repression and mRNA destabilization	miR-coding genes, introns (Mirtrons)	*C. Elegans* *D. Melanogaster* *A. Thaliana* *O. Sativa* *Mammals*
Endo-siRNAs	21-25	Transposon silencing and mRNA degradation	Transposons, pseudogenes	*D. Melanogaster* *C.Elegans* *Mammals*
siRNAs	21-25	mRNA degradation and transposon silencing	Intergenic regions, exons and introns	*C. Elegans* *D. Melanogaster* *S. Pombe* *A. Thaliana* *O. Sativa*
piRNAs	24-31[1]	Transposon silencing, spermatogenesis	Transposons and other repeat sequences	*C. Elegans* *D. Melanogaster* *Danio rerio* *Mammals*

1. piRNAs in *C.elegans* are ~21nt in length.

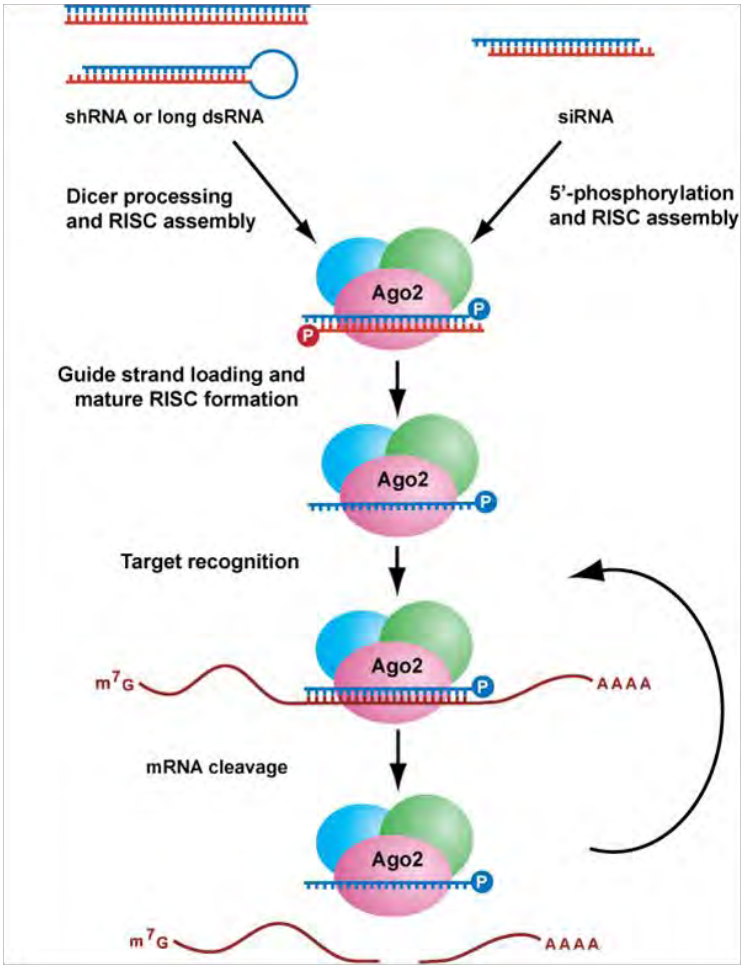
function are not fully understood. In this review, we focus on our current understanding of the structure and function of RNAi triggers and how this knowledge contributes to our understanding of miRNA function in mammalian cells.

1.2 The RNAi Triggers

A variety of RNA molecules are able to induce RNAi, including hairpin RNAs, long double-stranded RNAs, RNA viruses, transposon elements, and exogenously introduced siRNAs (Rana, 2007). Hairpin RNAs and long dsRNAs induce RNAi after processing by the enzyme Dicer, an RNase III family endoribonuclease (Figure 1.1.). The products of Dicer activity are small RNAs with a 2-nt overhang at the 3‘ end of each strand, and a monophosphate at the 5‘ end. Dicer binds to both linear dsRNAs and hairpin RNAs; thus, these molecules could be expressed by DNA vectors in target cells to induce efficient gene silencing. After cleavage by Dicer, the resulting ~21 nt RNAs are loaded into an RNA-protein complex called the RNA-induced silencing complex (RISC). Alternatively, exogenous siRNAs of the same length can be directly introduced into cells and loaded into RISCs without Dicer processing (Rana, 2007); this has become the standard experimental method to induce transient gene silencing in mammalian cells. Depending on the original source of the small RNAs, RISCs are termed miRISCs or siRISCs. Once loaded into RISCs, the two strands of the RNA duplex have distinct fates. The sense (passenger) strand that has the same

Figure 1.1. Steps in RISC function

Double-stranded (ds) or short hairpin (sh) RNAs are first bound and cleaved by Dicer into small interfering RNAs (siRNAs; ~21nt) with 2-nt overhangs and 5‘ phosphates. These siRNAs are then loaded into protein complexes termed RNA-induced silencing complex (RISCs). Ago2, a component of RISCs, binds the double-stranded siRNAs and cleaves the passenger strand, which induces its dissociation from the RISC complex and degradation. The remaining guide strand then leads the activated RISCs to find target mRNAs that contain perfectly matched complementary sequences to the guide strand. Binding of RISCs to the target mRNAs induces conformational changes and results in cleavage of the mRNA by Ago2. Cleaved mRNAs are then subject to mRNA decay or degradation, thus silencing the target gene expression.



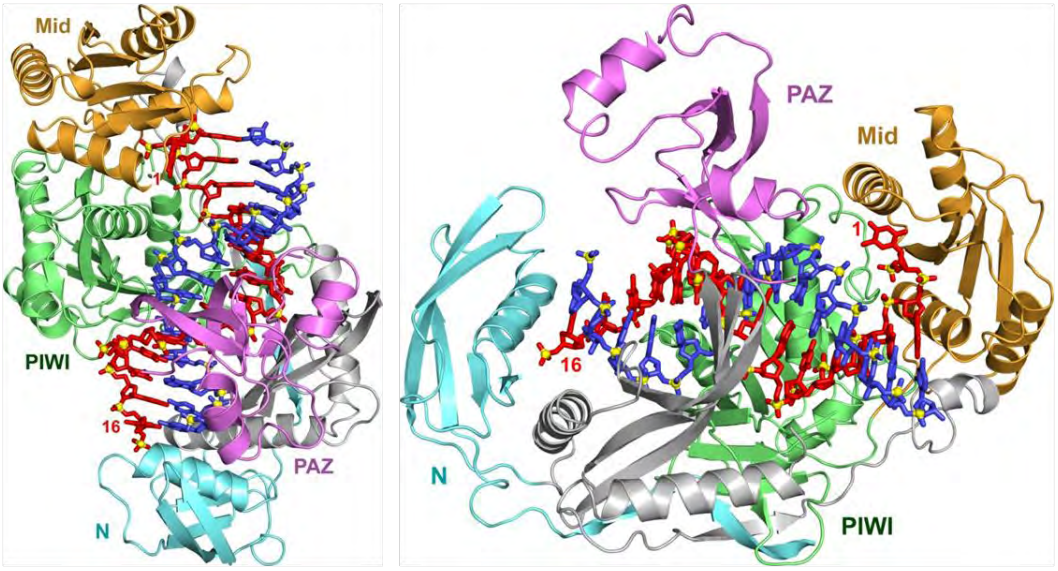
sequence as the target mRNA will be cleaved and degraded. In contrast, the antisense (guide) strand that has the complementary sequence to the target mRNA will remain in the RISC and direct recognition and cleavage of the target mRNA (Figure 1.1.). Target gene expression is silenced by cleaving the mRNA 10–11 nt upstream of the 5' end of the guide strand. This is mediated through the activity of Ago2, which is one of the main components of RISCs and contains an enzymatically competent RNase H-like domain. Ago2 lies at the heart of RNAi pathways and is the catalytic center of RISC function. After the target mRNA is cleaved, the RISCs are recycled and proceed through several rounds of cleavage events.

Not all siRNAs are loaded into RISC with the same efficiency. Several studies have uncovered some key siRNA features that considerably affect their RISC loading efficiency, and thus also affect the downstream potency of RNA interference. One important feature is the RNA structure. The ideal RNAi triggers adopt an A-form helix, which is different from the typical B-form helix of DNA molecules. This helical geometry leads to a more tightly packaged RNAi molecule with a narrower and deeper major groove, making it more stable than the B-form helix. These observations are supported by the results of experiments with mutant siRNAs that contain internal bulge structures (Chiu & Rana, 2002; Chiu & Rana, 2003) or residues with chemical modifications on functional groups (Amarzguioui et al, 2003). The bulge structures may distort the A-form helix by widening the major groove and increasing the accessibility of its functional

groups (Neenhold & Rana, 1995). Consistent with this, introducing bulge structures into the guide strand was found to completely abolish the RNAi activity of mutant siRNAs. These results, together with those using chemical modification of siRNA, have established the essential role of A-helical geometry in siRNA-mediated gene silencing (Rana, 2007). Recent crystal structures of Ago bound to a guide strand and its target RNA further highlighted the significance of the A-form helix in RISC catalysis (Figure 1.2.) (Wang et al, 2009). High-resolution crystal structures have been reported of *T. thermophilus* Ago catalytic mutant proteins bound to 5′-phosphorylated 21-nt guide DNA and complementary target RNAs of 12, 15, and 19 nt in length (Wang et al, 2009). These structural and biochemical studies provide insight into the guide-strand–mediated recognition and cleavage of target RNA by Ago, as well as the importance of divalent metal ions in catalysis (Wang et al, 2009). Ternary structures have determined that both ends of the guide strand are anchored forming one helical turn of the A-form helix with the 12-nt target RNA spanning the seed region and cleavage site. Analysis of base stacking between RNA and protein showed interesting interactions: the base at position 16 of the guide strand stacked on the aromatic ring of Tyr43 while the base at 16‘ of the target strand stacked over the Pro44 ring. Base-pair stacking is disrupted for bases 17, 18, and 19, leading to separation of guide and target strands (Figure 2). These interactions demonstrate an unexpected role of the N domain in blocking the propagation of the guide strand-target RNA duplexes beyond position 16 in the 19-nt target

Figure 1.2. Crystal Structure of *T. thermophilus* Ago (Asn478) bound with 21-nucleotide guide DNA and 19-nucleotide target RNA

Two views of the 2.8 A crystal structure of the ternary complex. The structure was generated using mutant Ago of *T. thermophilus*, which is unable to cleave the target RNA thus facilitating detailed examination of the cleavage site at position 10–11. The guide strand DNA (red) is traced for nucleotides 1–16, which are perfectly matched with its target mRNA (blue). Target RNA is traced for nucleotides 2‘–16‘. Only the 5‘ end of the guide strand is anchored in this ternary complex. The two strands retain the conformation to one turn of A-form helix (12 nt) upon binding (Chiu & Rana, 2002; Chiu & Rana, 2003), and the cleavage site of nt 10–11 stack on each other in a catalytically competent conformation. The N-domain of Ago seems to block the interaction between the guide strand and target mRNA beyond position 16, thus the 3‘ end could be released from the PAZ domain. Adapted from Wang et al. (Wang et al, 2009)



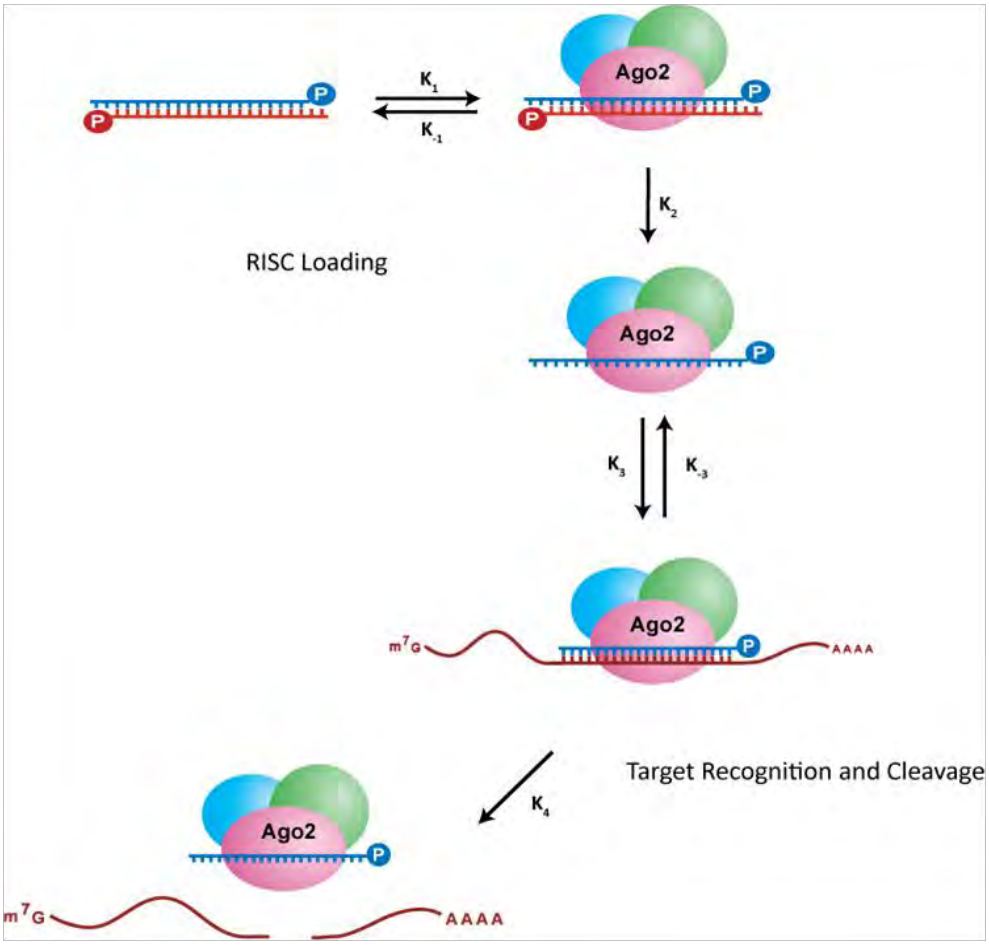
ternary complexes (Figure 1.2.). A second RNAi feature that influences efficient RISC loading and RNAi is the requirement for 5' phosphorylation of the guide strand. siRNAs generated from long dsRNAs by Dicer all contain 5' monophosphates, while exogenously introduced siRNAs often have 5' hydroxyl groups. This suggests that loading of siRNAs into functional RISCs may require 5' phosphorylation of siRNAs. Indeed, this is supported by the observation that RNAi activity can be abolished by chemical modification of the 5' end of siRNAs with amino groups and 3-carbon linkers to block phosphorylation (Chiu & Rana, 2002). This modification could also block the binding of cellular factors that recognize the 5' hydroxyl group. Interestingly, unlike the guide strand, modifications of the passenger strand, such as chemical modification and introduction of bulge structures, are well tolerated. Most passenger strand modifications will not negatively affect RNAi activity as long as the A-form helix structure of the siRNA duplex is maintained (Rana, 2007). This includes capping the 5' hydroxyl of the passenger strand to facilitate loading of the guide strand into functional RISCs.

1.3 Kinetics of the Catalytic Engine Assembly

The assembly of RISCs requires a series of kinetic processes, and can be divided into at least two catalytic steps: (1) RISC loading, and (2) target recognition, cleavage, and release. These two events each contain several further steps, such as dsRNA binding, target recognition, cleavage, product release, and RISC recycling (Figure 1.3.). For simplicity, only two checkpoints

Figure 1.3. Kinetics of RISC assembly and function

The assembly and function of RISCs can be divided into at least two catalytic steps; for simplicity only two checkpoints are considered here. The first checkpoint is RISC loading. siRNA binding by RISCs is denoted as K_1 and assembly of functional activated RISCs is denoted as K_2. K_2 can be affected by the thermodynamics of siRNAs. The second checkpoint involves target recognition and cleavage. After guide strands of siRNAs are loaded into RISCs, the protein complex is activated and led by the guide strand to target mRNAs. Target mRNAs are bound by functional RISCs, change their conformation to A-form helices and are finally cleaved by Ago2 at nt position 10–11 from the 5‘ end of the guide strand. The target mRNA recognition by RISCs is denoted as K_3 and mRNA cleavage is denoted as K_4. K_3 could be affected by several factors such as the secondary structure of target mRNAs.



steps are considered here. The overall catalytic efficiency of RISC assembly can be represented by K_{cat}, which is the turnover number or the number of reactions that occur at the catalytic site per unit of time. The K_{cat} for RISC loading is designated as K_2 while that of the second catalytic step, target recognition and cleavage, is designated as K_4. Therefore, RISCs with high catalytic potentials would have high K_2 and K_4.

Several parameters may affect the rates of K_2 and K_4 and thus result in the assembly of RISCs with different performance characteristics. For example, the thermodynamics of double-stranded siRNAs could determine which strand gets loaded into the RISCs. siRNA duplexes with unstable 5‘ ends in the guide strands will enable efficient incorporation of the guide strand into the functional RISCs (activated RISCs). During this process, the passenger strand will be cleaved by Ago2 and subjected to further destruction. Removal of the passenger strand facilitates RISC formation (Matranga et al, 2005). Therefore, siRNAs with unstable 5‘ ends in the guide strands will likely have a high K_2, which indicates that it will be more efficiently incorporated into the functional RISCs. As mentioned previously, modification of siRNAs would affect their loading efficiency into RISCs, which is another factor that could affect K_2. However, K_2 is not the only factor to consider for achieving efficient downstream target gene silencing. In reality, not all of the activated RISCs would have the same target mRNA recognition and cleavage, thus the second K_{cat}, K_4, is postulated to be crucial as well. At least two parameters could control K_4. One is the accessibility of target

mRNAs. The local environment of a target mRNA could indeed have a profound impact on the silencing efficiency of the same RISCs. Recent studies have shown that mRNA regions with strong secondary structures, such as hairpin and stem loops, are resistant to targeting by RISCs (Brown et al, 2005; Overhoff et al, 2005; Schubert et al, 2005). In this case then, high K_4 represents high accessibility of mRNAs for activated RISCs. Another factor that could affect K_4 is the structural flexibility of the RISC complex. Various studies have shown that RISCs formed in vivo (holo-RISCs) by delivery of exogenous siRNAs into the cell have lower K_{cat} (K_4) than RISCs formed in cell lysates (minimal RISCs), or recombinant RISCs (Brown et al, 2005). This could be due to binding of additional cellular factors to the RISCs, thus restricting the structural flexibility of the assembled protein complex. It should be noted that RISCs with high K_4 might not be advantageous since the high structural flexibility could increase the risk of non-specific mRNA destruction in cells.

There are two specific steps of RISC assembly that are rate-limiting (Figure 1.3.). The first is the binding and loading of siRNAs into RISCs, and the second is the target recognition process. Two mechanisms have been envisioned by which activated RISCs could recognize target mRNAs. One is a mechanism similar to that used by ribosomes to locate the translation initiation site by scanning across the target mRNA and stopping at the first suitable site. Alternatively, RISCs could recognize the target mRNAs by a diffusion-controlled —hit-and-run" mechanism. To test the scanning model, 2'-O-methyl

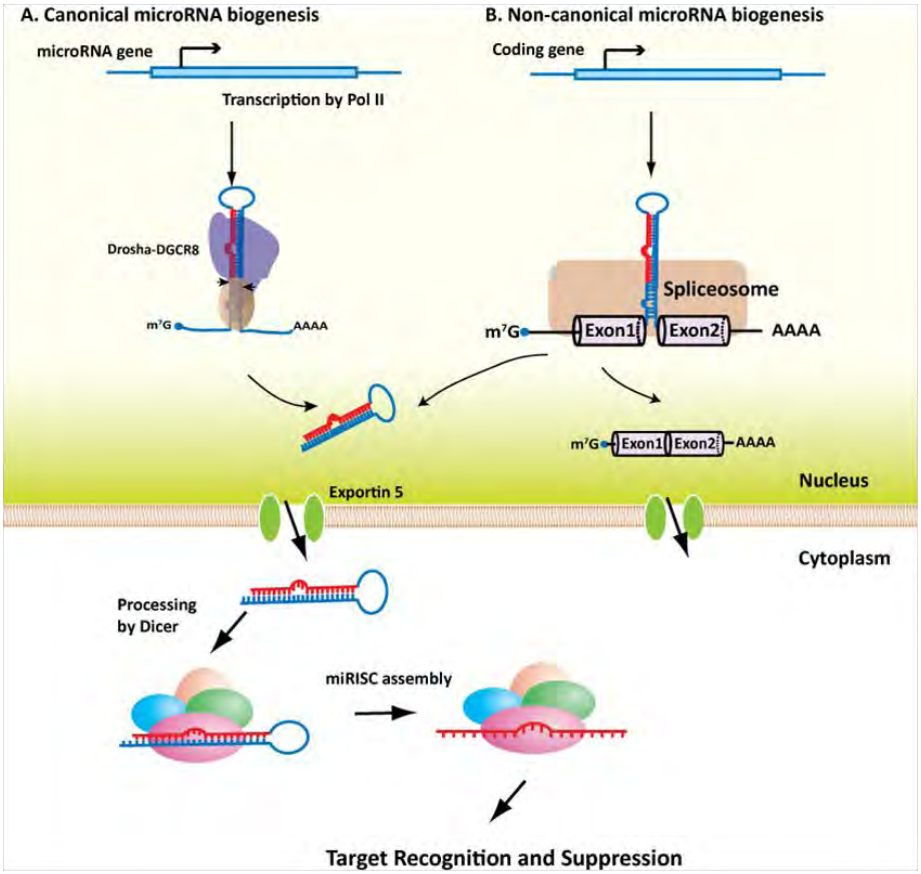
oligonucleotides were used to create blocks near the target site on the mRNAs. If the model is correct, RISCs will be arrested at these blocks due to high-affinity binding of the oligos on the mRNAs, which will prevent RISCs from further scanning. However, the 2‘-O-methyl oligonucleotides were found to enhance cleavage of target mRNAs by RISC due to the removal of nearby secondary structures and increased accessibility of the targets (Brown et al, 2005). Thus, target recognition of RISCs follows the diffusion-controlled model, where antisense strand-guided RISCs are continuously binding to different target mRNAs. Once a perfectly matched mRNA is bound, the complementary strands form an A-form helix and induce conformation changes in the RISCs, resulting in target mRNA cleavage. Interestingly, RISC is about three fold more active in the absence of translation and blocking scanning from both the 5‘ and 3‘ ends of an mRNA does not interfere with RISC function (Gu & Rossi, 2005).

1.4 The Origin of Natural Triggers

Currently, at least two pathways have been identified for miRNA biogenesis (Figure 1.4.). The canonical miRNA biogenesis pathway starts with the transcription of independent miRNA-encoding transcripts. These primary miRNA transcripts (pri-miRNAs) fold into hairpin structures and are processed in the nucleus by Drosha and its associated protein complex. Drosha is a member of the RNase III family of enzymes and, together with its cofactor DGCR8, cuts the pri-miRNA hairpins to generate ~70 nt miRNA precursors (pre-miRNAs). By

Figure 1.4. Canonical and non-canonical microRNA biogenesis pathways

Depending on the origin of miRNAs, two pathways have been proposed for miRNA biogenesis in vivo. (a) Canonical miRNA biogenesis. In this pathway, miRNA-encoding genes are first transcribed, usually through the Pol II promoter, into primary-miRNA–containing mRNAs. Hairpin structures within these mRNAs are then detected and bound by the Drosha-DGCR8 protein complex. Drosha cleaves the hairpin and generates ~70 nt long miRNA precursors, called pre-microRNAs. Pre-miRNAs are then transported from the nucleus into the cytoplasm through exportin 5, and are further processed by the Dicer complex. Processing by Dicer generates ~21 nt mature miRNAs which are then loaded by Ago2 to form functional RISCs and carry out downstream functions. (b) Non-canonical microRNA biogenesis. In this pathway, miRNAs are usually encoded in the intron regions of protein-coding genes, called mirtrons. After transcription, primary mRNAs are bound and processed by spliceosome protein complexes, which give rise to mature protein coding mRNAs and ~70-nt pre-miRNAs after debranching. Pre-miRNAs generated in this way then join the ones from the canonical pathway for transportation and Dicer processing.



contrast, the non-canonical miRNA biogenesis pathway is Drosha-independent. Instead, miRNAs generated through this pathway are usually encoded in the intron regions of protein-coding genes which are often referred to as mirtrons. Mirtron-containing primary transcripts are processed by spliceosomes to generate pre-miRNAs. Pre-miRNAs from both canonical and non-canonical biogenesis pathways are then exported into the cytoplasm by the exportin 5 complex and are further processed by Dicer to generate mature miRNAs. Finally, the miRNAs are loaded into Ago-containing RISC complexes (miRISCs) to carry out their downstream functions.

The regulation of miRNA biogenesis mainly relies on transcriptional regulation of miRNA-encoding genes. However, recent progress provides evidence that other steps in miRNA biogenesis are also tightly regulated (Fabian et al, 2010; Krol et al, 2010). In the canonical pathway, Drosha and DGCR8 can cross-regulate each other‘s expression. Binding of DGCR8 to Drosha's middle domain has a stabilizing effect, but excessive amounts of DGCR8 significantly compromise the processing activity of Drosha *in vitro*. It is likely that maintaining the correct ratio of Drosha to DGCR8 is crucial for optimal processing activity of the complex and for miRNA biogenesis. In addition to Drosha and DGCR8, Dicer is also regulated by its binding partner TRBP, as a decrease in TRBP levels results in destabilization of Dicer and defects in pre-miRNA processing. This is particularly important in certain diseases such as human carcinomas, where TRBP expression is diminished and causes impaired Dicer function. Since many

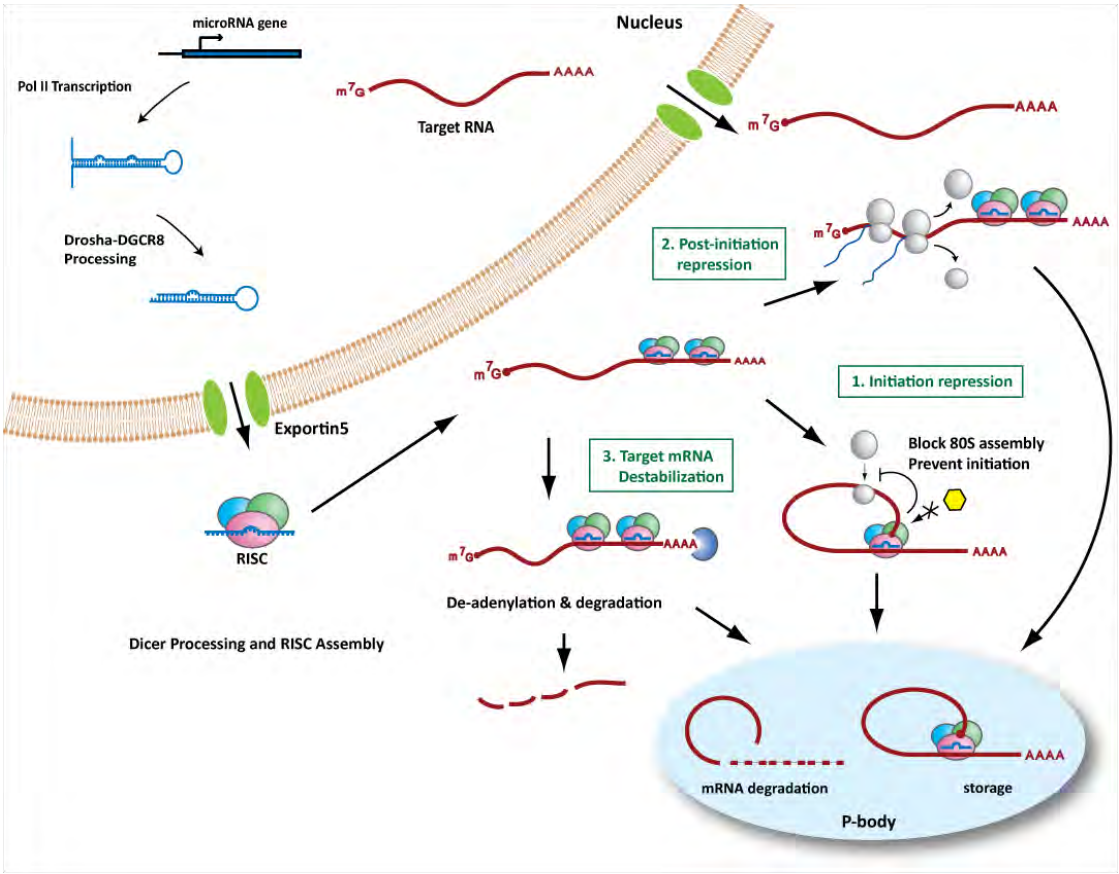
miRNAs act as potent tumor suppressors, impaired miRNA biogenesis could contribute to the progression of these carcinomas. Collectively, these findings point to a sophisticated network that tightly regulates miRNA biogenesis.

1.5 MicroRNA-mediated Post-transcriptional Gene Regulation

Once formed, pre-miRNAs are exported into the cytoplasm where they are further processed by Dicer and loaded into functional miRISCs (Figure 1.5.). The unique features of miRNAs results in miRISCs having different functions than siRISCs. While siRISCs induce gene silencing by cutting target mRNAs with perfectly complementary sequences to the guide strand, miRNAs induce gene silencing without cleaving the target mRNA, although cleavage activity is retained when a perfectly matched target is present. The seed region of miRNAs, 2–7 nt at the 5‘ end of mature miRNAs, plays a key role in determining which target mRNAs are regulated by a given miRNA (Lewis et al, 2005). The first translation repression mechanism by an miRNA was shown when miRNA Lin-4 in *C. elegans* inhibited Lin-14 expression without causing a reduction in Lin-14 mRNA levels (Lee et al, 1993; Olsen & Ambros, 1999; Wightman et al, 1993). Based on recent developments in understanding miRNA biology and mechanisms, at least three main models can be proposed by which miRNAs could modulate gene expression post-transcriptionally: (1) inhibition of translation initiation, (2) post-initiation inhibition of translation, and (3) mRNA destabilization (Figure 5).

Figure 1.5 microRNA function

After loading with Ago proteins to form functional RISCs, miRNA-guided RISCs bind to the target mRNAs and inhibit target gene expression. Currently, there are at least three mechanisms that have been linked to miRNA-mediated gene silencing. (1) Repression of translation initiation. In this case, miRISCs inhibit initiation of translation by affecting the eIF4F-cap recognition, 40S small ribosomal subunit recruitment, and/or by inhibiting incorporation of the 60S subunit and formation of the 80S ribosomal complex. Some of the target mRNAs bound by miRISC is transported into P-bodies for storage and may re-enter the translation phase when induced. (2) Post-initiation translational repression. miRISCs could interfere after translation has been initiated by inhibiting elongation of ribosomes, causing ribosome drop-off from mRNAs, and/or by facilitating degradation of newly-synthesized nascent peptides. (3) Destabilization of target mRNAs. miRISCs could cause destabilization of target mRNAs by directly interacting with CCR4-containing deadenylation complexes and facilitating the deadenylation of poly A tails of target mRNAs. Following deadenylation, the 5‘ end capping structures of target mRNAs are also removed by the DCP1-DCP2 complex.



Inhibition of translation initiation: MicroRNA-mediated translation repression was observed in HeLa cells in which reporter expression was regulated by let-7 miRNA (Pillai et al, 2005), and no decrease of reporter mRNAs was detected. In addition, reporter mRNAs containing let-7 target sites shifted to a lighter fraction of polysomal gradients, suggesting that repression could be modulating translation initiation (Pillai et al, 2005). There are some observations suggesting that this inhibition of translation initiation could be cap-dependent as mRNAs with non-traditional cap structures (ApppG) were less repressed by Cxcr4 miRNA mimics in HeLa cells (Krol et al, 2010). This was further supported by in vitro experiments using cell-free extracts (Krol et al, 2010). miR-2–mediated repression was shown to be linked with inhibition of 40S ribosomal subunit recruitment and formation of 80S initiation complexes in fly embryo extracts. Additional evidence came from experiments where target mRNA with modified 5‘ caps exhibited increased repression by miRNAs. Similarly, supplementing the protein extracts with eIF4F complexes, which directly recognize cap structures of mRNAs, also increased let-7–mediated translational repression of reporter mRNAs. Finally, there is additional evidence demonstrating that joining of 60S ribosomal subunits could also be inhibited by miRNAs (Fabian et al, 2010). Together, these results show that miRNA-mediated repression of target mRNA is cap-dependent and results from multiple inhibitory effects on translation initiation.

Post-initiation Inhibition: Several studies provide evidence that inhibition of target gene expression by miRNAs can occur at post-initiation steps (Fabian et al,

2010). Despite the observation that certain miRNAs and Ago proteins can be detected in polysomal fractions, IRES-containing target mRNAs have been reported to be repressed by miRNAs as well (Lytle et al, 2007; Petersen et al, 2006). Some IRES-bearing mRNAs even showed cap-independent translation while still being efficiently repressed by miRNAs (Petersen et al, 2006). One model proposed for post-initiation inhibition is ribosomal run-off, in which ribosomes fall off the mRNA prematurely. Although no direct evidence exists, this model is supported by observations in vitro that inhibition of translation initiation causes a more rapid loss of ribosomes on mRNAs targeted by miRNAs (Petersen et al, 2006). Premature termination is the simplest explanation for such observations.

mRNA destabilization and decay: mRNA degradation by miRNA was reported in *C. elegans* where partial base pairing of let-7 miRNA resulted in degradation of its lin-41 target mRNA (Bagga et al, 2005). This report raised the possibility that mRNAs containing partial miRNA complementary sites can be targeted for degradation in vivo. Destabilization of target mRNA by miRISCs in mammalian cells has recently been proposed as the main mechanism of miRNA gene regulation in mammalian cells (Guo et al, 2010). This destabilization is likely due to deadenylation of target mRNAs.

How do miRNAs cause deadenylation of target mRNAs? Recent studies have revealed the molecular mechanism of miRNA-mediated mRNA

deadenylation, in which one key protein, GW182, is centrally involved (Fabian et al, 2010; Krol et al, 2010). GW182 directly interacts with all members of the Ago protein family and is localized within P-bodies in the cytoplasm of mammalian cells (Fabian et al, 2010). Another P-body protein, RCK/p54, a DEAD box helicase, has been shown to interact with the argonaute proteins, Ago1 and Ago2, and modulate miRNA function (Chu & Rana, 2006). RCK/p54 facilitates formation of P-bodies and is a general repressor of translation, suggesting that miRNAs are transferred to P-bodies for further decay or storage (Chu & Rana, 2006). GW182 binds Ago proteins through GW repeats, and tethering of GW182 to the target mRNA promotes mRNA deadenylation (Behm-Ansmant et al, 2006; Eulalio et al, 2008) through GW182-dependent recruitment of the CCR4-containing deadenylation complex (Behm-Ansmant et al, 2006). In addition, GW182 also interacts with poly(A) binding proteins (PABP) through its C terminal domain (Fabian et al, 2009). PABP has previously been reported to be involved in translation initiation by interacting with eIF4G; thus, interactions between GW182 and PABP may interfere with this process and have multiple effects on target gene expression. It is worth noting that mRNA de-capping complexes such as DCP1-DCP2 may also be involved in miRNA-mediated gene silencing, as knockdown of DCP-1 and DCP-2 stabilizes deadenylated mRNAs and thus compromises miRNA-mediated inhibition of expression (Behm-Ansmant et al, 2006; Fabian et al, 2010).

1.6 Therapeutic applications

Catalytic silencing of specific genes by RNA provided the rationale for RNAi-based therapeutic agents because siRNAs could be designed to treat diseases by lowering concentrations of disease-causing gene products. Similarly, disease-related miRNA dysregulation can be treated either by expressing miRNA mimics to enhance miRNA levels or by inhibiting high levels of disease-related miRNAs in cells. Development of such RNA-based therapies requires chemically stabilized RNA and vehicles for targeted delivery *in vivo*. Recent advances in understanding the rules for chemically modifying siRNA and miRNA sequences without compromising their gene-silencing efficiency have allowed the design and synthesis of therapeutically effective RNA molecules that can silence target genes *in vivo* (Burnett et al, 2011). The second remaining challenge to deliver RNA-based drugs to diseased organs is being addressed by rapid developments in bioengineering and nanotechnologies to design RNA cargo vehicles that can efficiently deliver and release RNA compounds at their target sites (Burnett et al, 2011). Based on this rapid progress in understanding RNA structure and function in gene silencing and their applications in disease models, it is likely that RNA-based therapeutics will become a reality in the very near future. It is remarkable to witness that in the short period of time since the discovery of RNAi, a myriad of biotechnology and drug discovery companies using RNAi have been formed, and a number of RNA therapeutics are being tested in clinical trials.

1.7 Future Perspectives

Given the fundamental roles of miRNAs in regulating a variety of processes, our current understanding of the biogenesis, regulation, and function of miRNAs will no doubt expand considerably in the coming years. One area of particular interest is miRNA editing and modification. Several emerging lines of evidence suggest that modifications on miRNA termini could have a broad impact on their stability, downstream processing, and protein recruitment. In addition, variations have been observed in mature miRNA sequences from the same pre-miRNA, and addition of nucleotides to the miRNA 5‘ end could have dramatic effects on its function since the 5‘-end seed region determines the target mRNA population. Another potentially interesting area is the emerging role of long noncoding RNAs (lncRNAs) and possible crosstalk between lncRNAs and miRNAs. The lncRNAs could be the natural sponges for miRNAs, and the available miRISCs may be regulated by expression and binding of their corresponding lncRNAs. Additionally, different Ago proteins may regulate each other‘s function by competing for the available miRNAs.

Studying these small noncoding RNAs and their potential relationship with protein-coding genes or lncRNAs should shed light on the complexity of gene regulation and lead to the development of new technologies and therapeutics.

CHAPTER II: SMALL RNA-MEDIATED REGULATION OF IPSC GENERATION

2.1. Abstract

Somatic cells can be reprogrammed to an ES-like state to create induced pluripotent stem cells (iPSCs) by ectopic expression of four transcription factors, Oct4, Sox2, Klf4 and cMyc. Here we show that cellular microRNAs regulate iPSC generation. Knock-down of key microRNA pathway proteins resulted in significant decreases in reprogramming efficiency. Three microRNA clusters, miR-17~92, 106b~25 and 106a~363, were shown to be highly induced during early reprogramming stages. Several microRNAs, including miR-93 and 106b, which have very similar seed regions, greatly enhanced iPSC induction and modulated mesenchymal-to-epithelial transition step in the initiation stage of reprogramming, and inhibiting these microRNAs significantly decreased reprogramming efficiency. Moreover, miR-iPSC clones reached the fully reprogrammed state. Further analysis revealed that Tgfbr2 and p21 are directly targeted by these microRNAs and that siRNA knock-down of both genes indeed enhanced iPSC induction. Here, for the first time, we demonstrate that miR-93 and its family members directly target TGF-β receptor II to enhance iPSC reprogramming. Overall, we demonstrate that microRNAs function in the reprogramming process and that iPSC induction efficiency can be greatly enhanced by modulating microRNA levels in cells.

2.2. Introduction

Induced pluripotent stem cells (iPSCs), which exhibit properties similar to embryonic stem (ES) cells, were originally generated by ectopic expression of four transcription factors, Oct4, Sox2, Klf4, and cMyc, in mouse somatic cells (Takahashi & Yamanaka, 2006). In human and mouse somatic cells, besides these factors (Lowry et al, 2008; Park et al, 2008a; Takahashi et al, 2007), iPSCs can be generated with an alternative set of four factors, namely, Oct4, Nanog, Lin28, and Sox2 (Yu et al, 2007). Although cell types from several different tissues are confirmed to be reprogrammable (Aoi et al, 2008; Eminli et al, 2008; Giorgetti et al, 2009; Hanna et al, 2008; Meissner et al, 2007), a major bottleneck in iPSC derivation and therapeutic use is low reprogramming efficiency, typically from 0.01% to 0.2% (Aoi et al, 2008; Meissner et al, 2007; Nakagawa et al, 2008; Takahashi & Yamanaka, 2006). Although tremendous effort has been focused on screening for small molecules to enhance reprogramming efficiency and on developing new methods for iPSC derivation (Ichida et al, 2009; Li et al, 2009b; Lyssiotis et al, 2009; Maherali & Hochedlinger, 2009; Shi et al, 2008a; Shi et al, 2008b), mechanisms underlying reprogramming of primary fibroblasts to an ES cell-like state are still largely unknown.

Several elegant approaches have been employed to improve reprogramming efficiency. Small molecule-based methods have been developed based on observation that treatment of cells with DNA methyltransferase 1

(Dnmt1) inhibitors accelerates reprogramming (Mikkelsen et al, 2008). TGFβ inhibition also enables more efficient iPSC induction, as does omission of Sox2 and cMyc (Ichida et al, 2009; Maherali & Hochedlinger, 2009). In addition, array analysis shows that partially reprogrammed iPS cells can be created and then pushed to become fully reprogrammed following treatment with factors such as methyl transferase inhibitors (Mikkelsen et al, 2008). Genome-wide analysis of promoter binding and induction of gene expression by the four reprogramming factors demonstrates that they bind to similar targets in iPS and mES cells and likely regulate similar sets of genes, and also that targeting of reprogramming factors is altered in partial iPS cells (Sridharan et al, 2009). More recently, several groups showed that p53-mediated tumor suppressor pathways may antagonize iPSC induction (Banito et al, 2009; Hong et al, 2009; Kawamura et al, 2009; Li et al, 2009a; Utikal et al, 2009). Both p53 and its downstream effector p21 are induced during reprogramming, and minimizing expression of both enhances iPSC colony formation. Since these proteins are up-regulated in most cells expressing the four reprogramming factors, and cMyc reportedly blocks p21 expression (Gartel et al, 2001; Seoane et al, 2002), it is unclear how ectopic expression of these four factors overcomes the cellular responses to oncogenes/transgenes overexpression and why only a very small population of cells becomes fully reprogrammed.

microRNAs are 18-24 nucleotide single-stranded RNAs associated with a protein complex called the RNA-induced silencing complex (RISC). Small RNAs

are usually generated from noncoding regions of gene transcripts and function to suppress gene expression by translational repression (Ambros, 2004; Bartel, 2004; Kim et al, 2009a; Rana, 2007). In recent years, microRNAs have been found to function in many important processes, such as expression of self-renewal genes in human ES cells (Xu et al, 2009), cell cycle control of ES cells (Wang et al, 2008), alternative splicing (Makeyev et al, 2007) and heart development (Latronico & Condorelli, 2009). Furthermore, it was recently reported that ES cell-specific microRNAs enhanced mouse iPSC derivation and replaced the function of cMyc during reprogramming (Judson et al, 2009) and hES-specific miR-302 could alleviate the senescence response due to four factor expression in human fibroblast (Banito et al, 2009). However, since these microRNAs are not highly expressed until very late stages of reprogramming, whether microRNAs mediating regulation of gene expression play an important role in iPSC induction remains unknown.

Here we show that microRNAs function directly in iPSC induction and that interference with the microRNA biogenesis machinery significantly decreases reprogramming efficiency. We also identified three clusters of microRNAs, miR-17~92, miR-106b~25 and miR-106a~363, which are highly induced during early stages of reprogramming. Functional analysis demonstrated that introducing these microRNAs into MEFs enhanced Oct4-GFP+ iPSC colony formation. We also found that Tgfbr2 and p21, both of which inhibit reprogramming, are directly targeted by these microRNAs and that blocking their activity significantly

decreased reprogramming efficiency. Overall, we propose that miR-93 and 106b are key regulators of reprogramming activity.

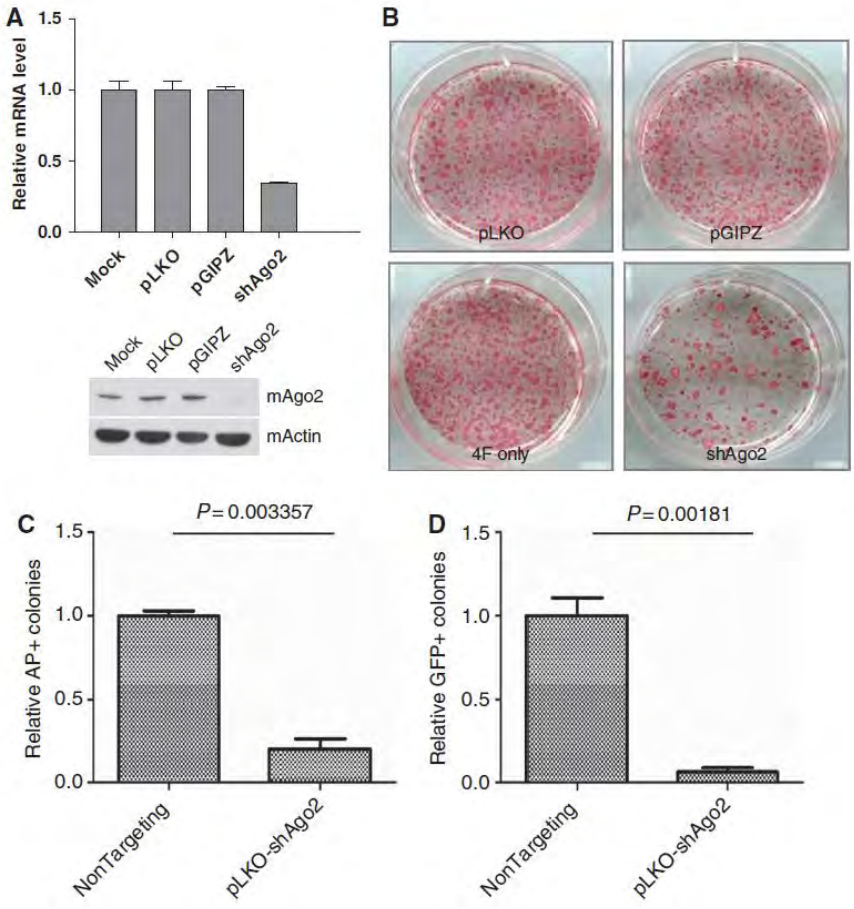
2.3. Results

2.3.1. Post-transcriptional regulation functions in reprogramming of MEFs to iPS cells

To investigate the role of post-transcriptional gene regulation in iPSC induction, we used lentiviral shRNA vectors targeting mouse Ago2 as well as Dicer and Drosha to stably knock-down these factors in primary Oct4-GFP MEFs. Knock-down efficiency of shRNA constructs was verified both by Western analysis and RT-qPCR (Figure 2.1.A, Figure 2.S1.A,B). For each shRNA, we routinely observed ~70%-80% mRNA level knock-down, as well as significant decreases in protein levels. We then transduced MEFs with each of these shRNAs separately along with viruses expressing the four OSKM (Oct4, Sox2, Klf4 and cMyc) factors at a volume ratio of 1:1:1:1:1. After 14 days, colonies were fixed and stained for alkaline phosphatase (AP) activity, a widely used ES cell marker. We found that knock-down of either Dicer, Drosha or Ago2 resulted in a dramatic decrease in the number of AP+ colonies compared with pLKO and pGIPZ controls (Figure 2.1.B, Figure 2.S1.C). We also observed similar results by using

Figure 2.1. The RNAi machinery functions in mouse iPSC induction.

(A) Knock-down of mouse RNAi machinery gene Ago2 by shRNAs. MEFs were transduced with lentiviral shRNAs plus 4ug/ul polybrene, and total RNAs or proteins were harvested at day 3 post-transduction. mRNA and protein levels of targeted genes were analyzed by RT-qPCR and Western blotting, respectively. pLKO is the empty vector control for the shRNA lentiviral vectors. pGIPZ is a lentiviral vector expressing a non-targeting shRNA. (B) Knock-down of Ago2 dramatically decreases iPS induction by 4F. Primary MEFs were transduced with the four reprogramming factors (OSKM (4F)) plus shRNA Ago2. Colonies were stained at day14 post transduction for alkaline phosphatase, which is a marker for mES/iPS cells. pLKO and pGIPZ vectors served as negative controls. (C) Knock-down of Ago2 decreases iPS induction by OSK. Colonies were stained and quantified for AP at day21 post transduction. Error bar represent standard deviation from duplicate wells. (D) GFP+ colony quantification of iPSC with shAgo2. GFP+ colonies were quantified at day21 post transduction. Error bar represent standard deviation from duplicate wells.



OSK (3F) transduction. Both GFP+ and AP+ colony quantification verified that knocking down Ago2 dramatically decrease reprogramming efficiency while proliferation of transduced fibroblasts were not affected (Figure 2.1.C,D, Suppl. Figure 2.S2.).

Despite the decrease in reprogramming efficiency upon Ago2 knockdown, we observed some GFP+ colonies in shAgo2 infected MEFs and further characterization determined that these colonies were positive for shRNA integration and shRNAs were actively expressed (Figure 2.S3.A, B). These cells also expressed all the tested ES-specific markers and had turned on the endogenous Oct4 locus as well as low expression of p21 and Tgfbr2 (Figure 2.S3.C, D, E). However, they seemed to have compromised differentiation tendency as they were not as responsive to retinoid acid treatment as mouse ES cells (Figure 2.S3.F). Understanding the detailed mechanism of GFP+ colony formation in shAgo2 infected MEFs needs further investigation. Taken together, these data strongly suggest that post-transcriptional regulation, particularly that mediated by microRNAs, functions in the reprogramming process.

2.3.2. miR-17, 25, 106a and 302b clusters are induced during the early stage of reprogramming

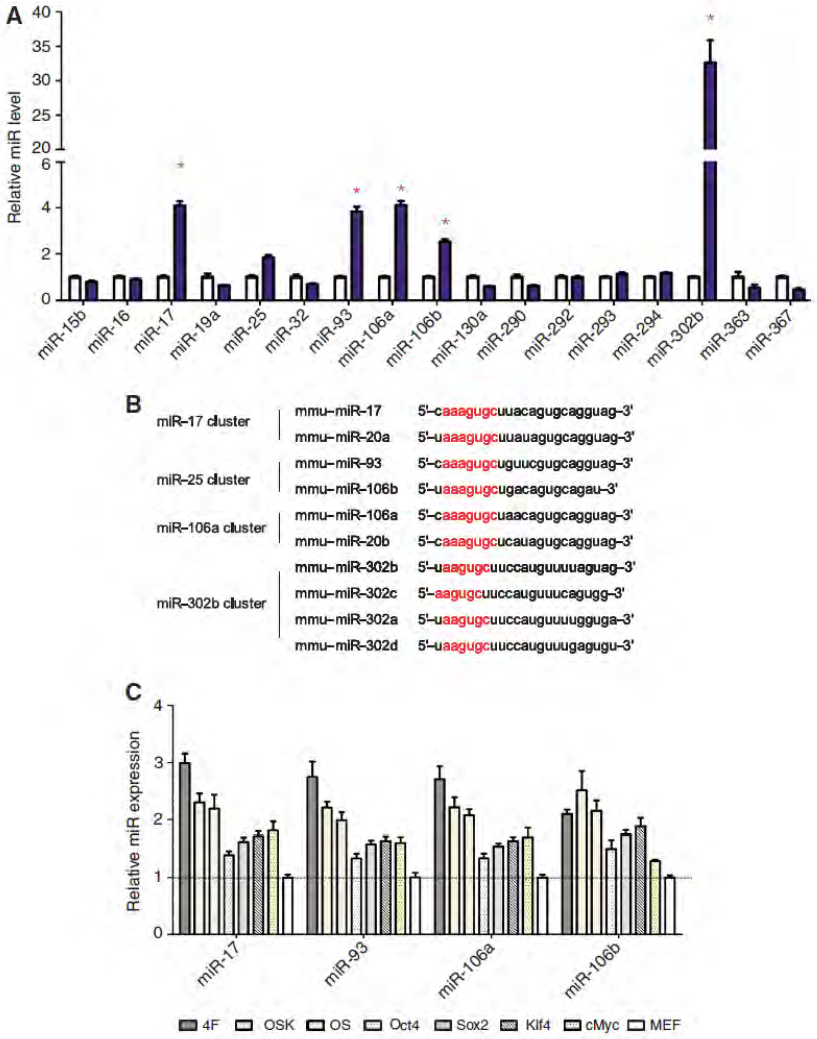
Expression of the four reprogramming factors induces numerous changes in gene expression during iPS induction (Mikkelsen et al, 2008; Sridharan et al,

2009). We hypothesized that some ES cell-specific microRNAs might be induced by these factors to facilitate reprogramming. Based on previously published results (Houbaviy et al, 2003; Landgraf et al, 2007), we analyzed nine microRNA clusters highly expressed in mouse ES cells. Two representative microRNAs from each cluster were evaluated using a miR qPCR-based method to quantify expression changes at different reprogramming stages—namely days 0, 4, 8 and 12—following transduction of the OSKM factors. Many ES cell-specific microRNAs, such as the miR-290 and miR-293 clusters, were not induced until day 8 (Figure 2.S4.), at which stage GFP+ colonies were already detectable. Interestingly, we found that several other clusters, including miR-17~92, 106b~25, 106a~363 and 302b~367, were expressed to varying extents by day 4 of induction (Figure 2.2.A). Among these four clusters, the level of miR-302b~367 in MEFs was the lowest (data not shown). It is noteworthy that of the three clusters highly induced at reprogramming day 4, many shared very similar seed regions (Figure 2.2.B). In general, seed region of a microRNA decides the target specificity, however, recent reports suggest other mechanisms could also play roles in microRNA targeting (Lal et al, 2009; Tay et al, 2008; Wang et al, 2010). Together, our findings suggest that these microRNAs function in reprogramming and could target similar sets of genes.

We next asked which of the four reprogramming factor(s) induced these

Figure 2.2. miR-17, 25, 106a and 302b clusters are induced during early stages of reprogramming.

(A) Induction of 10 microRNA clusters in the early stages after transduction with the four reprogramming factors. miR RT-qPCR was used to quantify expression levels of representative microRNAs from clusters highly expressed in ES cells. Total RNAs from day 0 MEFs and from MEFs transduced with reprogramming factors at day 4 post-infection were analyzed. Blue bars: day 4 MEFs; white bars: day 0 MEFs. Asterisks indicate induced microRNAs. (B) Seed region comparison of different miR clusters induced at day 4 post-reprogramming factor transduction. Red indicates similar seed regions. (C) Representative microRNAs can be induced with different combinations of reprogramming factors. microRNA expression was quantified at 4 days post transduction.



microRNAs by transducing MEFs with different combinations of OSKM factors at the same dose and undertaking miR qPCR analysis at day 4 post infection (Figure 2.2.C). This analysis confirmed that cMyc alone could induce miR-17~92, miR-106b~25, and miR-106a~363 cluster expression, as reported previously (Mendell, 2008). However, in each case, a combination of all four reprogramming factors induced the most abundant expression of microRNA clusters, and that robust expression was correlated with the highest reprogramming efficiency (Figure 2.2.C).

2.3.3. miR-93 and miR-106b enhance iPSC induction and mesenchymal-to-epithelial transition (MET) step of reprogramming

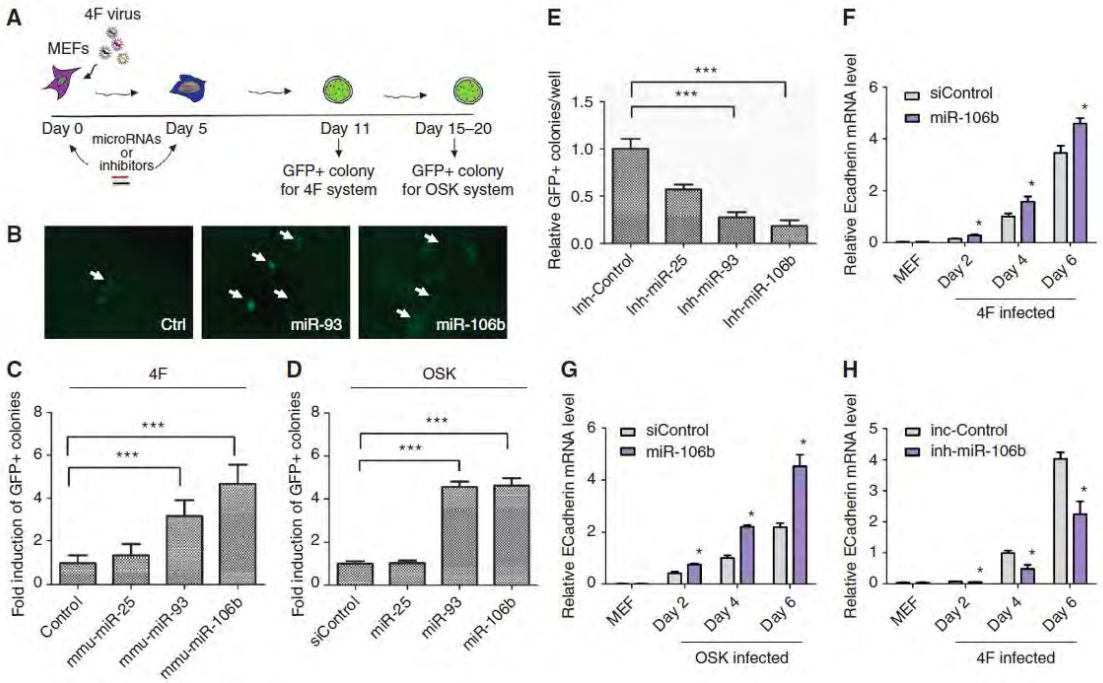
Since the four identified microRNA clusters contain several microRNAs with similar seed regions, we chose the miR-106b~25 cluster for further analysis because it contains only three microRNAs: miR-25, miR-93 and miR-106b. miR-93 and 106b have the identical seed regions, and both were highly induced by the four reprogramming factors (Figure 2.2.A). Thus we reasoned that we might observe more efficient iPSC induction if we ectopically expressed these microRNAs during reprogramming. Besides, microRNAs mimics could be transfected into MEFs with high efficiency and have a half-life of 4 days (Figure 2.S7., Figure 2.S9.B). To test this hypothesis we directly transfected microRNA mimics into MEFs harboring Oct4-GFP at days 0 and 5 with vectors expressing either all four factors (4F, OSKM) or only Oct4, Sox2, and Klf4 (OSK) and

assayed reprogramming based on GFP expression (Figure 2.3.A). GFP+ colonies were counted on day 11 to evaluate reprogramming efficiency (Figure 2.3.B). Transfection of miR-93 and 106b mimics promoted a 4~6 fold increase in the number of GFP+ colonies both in 4F and OSK transduction (Figure 2.3.C,D; Figure S22), confirming that these microRNAs, which are induced during iPSC induction, facilitate MEF reprogramming. Dose/response analysis showed that enhanced reprogramming efficiency occurred at as low as the 5-15nM range of miRs (Figure 2.S5.). To confirm that the enhancement by these microRNAs was from induction of bonafide iPS colonies, we further analyzed the expression of another marker Nanog in miR-106b transfected cells. In both 4F and OSK infected samples, miR-106b transfection consistently increased the relative Nanog expression (Figure 2.S6.A,B). Immunostaining and followed by Nanog+ colonies quantification further proved that almost every Oct4-GFP+ colony is also Nanog+ at that stage (Figure 2.S6.C) and miR-106b can enhance formation of both colonies (Figure 2.S6.D,E). Alkaline phosphatase staining showed no obvious increase in the number of AP+ colonies in miR mimic transfections (Figure 2.S8.A), suggesting that miR-93 and 106b facilitate maturation of iPS colonies. This idea was supported by our observation of the OSK system, in which many GFP+ colonies were apparent at day 15 post-OSK transduction in miR mimic-transfected cells, while control wells did not exhibit any mature iPS colonies at this stage (data not shown). To confirm that these microRNAs play an important role in iPSC induction, we used miR inhibitors (Hutvagner et al, 2004;

Figure 2.3. miR-93 and 106b greatly enhance iPS induction.

(A) Reprogramming assay timeline. MicroRNA mimics or inhibitors (or inhibitors) were transfected at a final concentration of 50nM on day 0 and day 5 of reprogramming. GFP+ colonies were quantified at day 11 for 4F induction and days 15-20 for OSK induction. (B) Representative images of GFP+ colonies from reprogrammed Oct4-GFP MEFs transfected with microRNA mimics. Arrows indicate GFP+ colonies. (C) miR-93 and 106b mimics enhance iPS induction with 4F induction. Oct4-GFP MEFs were transfected with 50nM of the indicated microRNAs at days 0 and 5 of reprogramming. GFP+ colonies were quantified at day 11 post-transduction. Fold-induction and error bars were calculated from three independent experiments using triplicate wells. *** $p<0.0001$. (D) The enhancing effect of miR-93 and 106b is observed using the OSK system. microRNA mimics were transfected as in 4F experiments. GFP+ colonies were quantified on days 15-20. Error bars represent standard deviation from three independent experiments in triplicate wells. *** $p<0.0001$. (E) miR-93 and 106b inhibitors dramatically decrease reprogramming efficiency. microRNA inhibitors were transfected at a final concentration of 50 nM. The experimental timeline was the same as in miR mimic transfections. Error bars represent standard deviation from three independent experiments in triplicate wells. *** $p<0.0001$. (F) miR-106 promotes the MET transition during 4F mediated reprogramming. miR-106b mimic was transfected into MEFs and cells were harvested at different time point to analyze E-Cadherin expression. Fold induction of ECad was normalized to day4 samples after 4F infection. Error bar represents standard deviation of three independent experiments. * $p<0.001$. (G) miR-106b promotes the MET transition in OSK infected cells. The experimental procedures were the same as in (f). Fold induction of ECad was normalized to day4 samples after OSK infection. Error bar represents standard deviation of three independent experiments. * $p<0.001$. (H) Inhibition of miR-106b decreases induction of MET process. The experimental procedures were the same as in (f), except anti-miR oligos were transfected instead of miR mimics. Fold induction of ECad was normalized to day4 samples after 4F infection. Error bar represents standard deviation of three independent experiments. * $p<0.001$.

Figure 2.3. miR-93 and 106b greatly enhance iPS induction.



Meister et al, 2004; Vermeulen et al, 2007) to knock down targeted microRNAs during the reprogramming process. All of the miR inhibitors could efficiently decrease target miR expression and their transfection did not affect proliferation (Figure 2.S9.A,C). Consistent with miR mimic experiments, miR-93 and 106b knock-down promoted a dramatic decrease in the number of GFP+ colonies (Figure 2.3.E). It is also noteworthy that although the miR-25 mimic did not enhance MEF iPS induction, knocking down this microRNA decreased reprogramming efficiency by ~40% (Figure 2.3.E). These results suggest that miR-25 could also function during reprogramming.

Recent reports have identified that during the initial stage of reprogramming, a mesenchymal-to-epithelial transition (MET) is required (Li et al, 2010; Samavarchi-Tehrani et al, 2010). E-Cadherin is one of the most important genes for MET process and we used it as the marker to determine whether miR-106b could facilitate this step of iPSC generation. We detected a significant increase of E-Cadherin expression in both 4F and OSK infected samples (Figure 2.3.F,G). In addition, knocking down of miR-106b also dramatically decreased the induction of E-Cadherin expression (Figure 2.3.H). Overall, these data indicate that miR-93 and 106 promote reprogramming of MEFs to iPSCs and modulate MET transition in the initiation step of reprogramming.

2.3.4. MicroRNA-derived clones are fully pluripotent

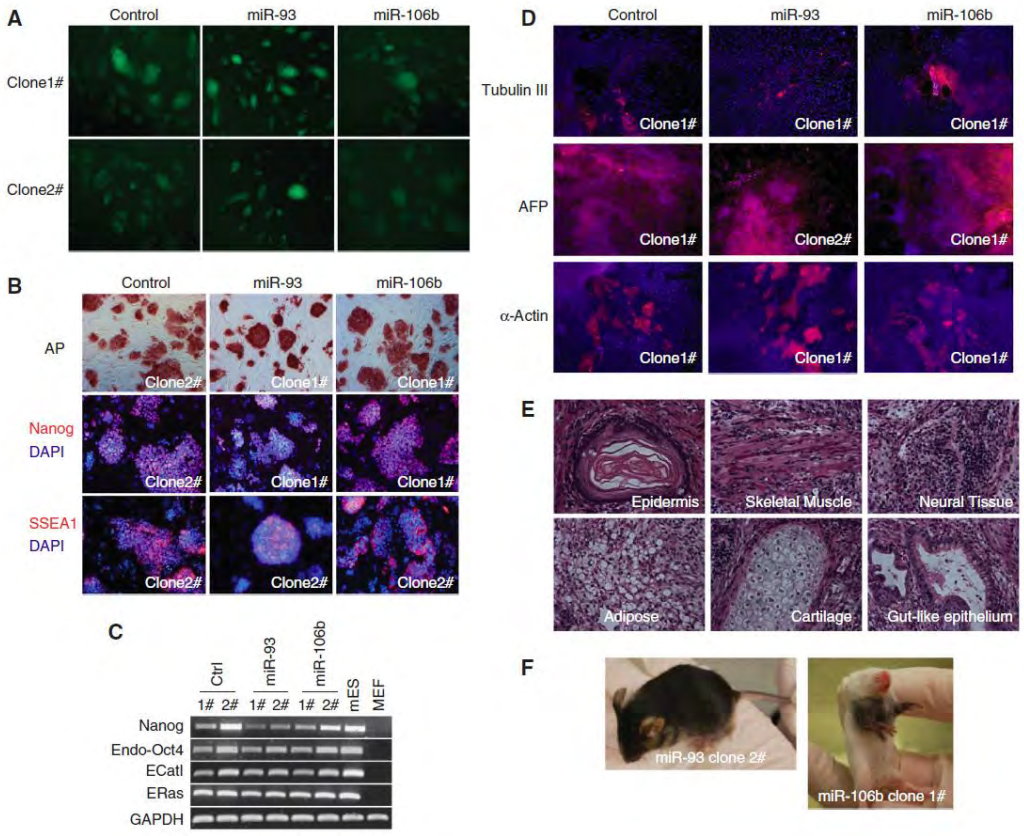
Next we asked whether induced cells reach a fully pluripotent state. To answer this question, several iPS clones for each microRNA as well as miR controls were derived and analyzed for expression of pluripotency markers. All clones were GFP+ indicative of reactivated Oct4 expression (Figure 2.4.A). Immunostaining confirmed that Nanog and SSEA1 were also activated in all clones (Figure 2.4.B). RT-qPCR for other mES markers such as Eras, ECat I and endogeneous Oct4 showed similar results (Figure 2.4.C). Whole genome mRNA expression profiling also indicated that derived clones exhibited a gene expression pattern more similar to mouse ES cells than MEFs (Figure 2.S10.A). Promoter methylation of endogenous Nanog loci was also analyzed, and all tested clones showed de-methylated promoters, as is observed in mouse ES cells (Blelloch et al, 2006) (Figure 2.10.B).

To investigate whether derived clones exhibit the full differentiation capacity of mES cells, we evaluated embryoid body (EB) formation. All derived clones showed efficient EB formation, and EBs showed positive staining for lineage markers such as such as β-tubulin III (ectoderm), AFP (endoderm) and α-actinin (mesoderm) (Figure 2.4.D). Beating EBs were also derived from these cells (Suppl. Video 1), indicating that functional cardiomyocytes can be derived from these miR-iPSC clones. When these miR-iPSCs were injected into athymus nude mice, teratomas were readily derived in 3-4 weeks (Figure 2.4.E). Finally, as a more stringent test, we injected miR-derived iPSC clones into albino/black B6 blastocysts and generated chimera mice (Figure 2.4.F).

Figure 2.4. Characterization of iPS clones derived from miR mimic experiments.

(A) Derived clones activate endogenous Oct4-GFP expression. Colonies were picked starting at day 12 post-OSKM transduction with microRNA mimics and maintained on irradiated MEF feeder plates. Green fluorescence is GFP signal from the endogenous *pou5f1* locus. (B) Clones shown in (a) are positive for alkaline phosphatase staining and immunostaining of ES-specific markers based on Nanog and SSEA1 staining. Hoechst 33342 was used for nuclear staining. (C) RT-PCR of endogenous ES markers. Total RNAs were isolated from iPS cell lines at day 3 post-passage. ES cell-specific markers such as Eras, ECat I, Nanog, and endogenous Oct4 expression were analyzed by RT-PCR. (D) Cells from all three germ layers can be obtained in embryoid body (EB) assays using derived iPS clones. iPS cells were cultured for EB formation at ~4000 cells/20ul drop for 3 days, and EBs were then reseeded onto gelatin coated plates for further culture until day 12-14, when beating cardiomyocytes were observed (Supplementary Video 1). Cells were immunostained with different lineage markers. β-tubulin III: ectoderm marker; AFP: endoderm marker; α-Actinin: mesoderm marker. (E) Teratomas form from injected iPS cells. 1.5 million cells were injected into each mouse, and tumors were harvested 3~4 weeks after injection for paraffin embedding and H&E staining. Structures representing different lineages are labeled. Representative pictures are from miR-106b clone 1#. (F) Derived clones can be used to generate chimeric mice. iPS cells were injected into blastocysts from albino or black C57B6 mice (NCI) and the contribution of iPSCs can be seen with agouti or black coat color.

Figure 2.4. Characterization of iPS clones derived from miR mimic experiments.



Furthermore, these cells could contribute to the genital ridge of derived E13.5 embryos (Figure 2.S11). Taken together, these results indicate that the enhancing effects of miR-93 and 106b on reprogramming do not alter differentiation capacity of induced pluripotent cells and that those derived clones can differentiate into all three germ lines.

2.3.5. miR-93 and 106b target Tgfbr2 and p21

To further understand the mechanism underlying miR-93 and 106b enhancement of reprogramming efficiency, we investigated cellular targets of these microRNAs. We chose miR-93 for analysis since it shares the same seed region as miR-106b. miR-93 mimics were transfected into MEFs, and total RNAs were harvested at day2 for mRNA expression profile analysis. That analysis identified potential functional targets of miR-93 that we compared with published expression profiles of MEFs and iPSCs (Sridharan et al, 2009). We found that genes significantly decreased upon miR-93 transfection showed a threefold enrichment of genes which are lowly expressed in iPSCs (Figure 2.S13.A), while genes which were increased upon miR-93 transfection did not show such enrichment. In addition, we undertook pathway ontology analysis of the expression profile of miR-93 transfected MEFs (data not shown). Interestingly, two important pathways for iPS induction were regulated by miR-93: TGF-β signaling and G1/S transition pathways.

For TGF-β signaling, Tgfbr2 is among one of the most significantly decreased genes upon miR-93 transfection. Tgfbr2 is a constitutively active receptor kinase that plays a critical role in TGF-β signaling, and recent small molecule screens indicate that inhibitors of its heterodimeric partner Tgfbr1 enhance iPSC induction (Ichida et al, 2009; Maherali & Hochedlinger, 2009). MicroRNA target site prediction suggested that there were two conserved targeting site for miR-93 and its family microRNAs in its 3‘UTR. Therefore we choose it as the candidate target for further investigation.

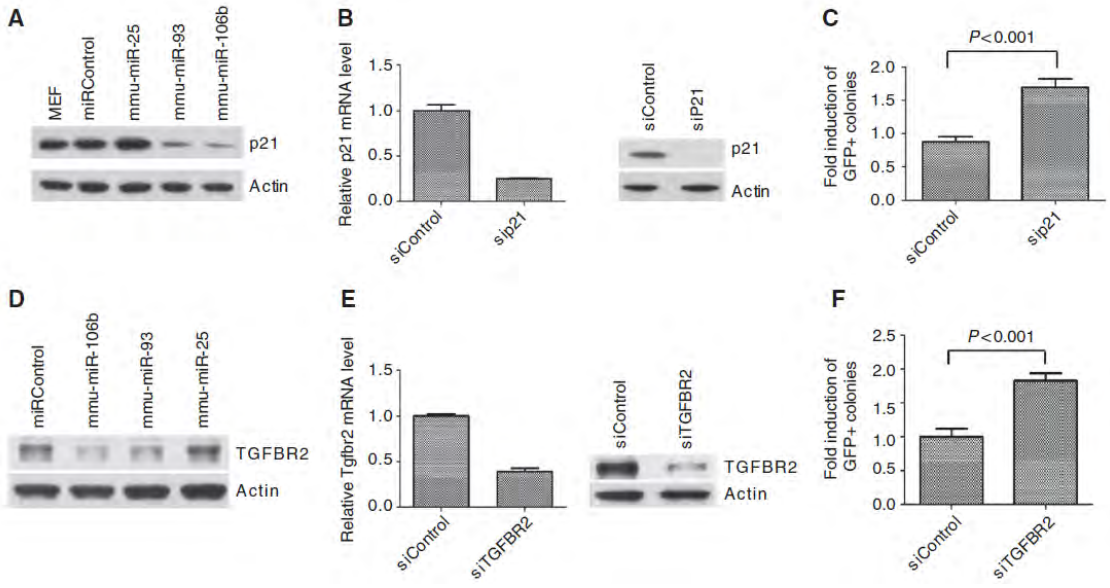
Regarding the G1/S transition, we choose p21 as the potential target because recent results in human solid tumor samples (breast, colon, kidney, gastric, and lung) and gastric cancer cell lines indicate that the miR-106b~25 cluster can target cell cycle regulators, such as the CDK inhibitors p21 and p57 (Ivanovska et al, 2008; Kim et al, 2009b) and that human and mouse p21 share a conserved miR-93/106b target site in the 3‘UTR. Furthermore, mouse ES cell-specific microRNA clusters, such as miR-290 and 293, reportedly target negative regulators of the G1-S transition, including p21 (Wang et al, 2008). miR-290 and 293 cluster microRNAs also share very similar seed regions with miR-93 and 106b (unpublished observations). p21 is also greatly induced by OSKM factors during early stages of iPSC induction, (Kawamura et al, 2009) an up-regulation that we confirmed in MEFs (Figure 2.S12.A). Detailed analysis revealed that p21 induction is primarily due to Klf4 and cMyc misexpression, as a combination of Oct4 and Sox2 only did not significantly alter p21 protein levels (Figure 2.S12.A).

To determine whether mouse Tgfbr2 and p21 are targeted by miR-93 and 106b, miR mimics were transfected into MEFs, and total cell lysates were analyzed by Western blotting 48 hours later. Indeed, miR-93 and 106b expression efficiently decreased both Tgfbr2 and p21 protein levels (Figure 2.5.A,D) and p21 mRNA levels were decreased by ~25-30% while Tgfbr2 was decreased by ~60-70% (Figure 2.S14.A,B). These levels of suppression were further confirmed in 4F and OSK infected MEFs (Figure 2.S15&S16). To determine whether p21 is a direct target of miR-93 and 106b, we constructed a luciferase reporter with p21 3‘UTR sequence inserted downstream of the firefly luciferase coding sequence. We observed consistent ~40% repression of luciferase activity following transfection of miR-93 and 106b mimics into co-transfected Hela cells, a repression lost when mutations were introduced into the seed region of conserved p21 3‘UTR target sites (Figure 2.S17.A,B). For Tgfbr2, luciferase assay also showed ~50% decrease of GL activity while miR-93 mutant did not have such effect (Figure 2.S18.A,B).

Cell cycle arrest promoted by p21 may inhibit epigenetic modifications required for reprogramming, since those modifications occur more readily in proliferating cells. To determine whether p21 expression compromises iPS, HA-tagged p21 cDNA was cloned into the pMX retroviral backbone and overexpressed in MEF cells. When HA-p21 virus was introduced into MEFs together with the four OSKM factors, an almost complete inhibition of iPS induction was observed, based on both alkaline phosphatase staining and Oct4-

Figure 2.5. miR-93 and 106b directly target mouse p21 and Tgfbr2.

(A) miR-93 and 106b transfection decreases p21 protein levels. Oct4-GFP MEFs were transfected with 50nM miR mimics and harvested 48hr after transfection for Western analysis. Actin was used as the loading control. (B) p21 is knocked down efficiently by siRNA. P21 siRNA- and control-transfected MEFs were harvested at 48hr and RT-qPCR, and western blotting was undertaken to verify p21 expression. p21 mRNAs were normalized to GAPDH. (C) Knock-down of p21 by siRNA enhances iPSC induction. MEFs were infected with 4F virus, and siRNAs were transfected following the same timeline as microRNAs mimic transfection. GFP+ colonies were quantified at day 11. Error bars represent three independent experiments using triplicate wells. (D) miR-93 and 106b transfection decreases TGFBR2 expression. Transfected cells were harvested at 48hr for western blotting. (E) Tgfbr2 is knocked down by siRNAs. Relative Tgfbr2 mRNA levels were normalized to those of Gapdh. (F) Knock-down of Tgfbr2 by siRNAs enhances iPSC induction. Error bars represent four independent experiments in triplicate wells.



GFP-positive colony formation (Figure 2.S19.A). Similar results were obtained when the three OSK factors were used for reprogramming (Figure 2.S19.B).

Since our analysis indicated that miR-93 and 106b efficiently repress both Tgfbr2 and p21 expression, we asked whether Tgfbr2 and p21 activity antagonizes reprogramming. To do so, we transfected Tgfbr2 or p21 siRNAs into MEFs using the same experimental time line employed with microRNA mimics. Western blotting and RT-qPCR confirmed that both protein and mRNA levels, respectively, were efficiently knocked down by siRNAs without virus transduction (Figure 2.5.B, E). MEF reprogramming was then initiated by OSKM transduction, and Oct4-GFP+ colonies were quantified at day 11 post-transduction. We observed a ~2-fold induction in colony number for each gene (Figure 2.5.C, F). TGFBR2 was also overexpressed in MEFs and iPS enhancement by miR-106b was compromised under such condition (Figure 2.S20). All together, our data identify that Tgfbr2 and p21 are the direct target of miR-93 and 106b and down regulation of these genes can enhance the reprogramming process.

2.3.6. Additional upregulated microRNAs enhance iPSC induction

As noted, we identified three microRNA clusters induced by reprogramming factors, and several microRNAs within these clusters have the same seed regions, suggesting that they target similar mRNAs (Figure 2.2.). To investigate whether other microRNAs that share the same seed region with miR-93 and 106b also enhance iPSC induction, microRNA mimics of miR-17 and 106a were

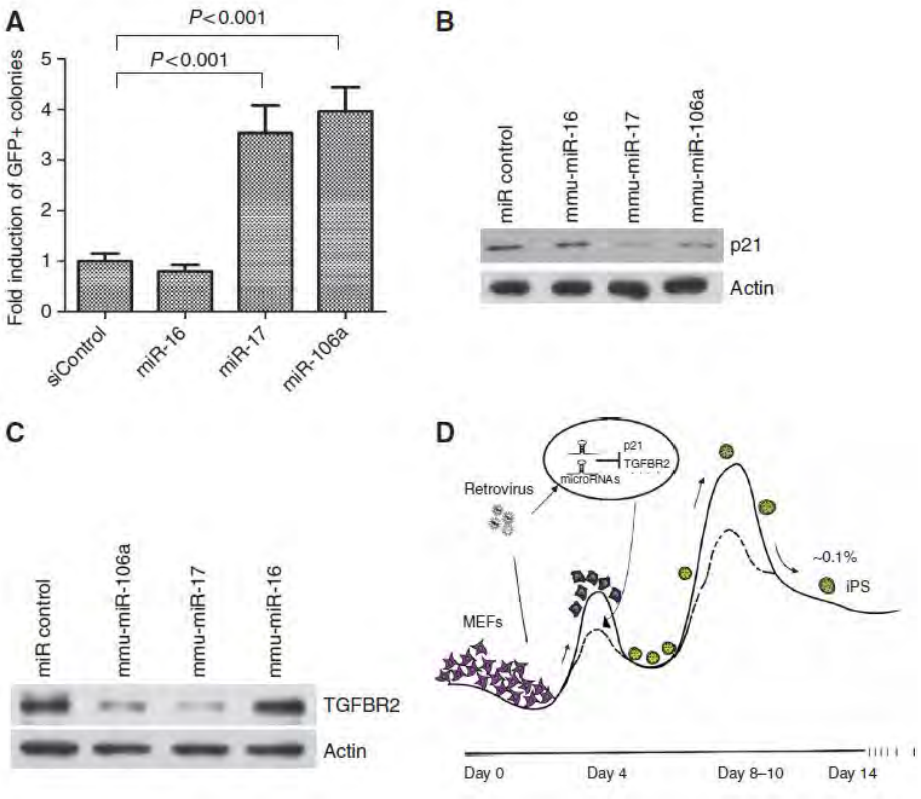
tested using an experimental procedure similar to that described above for miR-93 mimic treatment and iPSC induction. These microRNAs enhanced reprogramming in a manner similar to that seen with the miR-106b~25 clusters (Figure 2.6.A), and transfection of these miRs resulted in decreased TGFBR2 and p21 protein levels (Figure 2.6.B,C) as well as Tgfbr2 mRNA (Figure 2.S21.). Together, this evidence suggests that induction of miR-17~92, miR-106b~25 and miR-106a~363 clusters is important for proper reprogramming and that upregulation of these microRNAs lowers reprogramming barriers to the iPSC generation process (Figure 2.6.D).

2.4. Discussion

Since the discovery that MEFs can be reprogrammed to iPS cells, much effort has been directed toward understanding fundamental mechanisms underlying this process. Our results show for the first time that post-transcriptional gene regulation occurs during reprogramming and that interference with the RNAi machinery can significantly alter reprogramming efficiency. We identified three microRNAs clusters significantly up-regulated by the four factors used to induce iPS cells and found that microRNAs in those clusters likely target two important reprogramming pathways: TGF-β signaling and cell cycle control. While these experiments were in progress, several investigators also reported that the p53 pathway, which includes downstream tumor suppressors such as p21, is a major

Figure 2.6. Reprogramming is enhanced by other family microRNAs

(A) miR-17 and miR-106a can also enhance reprogramming efficiency. miR-17 and 106a mimics were transfected into MEFs at a final concentration of 50nM. GFP+ colonies were quantified at day 11 post-transduction. Error bars represent three independent experiments in triplicate wells. (B) miR-17 and 106a also target p21. p21 Western blotting was performed 2 days after transfection of microRNA mimics into MEFs. (C) miR-17 and 106a target TGFBR2 expression. microRNA mimics were transfected into MEFs at 50nM final concentration. Western blotting was performed 2 days post transfection. (D) Model for the role for microRNAs during iPS induction. Several microRNAs, including miR-17, 25 and 106a clusters, are induced during early stages of reprogramming. These microRNAs facilitate full reprogramming by targeting factors that antagonize the process, such as p21 and other unidentified proteins. Up and down represent the potential different stages and barriers during reprogramming process and dashed line indicates that barriers for reprogramming which are lowered upon microRNAs induction in reprogrammed cells.



barrier to iPSC induction (Banito et al, 2009; Hong et al, 2009; Kawamura et al, 2009; Li et al, 2009a; Utikal et al, 2009). Much evidence indicates that ectopic expression of the four factors (OSKM) readily up-regulates p53 and initiates cellular —defense programs," such as cell cycle arrest, apoptosis or DNA damage responses. These responses likely underlie low reprogramming efficiency, which we observe to be ~0.1%. However, these data do not explain how successfully reprogrammed cells overcome these barriers to become iPS cells. Our data suggest that cells do so in part by inducing expression of microRNAs that target pathways that antagonize successful reprogramming. By modulating microRNA levels in primary fibroblasts, we were able to achieve a significant increase in reprogramming efficiency.

TGF-β signaling is an important pathway that functions in processes as diverse as gastrulation, organ-specific morphogenesis and tissue homeostasis (Moustakas & Heldin, 2009). The current model of canonical TGF-β transduction indicates that TGF-β ligand binds the TGF-β receptor II (TGFBR2), which then heterodimerizes with TGFBR1 to transduce signals through receptor-associated Smads (Kahlem & Newfeld, 2009). TGF-β signaling reportedly functions in both human and mouse ES cell self-renewal, and FGF2, a widely used growth factor for ES cell culture, induces TGF-β ligand expression and suppresses BMP-like activities (Greber et al, 2007; Ogawa et al, 2007). Blocking TGF-β receptor I family kinases by chemical inhibitors compromises ES cell self-renewal (Ogawa et al, 2007). These findings are particularly significant for iPSC induction,

because those inhibitors seem to have completely different roles during reprogramming. Recent chemical screening has shown that small molecules inhibitors of the TGF-β receptor I (TGFBR1) actually enhance iPSC induction and can replace the requirement for Sox2 by inducing Nanog expression (Ichida et al, 2009). Moreover, treating reprogramming cells with TGF-β ligands has a negative effect on iPSC induction (Maherali & Hochedlinger, 2009). Therefore, although TGF-β signaling is important for ES cell self-renewal, it is a barrier for reprogramming. Our results determined that, in addition to TGFBR1, activity of the constitutively active kinase TGFBR2 also antagonizes reprogramming. Here, for the first time, we demonstrate that miR-93 and its family members directly target TGFBR2 to modulate it‘s signaling and reprogramming.

p21, a protein of only 165 amino acids, functions as a tumor suppressor by mediating p53-dependent G1 growth arrest and promoting differentiation and cellular senescence (Abbas & Dutta, 2009). Our data (Suppl. Fig. 11) and that of others (Kawamura et al, 2009) demonstrate that p21 expression is up-regulated when the four factors (OSKM) are introduced into MEFs and that this up-regulation antagonizes reprogramming, since p21 overexpression almost completely blocked iPSC induction (Figure 2.S16.). p21 induction in reprogramming cells could be dependent or independent of p53, as the Klf4 reprogramming factor reportedly binds to the p21 promoter and increases p21 transcription (Abbas & Dutta, 2009). This finding raises an interesting question regarding the function of the four reprogramming factors, since the same

transcription factor can both promote and antagonize iPSC induction. In fact, we cannot currently rule out the possibility that a certain level of p21 induction benefits the reprogramming process. Besides its well-known role in p53-dependent cell cycle arrest, p21 also reportedly has an oncogenic activity by protecting cells from apoptosis, a function unrelated to its usual role in cell cycle control (Abbas & Dutta, 2009). A potential benefit for p21 in reprogramming may depend on its ability to regulate gene expression through protein-protein interactions (Abbas & Dutta, 2009). For example, p21 directly binds to several proteins regulating apoptosis, such as caspases 8 and 10 and procaspase 3. It also suppresses pro-apoptotic activity of Myc by associating with the Myc N-terminus to block Myc-Max heterodimerization (Abbas & Dutta, 2009). Indeed, when Myc itself is overexpressed in MEFs, a significant increase in cell death is observed in cell culture, while in four-factor transduced cells, cell death is minimal compared to myc-only samples (data not shown). Therefore, p21 induction may not only serve as a barrier to reprogramming but may maintain levels of p21 necessary to reduce apoptosis and thus increase reprogramming efficiency. Our data serves as partial evidence to support this hypothesis, since miR-93 and 106b treatment had greater enhancing effects on reprogramming than did p21 siRNA transfection (Figure 2.3.C, Figure 2.5.A, C). It is also possible that this effect is due to targeting of multiple proteins such as TGFBR2 in addition to p21 by these microRNAs.

Finally, since the miR clusters identified here, such as miR-17~92, miR-106b~25 and miR-106a~363, are induced during iPS induction and are conserved between mouse and humans, the enhancing effects of miR-93 and 106b may apply to human reprogramming. Further studies should focus on the activity of these microRNAs in human cells and in various disease models.

Materials and Methods

Cell culture, vectors and virus transduction

Oct4-GFP MEFs were derived from mice carrying an IRES-EGFP fusion cassette downstream of the stop codon of *pou5f1* (Jackson lab, Stock#008214) at E13.5. MEFs were cultured in DMEM (Invitrogen, 11995-065) with 10% FBS (Invitrogen) plus glutamine and NEAA. Only MEFs at passage of 0 to 4 were used for iPS induction. pMX-Oct4, Sox2, Klf4 and cMyc were purchased from Addgene. pMX-HA-p21 was generated by inserting N-terminally tagged-p21 into the pMX EcoRI site. pLKO-shRNA clones were purchased from Open Biosystems. To generate retrovirus, PLAT-E cells were seeded in 10cm plates, and 9ug of each factor were transfected the next day using Lipofectamine (Invitrogen, 18324-012) and PLUS (Invitrogen, 11514-015). Viruses were harvested and combined 2 days later. For iPS induction, MEFs were seeded in 12-well plates and transduced with "four factor" virus the next day with 4ug/ml Polybrene. One day later, the medium was changed to fresh MEF medium, and 3 days later it was changed to mES culture medium supplemented with LIF (Millipore, ESG1107). GFP+ colonies were picked at day 14 post-transduction, and expanded clones were cultured in DMEM with 15%FBS (Hyclone) plus LIF, thioglycerol, glutamine and NEAA. Irradiated CF1 MEFs served as feeder layers to culture mES cells and derived iPS clones. To generate shRNA lentivirus, shRNA lentiviral vectors were cotransfected into 293FT cells together with the pPACK-H1 packaging system

(SBI, LV500A-1). Lentiviruses were harvested at day 2 after transfection and centrifuged at 4000 rpm for 5min at room temp. shRNA virus was added together with 4 factor virus at a volume ratio of 1:1:1:1:1.

MicroRNA and siRNA transfection of MEFs

MicroRNA mimics and inhibitory siRNAs were purchased from Dharmacon. To transfect MEFs, microRNA mimics or inhibitors were diluted in Opti-MEM (Invitrogen, 11058-021) to the desired final concentration. Lipofectamine 2000 (Invitrogen, 11668-019) was added to the mix at 2ul/well in 12 well plates, which were incubated 20 min at RT. For 12-well transfections, 80ul of the miR mixture was added to each well with 320ul of Opti-MEM. Three hrs later, 0.8ml of the virus mixture (for iPS) or fresh medium was added to each well and the medium was changed to fresh MEF medium the next day.

Western blotting

Total cell lysates were prepared by incubating cells in MPER buffer (PIERCE, 78503) on ice for 20 min, and then cleared by centrifugation at 13,000 rpm for 10 min. An equal volume of lysates was loaded onto 10%SDS-PAGE gels, and proteins were transferred to PVDF membranes (Bio-Rad, 1620177) using the semi-dry system (Bio-Rad). Membranes were blocked with 5% milk in TBST for at least 1hr at room temp or overnight at 4°C. Antibodies used include: anti-p21 (BD, 556430), anti-mNanog (R&D, AF2729), anti-h/mSSEA1 (R&D, MAB2156), anti-HA (Roche, 11867423001), anti-mAgo2 (Wako, 01422023), anti-Dicer

(Abcam, ab13502), anti-Drosha (Abcam, ab12286), anti-Actin (Thermo, MS1295P0), anti-AFP (Abcam, ab7751), anti-Beta III tubulin (R&D systems, MAB1368), anti-TGBR2 (Cell signaling, 3713s) and anti-alpha actinin (Sigma, A7811).

mRNA and microRNA quantitative PCR

Total RNAs were extracted using Trizol (Invitrogen). After extraction, 1ug total RNA was used for RT using Superscript II (Invitrogen). Quantitative PCR was performed using a Roche LightCycler480 II and the Sybr green mixture from Abgene (Ab-4166). Mouse Ago2, Dicer, Drosha, Gapdh and p21 primers are listed in Supplementary Table 2. Other primers were previously described (Takahashi & Yamanaka, 2006). For microRNA quantitative analysis, total RNA was extracted using the method above. After extraction, 1.5~3ug of total RNA was used for microRNA reverse transcription using QuantiMir kit following the manufacturer‘s protocol (SBI, RA420A-1). RT products then were used for quantitative PCR using the mature microRNA sequence as a forward primer together with the universal primer provided with the kit.

Immunostaining

Cells were washed twice with PBS and fixed with 4% paraformaldehyde at room temperature for 20min. Fixed cells were permeabilized with 0.1% Triton X-100 for 5min. Cells were then blocked in 5% BSA in PBS containing 0.1% Triton X-100 for 1hr at room temperature. Primary antibody was diluted from 1:100 to 1:400 in

2.5% BSA PBS containing 0.1% Triton X-100, according to the manufacturer‘s suggestion. Cells were stained with primary antibody for 1hr and then washed three times with PBS. Secondary antibody was diluted 1:400 and cells were stained for 45min at room temperature.

EB formation and differentiation assay

iPS cells were trypsinized into a single cell suspension and the hanging drop method was used to generate embryoid bodies. For each drop, 4000 iPS cells in 20ul EB differentiation medium were used. EBs were cultured in hanging drops for 3 days before being reseeded onto gelatin-coated plates. After reseeding, cells were further cultured until day 14 when beating areas could be identified.

Promoter methylation analysis

CpG methylation of the Nanog and Pou5f1 promoters was analyzed following procedures described elsewhere (Takahashi & Yamanaka, 2006). Briefly, genomic DNA of derived clones was extracted using a Qiagen kit. 1ug DNA was then used for genome modification analysis following the manufacturer‘s protocol (EZ DNA Methylation –Direct Kit, Zymo Research, D5020). After modification, PCR of selected regions was performed, and products were cloned into pCR2.1-TOPO (Invitrogen). Ten clones were sequenced for each gene.

Teratoma formation and chimera generation

To generate teratomas, iPS cells were trypsinized and resuspended at a concentration of $1x10^7$ cells/ml. Athymus nude mice were first anesthetized with Avertin, and then approximately 150 ul of the cell suspension was injected into each mouse. Mice were checked for tumors every week for 3~4 weeks. Tumors were harvested and fixed in zinc formalin solution for 24hrs at room temp before paraffin embedding and H&E staining. To test the capacity of derived iPSC clones to contribute to chimeras, iPS cells were injected into C57BL/6J-$Tyr^{(C-2J)/J}$ (albino) blastocysts. Generally, each blastocyst received 12-18 iPS cells. ICR recipient females were used for embryo transfer. The donor iPSC cells are either in agouti or black color.

mRNA microarray analysis

miR-93 and siControl were transfected into MEFs and total RNAs were harvested at 48hrs post transfection. mRNA microarray was carried out by Microarray facility in Sanford-Burnham institute. Gene lists for both potential functional targets (fold change >2, $p<0.05$) and total targets (fold change >25%, $p<0.05$) were generated by filtering through volcano maps. Gene lists were then used for ontology analysis using GeneGo software following guidelines from the company.

Dual luciferase assay

3‘UTR of both p21 and Tgfbr2 were cloned into XbaI site of pGL3 control vectors. For each well of 12-well plates, 200ng of resulted vectors and 50ng of pRL-TK (renilla luciferase) were transfected into $1x10^5$ Hela cells which were seeded one

day before the transfection. 50nM of microRNAs were used for each treatment and cell lysates were harvested at day 2 post transfection. 20ul of lysates were then used for dual luciferase assay following manufacturer‘s protocol (Dual-Luciferase® Reporter Assay System Promega, E1910)

Cell proliferation assay

3000 MEFs were seeded in each well in 96-well plates and transduced with 4F virus and shRNA lentivirus (or transfected with microRNA inhibitors). Starting from day 1 post transduction/transfection, every two days, cells were incubated with mES medium containing Celltiter 96 Aqueous one solution (Promega, G3580) for 1hr in tissue culture incubator. Absorbance at 490nm was then measured for each well using plate reader and collected data was used to generate relative proliferation curve using signal from day 1 post transduction/transfection as the reference.

Acknowledgments

We thank Greg Hannon, Craig Mello, Mark Mercola, Evan Snyder, and Rana laboratory members for helpful discussions and support. We are grateful for the following shared resource facilities at the Sanford-Burnham Institute: genomics and informatics and data management core for miRNA and mRNA array experiments and data analysis; the animal facility for the generation of teratomas and chimeric mice; and the histology and molecular pathology core for characterization of various tissues.

Figure 2.S1. Knock-down of Dicer and Drosha decrease iPS induction

(A) Drosha can be efficiently knocked down by shRNAs. MEFs were transduced with lentiviral shRNAs plus 4ug/ul polybrene, and total RNAs or proteins were harvested at day 3 post-transduction. mRNA and protein levels of targeted genes were analyzed by RT-qPCR and Western blotting, respectively. (B) Dicer can be efficiently knocked down by shRNAs. The procedure is the same as Ago2 and Drosha shRNA experiments. (C) Knock-down of Dicer and Drosha decrease iPS colony formation. Alkaline phosphatase staining was performed at day14 post transduction.

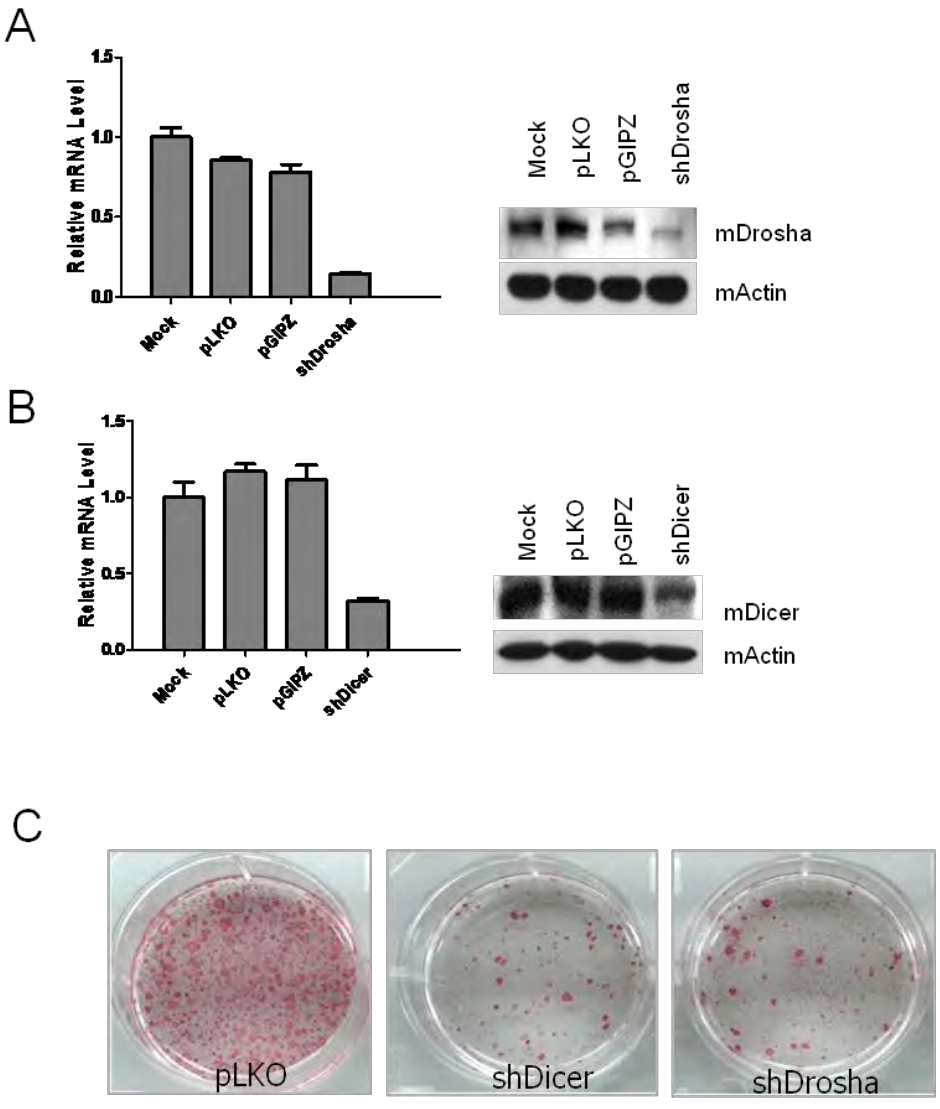


Figure 2.S2. Proliferation curve of MEFs with 4F and shRNAs

MEFs were seeded in 96-well plates and incubated with Celltiter 96 Aqueous One solution for 1hr at 37°C before absorbance reading at 490nm. Signal from day1 samples was used as the reference to calculate relative proliferation curve. Error bar represents standard deviation of 6 wells for each treatment. Drosha and Dicer knock-down by shRNAs resulted in a gradual loss of transduced cells while shAgo2 transduced ones had similar proliferation rate as control non-targeting samples.

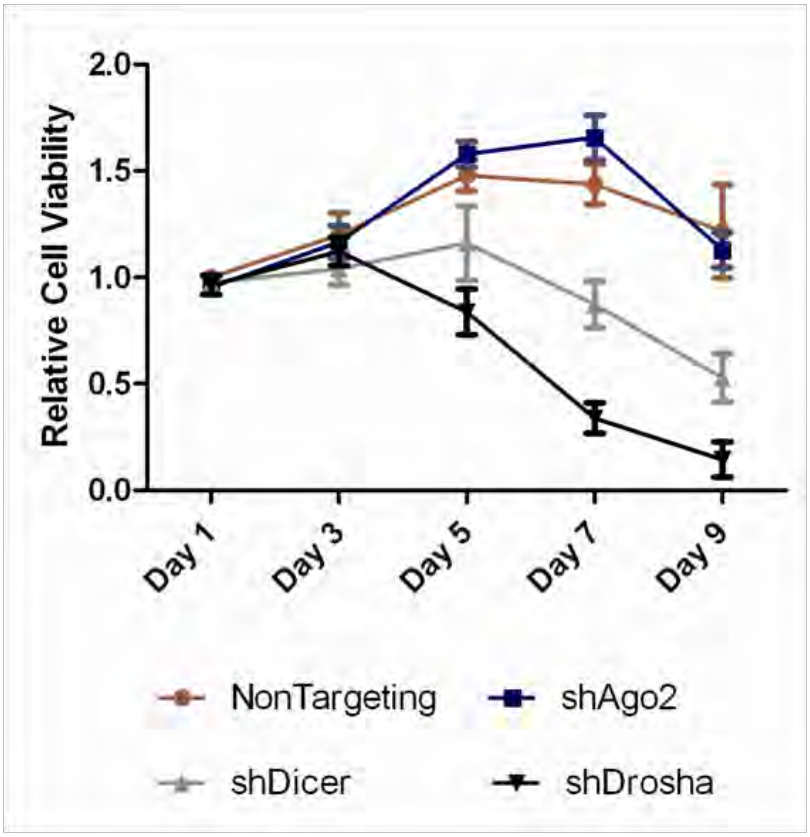


Figure 2.S3. Characterization of shAgo2 iPSC clones

(A) Picked shAgo2 clones were positive for shRNA integration in the genome. Genomic DNAs from shAgo2 iPSCs were extracted and analyzed by PCR for integration of puromycin, which is the drug resistant gene contained in the shRNA virus. Primers for endogenous PGK locus were used as the loading control. (B) shAgo2 iPSCs were actively expressing shRNAs. Northern blotting was used to detect the expression of shAgo2 shRNAs. rRNA was used as the loading control. (C) shAgo2 iPSCs were expressing all the tested mES markers. Nanog, Oct4, Eras and ECatI expression were analyzed in picked clones and they were all positive for those tested markers. (D) shAgo2 iPSCs have activated the endogenous Oct4 locus and are positive for immunostaining of Nanog. (E) shAgo2 iPSCs have similar expression level of Tgfbr2 and p21 as mES cells. (F) shAgo2 iPSCs have compromised tendency to differentiation. mES cells and shAgo2 iPS clones were treated with RA for two days to analyze their differentiation tendency. Nanog was used as the self renewal marker. Mouse embryonic stem cell line CCE was used as control. Error bar represents data from duplicate samples.

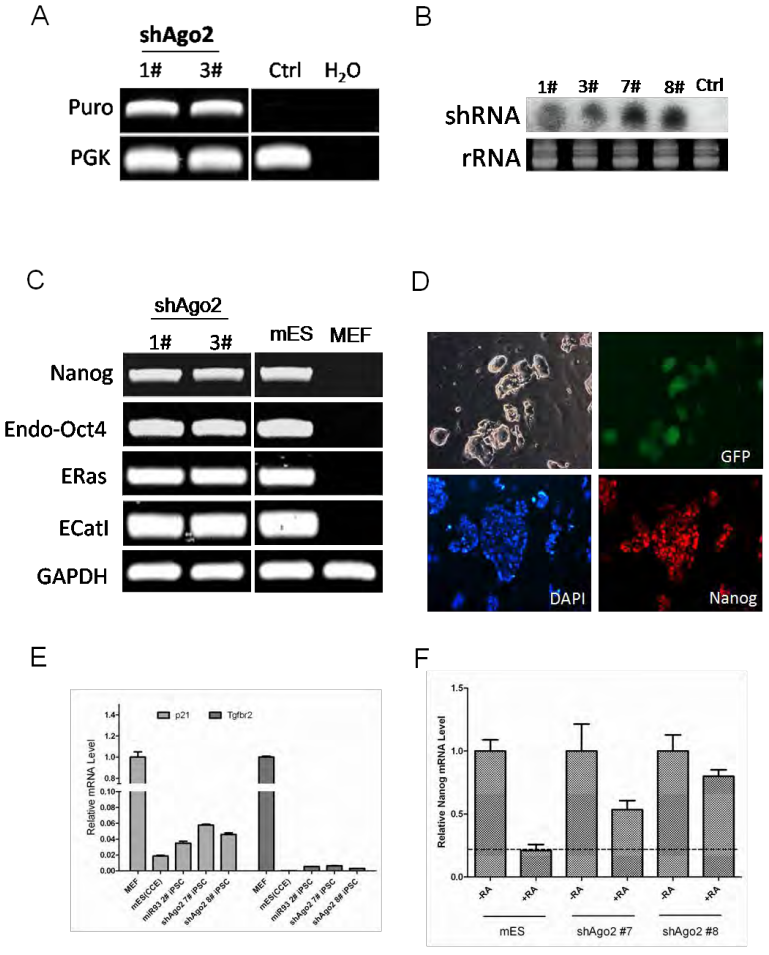


Figure 2.S4. miR expression profile at different reprogramming stages

MEFs were transduced with 4F and harvested at different time points (days 0, 4, 8 and 12). Total RNAs were extracted using Trizol and microRNA RT-qPCR was used to evaluate expression changes of different miRs. Data was normalized to U6 expression, and expression data at day 12 was used as the reference (100%).

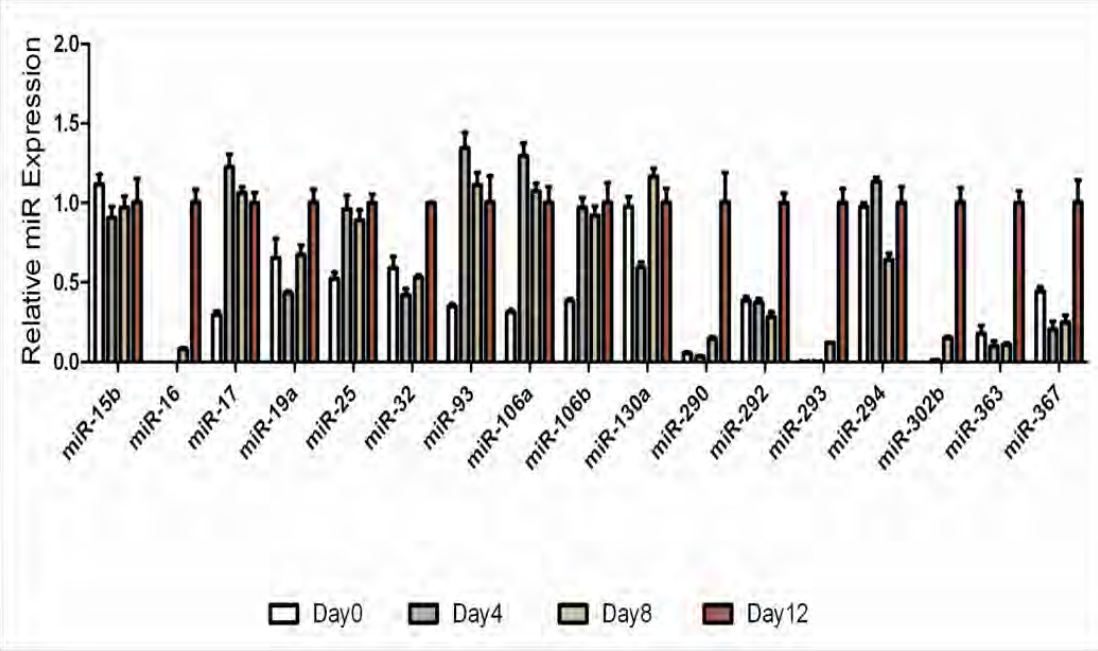


Figure 2.S5. Dose/response analysis of the effect of miR-93 and 106b on mouse iPS induction

Oct4-GFP MEFs were transfected with different concentrations (5, 15 and 50nM) of indicated microRNAs. Nontargeting siRNA was used as the control. GFP+ colonies were quantified at day 11 post-transduction. Data represents triplicate wells of 12-well plates.

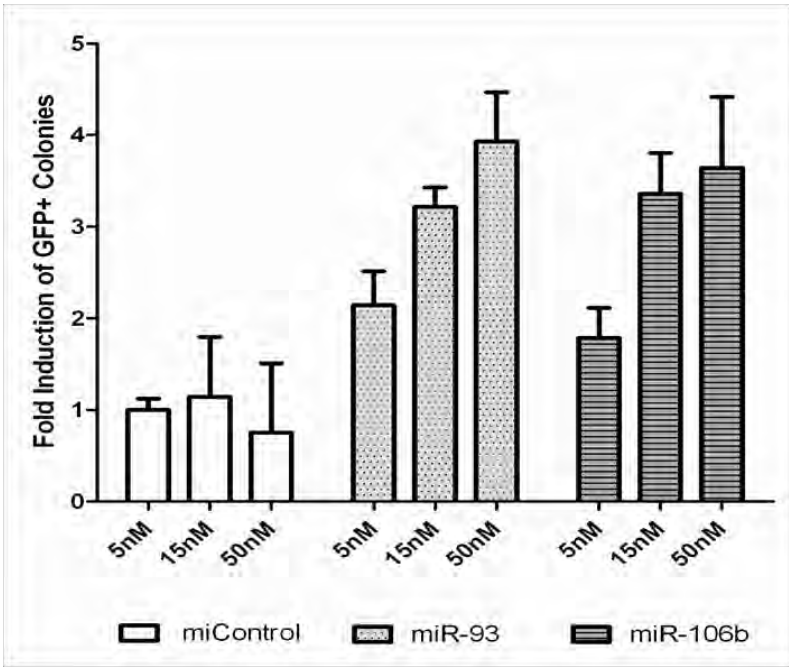


Figure 2.S6. miR-106b enhances reprogramming of MEFs to iPSCs

(A) miR-106b transfection increased endogenous Nanog expression in 4F transduced cells. Cells were harvested at day12 post transduction for RT-qPCR analysis of Nanog gene. Data was normalized to Gapdh expression. Error bar represents standard deviation of three separate wells. (B) miR-106b increased Nanog expression in OSK transduced cells. Cells were harvested at day15 for RT-qPCR analysis of Nanog. Error bar represents standard deviation of three separate wells. (C) Oct4-GFP+ colonies were also positive for Nanog staining. GFP+ colonies from 4F (day12) and OSK (day15) infected cells were fixed and stained with Nanog antibody to detect activation of Nanog expression. (D,E) Quantification of both Oct4-GFP+ and Nanog+ colonies generated by 3 factors OSK (d) or 4 factors OSKM confirmed that miR-106b could enhance the reprogramming of MEFs to bonafide iPSCs. Error bar represents standard deviation of three separate wells.

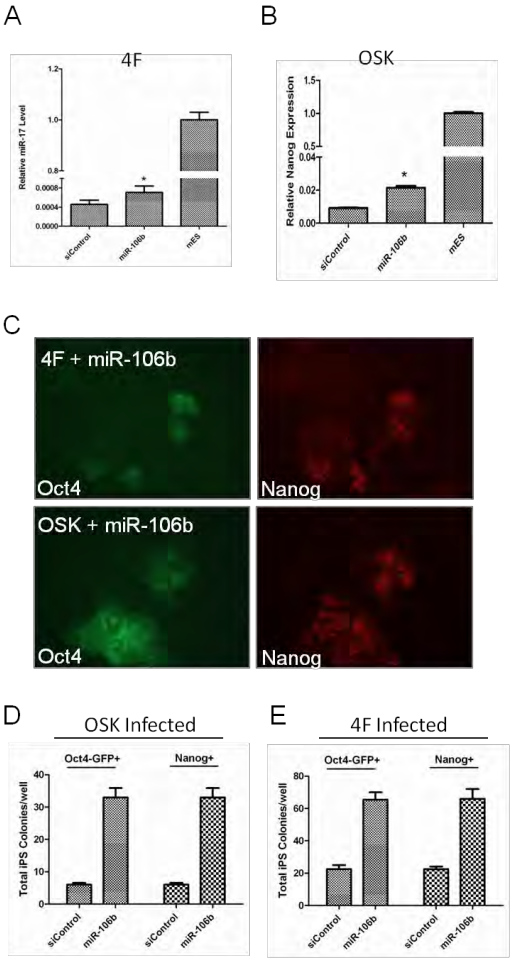


Figure 2.S7. miR mimic level in transfected MEFs

(A) miR-93 level in transfected MEFs. MEFs were transfected with different concentration of miR-93 and the level of miR mimics were analyzed at day 3 post transfection by RT-qPCR. (B) microRNAs can be efficiently delivered to MEFs. Fluorescence labeled siRNAs were tranfected into MEFs and transfection efficiency was monitored at day1 post transfection. Almost all the cells were positive for Cy3 signal. (C) Decay of transfected microRNA mimics in MEFs. miR-17 was transfected into MEFs and cells were harvested at different time points and qPCR was used to quantify the relative level of miR-17 in these cells. microRNA mimics showed a half life of ~4 days in MEF cells.

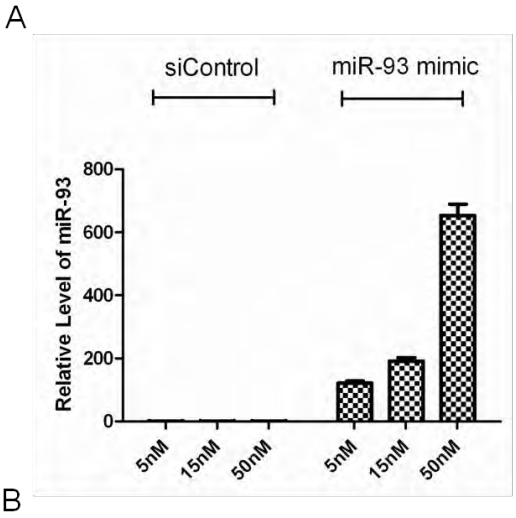


Figure 2.S8. microRNA mimics do not seem to alter overall AP+ colony formation, while microRNA inhibitors do

(A, B) Reprogrammed cells at day 12 were stained using alkaline phosphatase substrates. Transfection with miR mimics did not significantly alter the number of AP+ colonies; however, miR-93 and 106b knock-down resulted in significant loss of AP+ and GFP+ colonies. Non-targeting siRNA served as control for miR mimic experiments, and non-targeting hairpin inhibitors were used as controls for the inhibitor experiments (labeled as inh-Control).

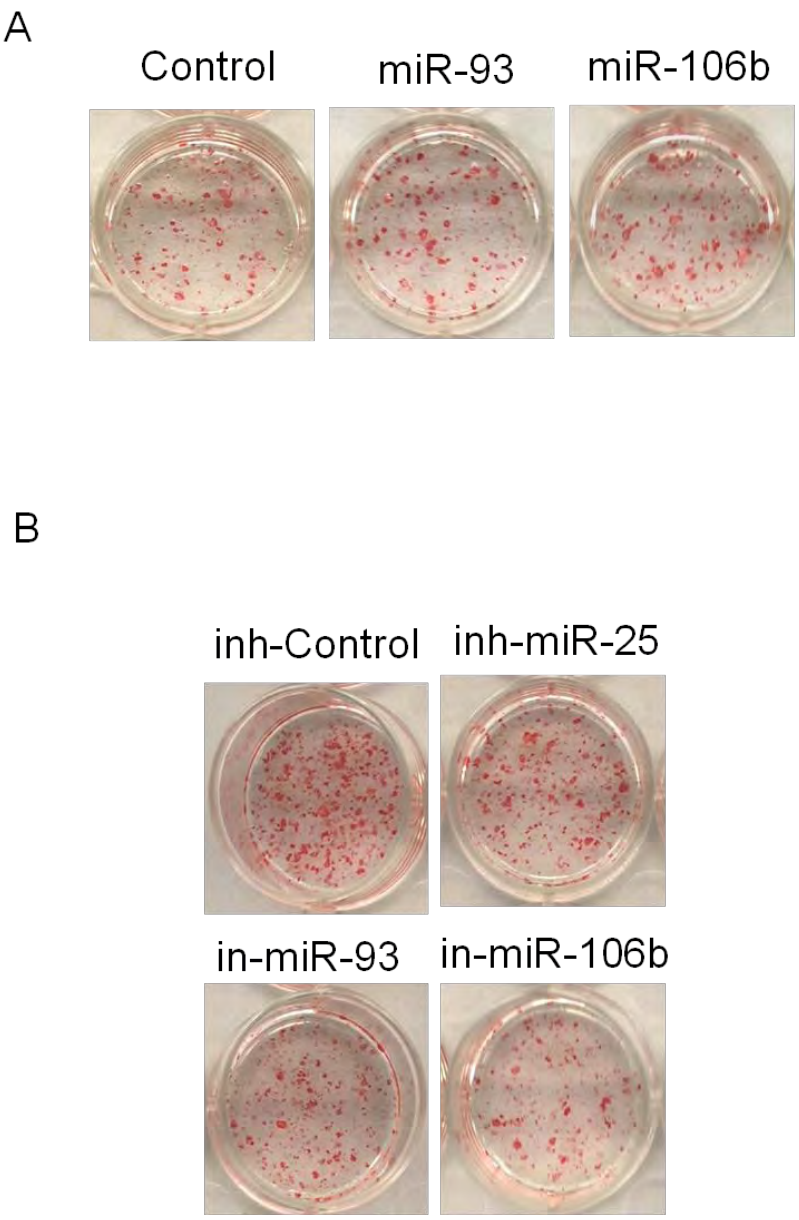


Figure 2.S9. Efficacy of microRNA inhibitors in transfected MEFs

(A) microRNAs can be efficiently knocked by hairpin inhibitors. miR inhibitors were transfected into MEFs at 50nM and total RNAs were harvested at day4 post transfection for miR expression analysis by RT-qPCR. (B) microRNAs could be transfected into MEFs efficiently. miR mimics were transfected into MEFs at 50nM and total RNAs were harvested at day2 post transfection for miR quantification. Error bar represents duplicate samples. (C) Transfection of miR inhibitors did not change the proliferation of MEFs. MEFs were seeded in 96-well plates and transfected with 50nM inhibitors. Starting from day1, cells were incubate with Celltiter 96 aqueous one solution for 1hr at 37°C before absorbance reading at 490nm. All signals were normalized using day1 signal as the reference. Error bar represents standard deviation of 6 wells for each treatment.

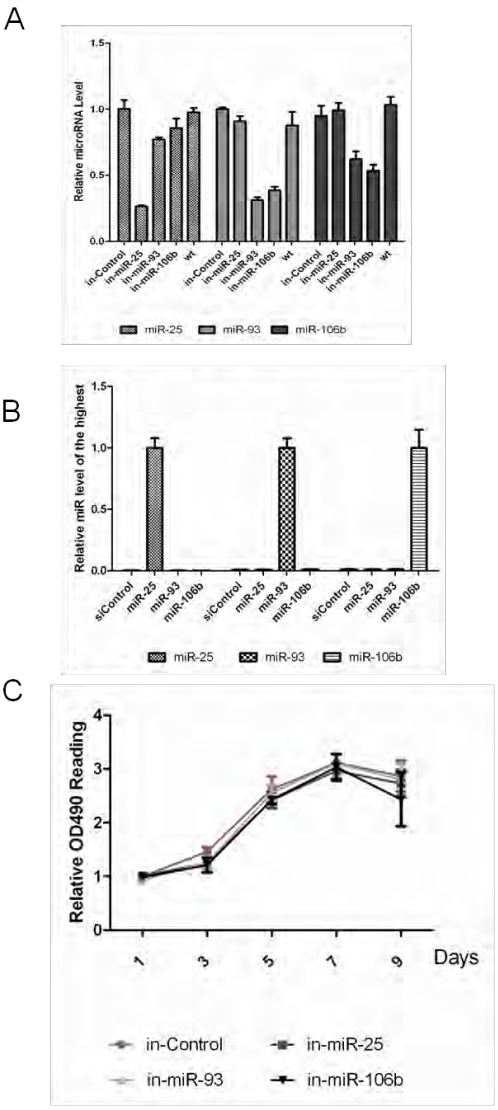


Figure 2.S10. Characterization of derived iPS clones

(A) Global mRNA comparison of derived iPS clones versus a mouse CCE ES cell line confirmed that all derived iPS lines were clustered with mES rather than starting MEFs. Total RNAs were isolated from iPS cell lines at day 3 post-passage and analyzed for global mRNA expression profiling using Illumina mRNA array chips. (B) The promoter of the ES cell marker gene Nanog is demethylated in derived iPS clones. Genomic DNAs from starting MEFs and iPS lines were isolated and then treated with bisulfite reagent using the Zymo DNA Methylation Direct kit. Modified DNAs were analyzed by PCR and subcloned into the pTOPO vector for sequencing.

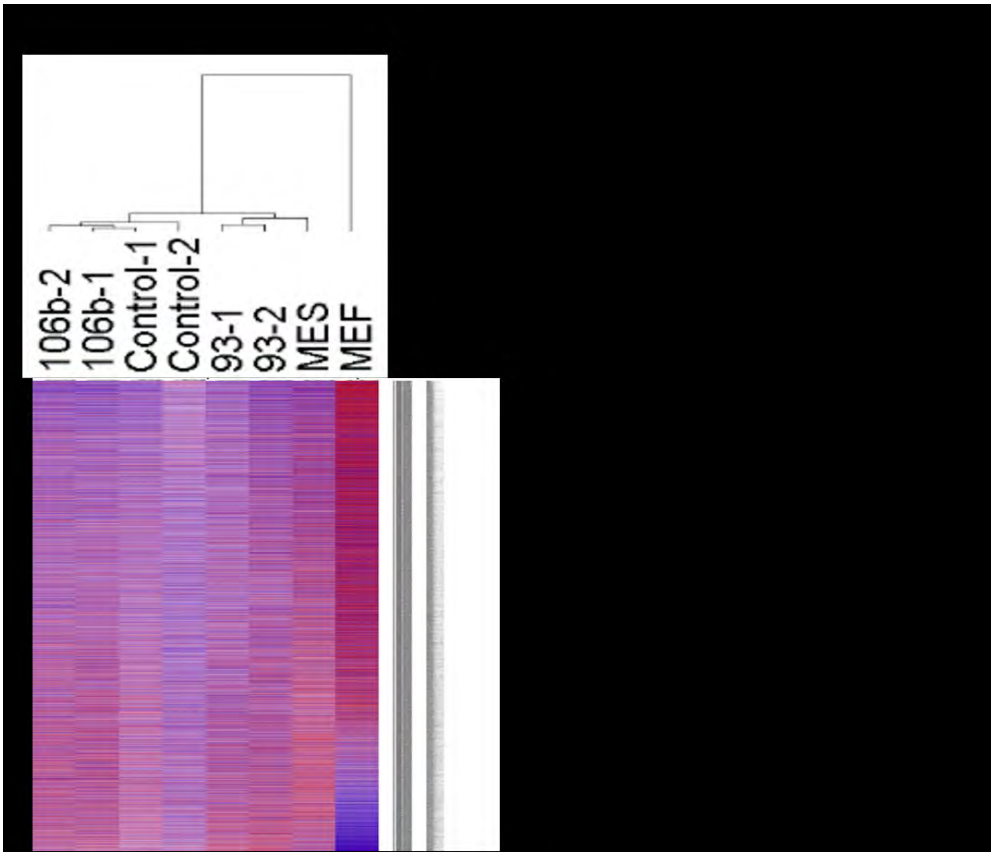


Figure 2.S11. miR-iPSCs contribute to the germline of derived embryos

Embryos were collected at E13.5 and fixed in 4% paraformaldehyde for O.N. and genital ridges were dissected out for fluorescence analysis. Oct4-GFP wild type embryos were used as the positive control.

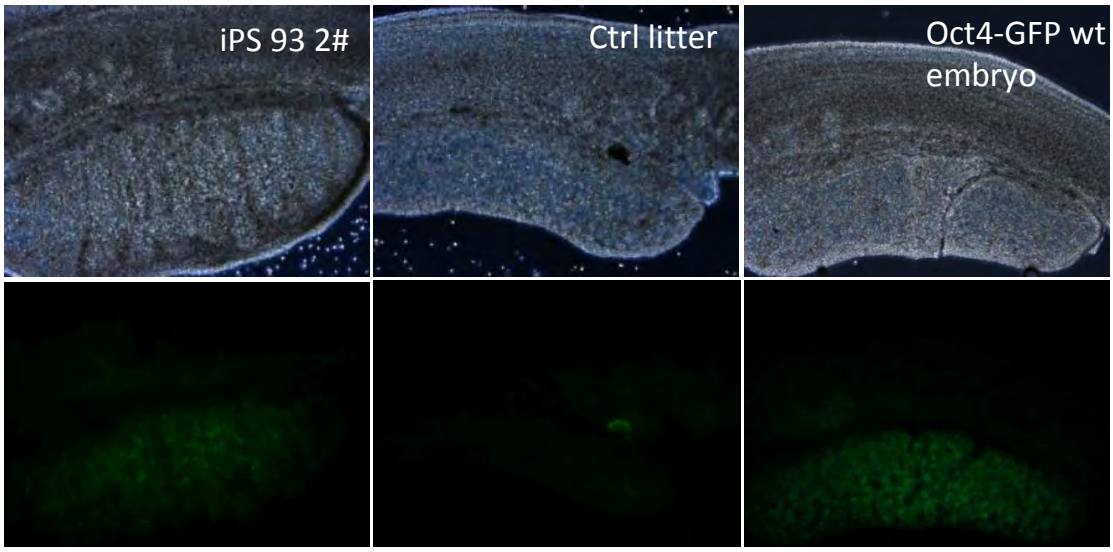
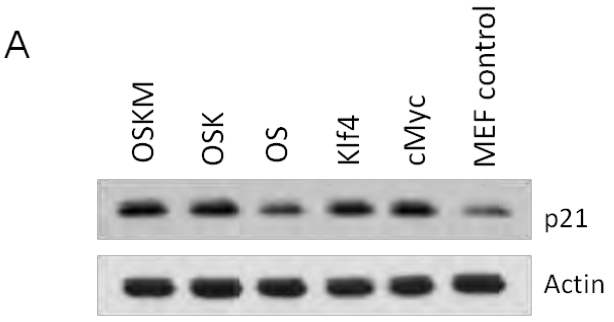


Figure 2.S12. p21 expression is induced during iPS induction

(A) p21 expression is induced by Klf4 and cMyc. MEFs infected with 4F, OSK, OS, Klf4 and cMyc were harvested at day 5 post-transduction for Western analysis. (B) Confirmation of transgene expression in MEFs. Total RNAs were isolated from transduced MEFs and analyzed by RT-qPCR for transgene expression.



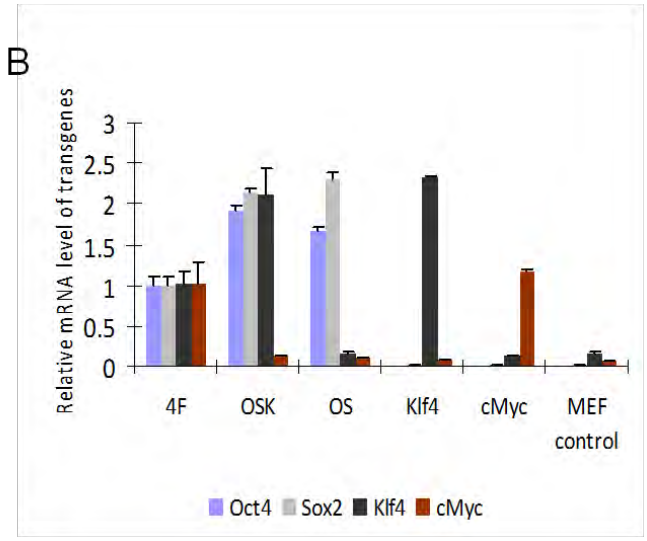
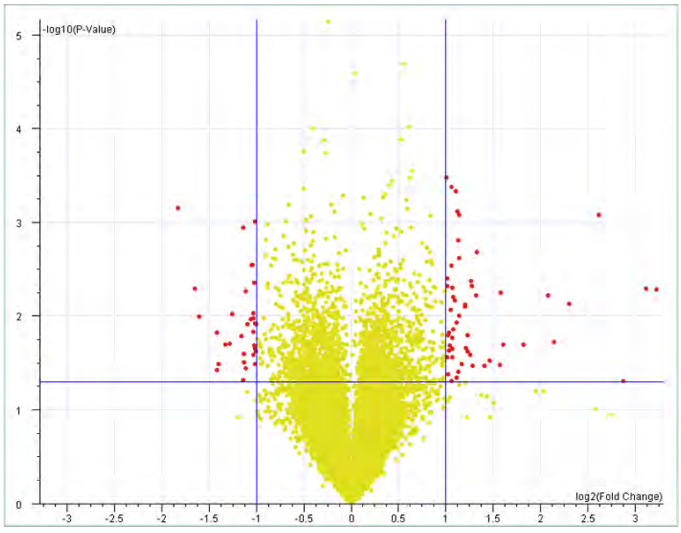


Figure 2.S13. Gene expression pattern analysis of miR-93-transfected MEFs

(A) mRNA microarray analysis identified genes showing significant mRNA level changes upon miR-93 transfection. A Volcano map was generated with a 2-fold expression change and p value <0.05. Red dots indicate the genes which are significantly changed in such settings (altered by more than two fold). (B) Genes significantly altered were divided into two groups (increased or decreased expression upon miR-93 introduction) and compared with published iPS/MEF expression profiles (Sridharan et al., 2009). —Correlated" indicates genes with expression changes similar to published iPS/MEF expression profile (Sridharan et al., 2009) (i.e,genes exhibiting decreased expression in miR-transfected MEFs show decreased expression in iPS cells vs MEFs).

A



B

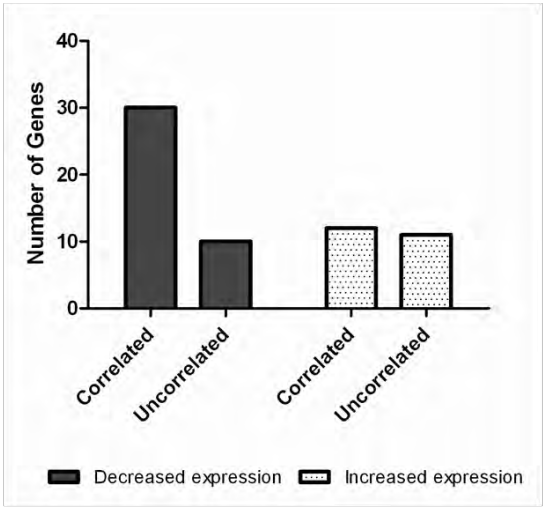
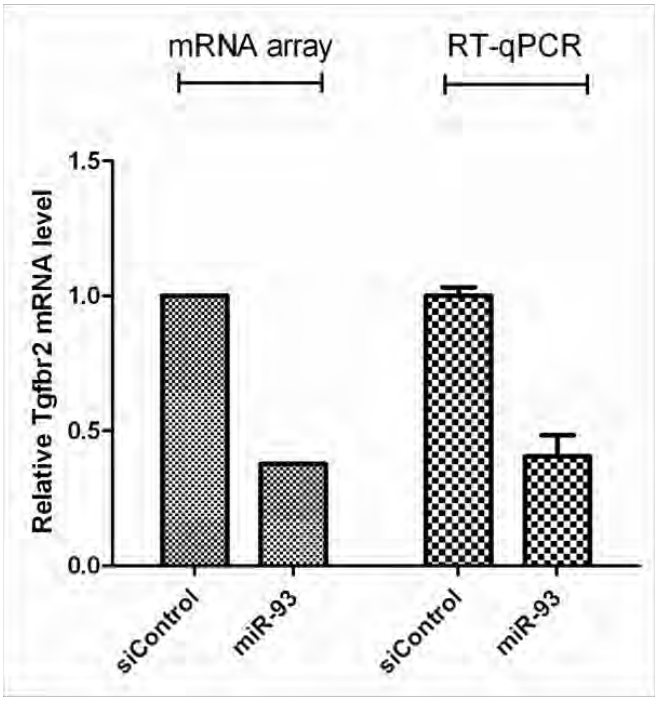


Figure 2.S14. Tgfbr2 and p21 mRNAs decrease upon miR transfection

(A) Global mRNA expression analysis indicates that Tgfbr2 mRNA levels decrease ~60%-70% upon miR-93 transfection, a finding confirmed by RT-qPCR. Error bars represent two independent experiments in duplicate wells. (B) p21 mRNA level is decreased ~25-30% by miR-93 and 106b. Total RNAs of transfected cells were harvested at 48hrs post-transfection. While miR-93 and 106b decreased the mRNA level of p21, miR-25 did not have any effect. Error bars represent three independent experiments using duplicate wells.

A



B

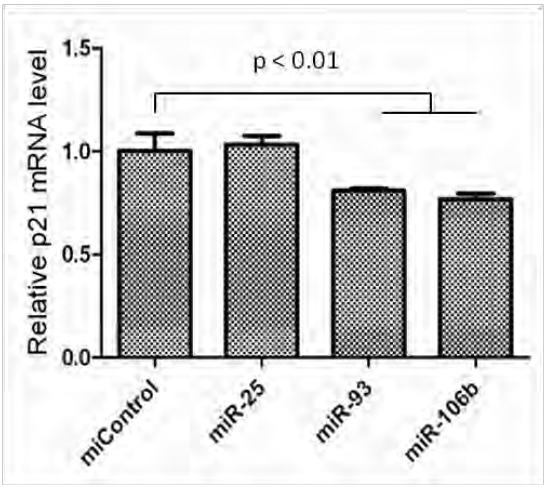
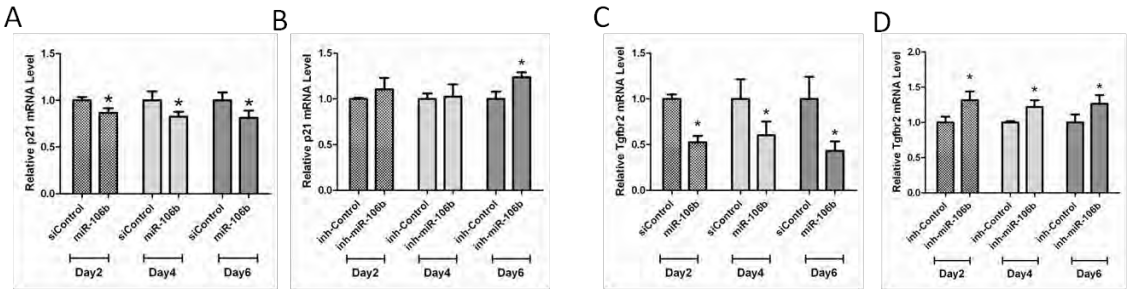


Figure 2.S15. p21 and Tgfbr2 expression is regulated by miR-106b in 4F transduced MEFs

(A) p21 mRNA level decreased upon miR-106b transfection. miR-106b was transfected into 4F transduced MEFs at day0 and day5 and cells were harvested at indicated time points. RT-qPCR was used to analyze the expression of p21 and data was normalized to Gapdh. Error bar represents standard deviation of three independent experiments. * $p<0.001$. (B) p21 mRNA increased upon miR-106b inhibition. miR-106b inhibitor was transfected the same way as miR mimics. Error bar represents standard deviation of three independent experiment. * $p<0.001$. (C) Tgfbr2 mRNA level decreased upon miR-106b transfection. The experimental procedures were the same as for p21 experiment. Error bar represents standard deviation of three independent experiments. * $p<0.001$. (D) Tgfbr2 mRNA increased upon miR-106b inhibition. The experimental procedures were the same as for p21 experiment. Error bar represents standard deviation of three independent experiments. *$p<0.001$. (E) P21 and TGFBR2 were regulated by miR-106b in 4F transduced MEFs. Both miR mimic and inhibitor were transfected into 4F infected MEFs. Cells were harvested at indicated time points for immunoblotting and analysis of P21 and TGFBR2 levels.



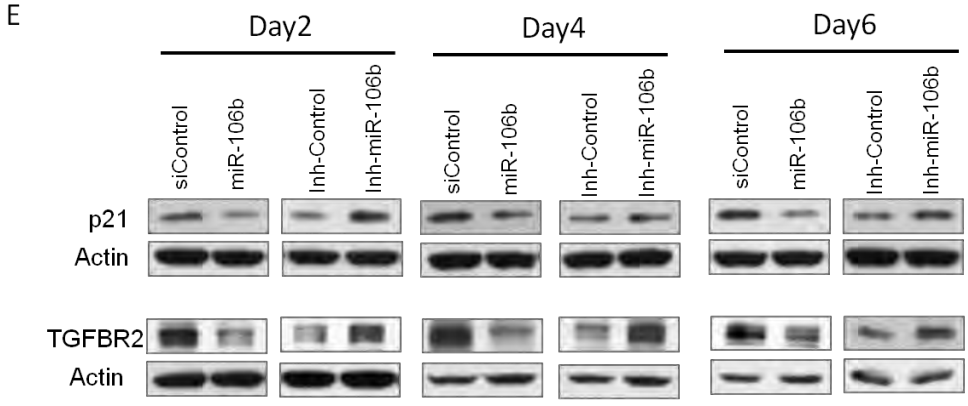


Figure 2.S16. p21 and Tgfbr2 expression is regulated by miR-106b in OSK infected MEFs

(A) p21 mRNA level decreased upon miR-106b transfection. The procedures were the same as 4F experiments. Error bar represents standard deviation of three independent experiments. *p<0.001. (B) p21 mRNA level decreased upon miR-106b transfection. Error bar represents standard deviation of three independent experiments. *p<0.001. (C) P21 and TGFBR2 proteins decreased upon miR-106b transfection.

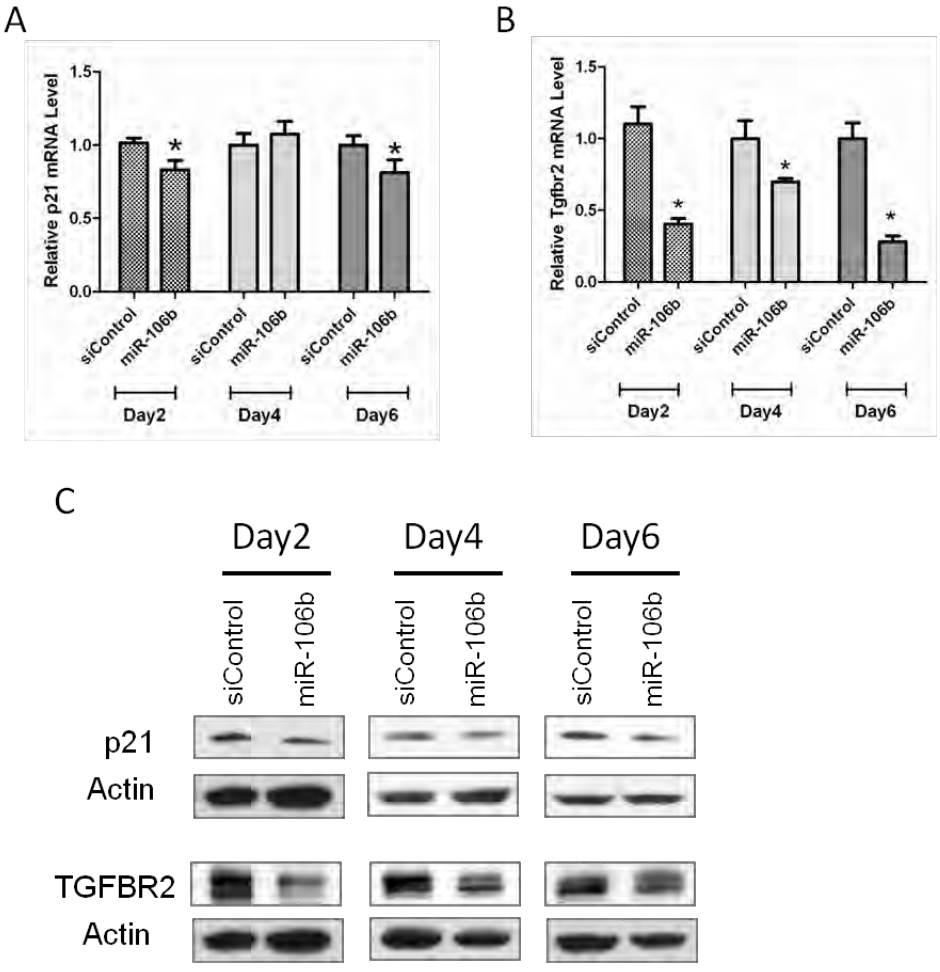


Figure 2.S17. p21 expression is directly regulated by the miR-106b~25 cluster

(A) (upper) Two potential miR-93 and 106b binding sites were found in the p21 mRNA 3'UTR. Blue box stands for the coding region of p21 mRNA. (lower) Mutations were introduced into the first conserved site to disrupt binding. (B) Quantification of pGL3-p21 luciferase reporter expression in Hela cells. Cells were transfected with the luciferase reporter pGL3-p21 and the renilla luciferase control vector (pRL-TK), as well as microRNAs for 48hrs before harvesting. Results were normalized to pRL-TK levels in transfected cells.

A

Site1:

wt: 5'- ...AATAGCACTTTG... -3'

mut: 5'- ...AATAGGTCATAG... -3'

B

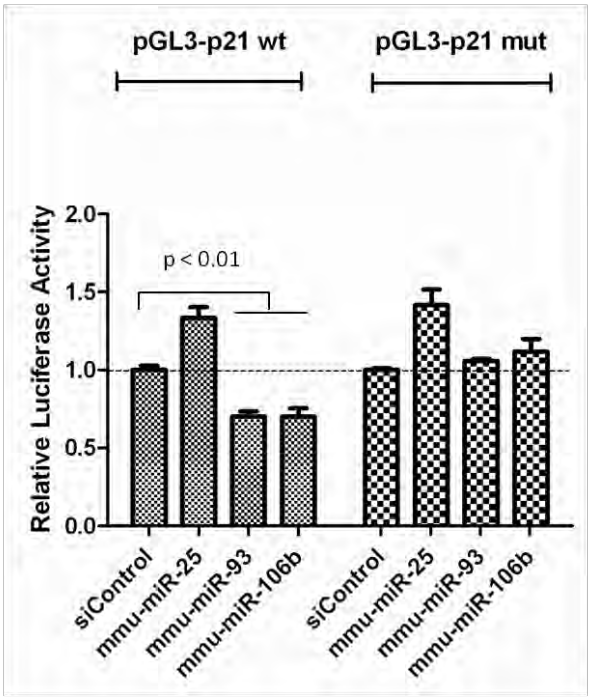


Figure 2.S18. TGFBR2 expression is directly regulated by the miR-106b~25 cluster

(A) Two potential miR-93 and 106b binding sites were found in the Tgfbr2 mRNA 3'UTR. Blue box stands for the coding region of Tgfbr2 mRNA. (B) Quantification of pGL3-Tgfbr2 luciferase reporter expression in Hela cells. Cells were transfected and harvested the same as p21 dual luciferase assay.

A

B

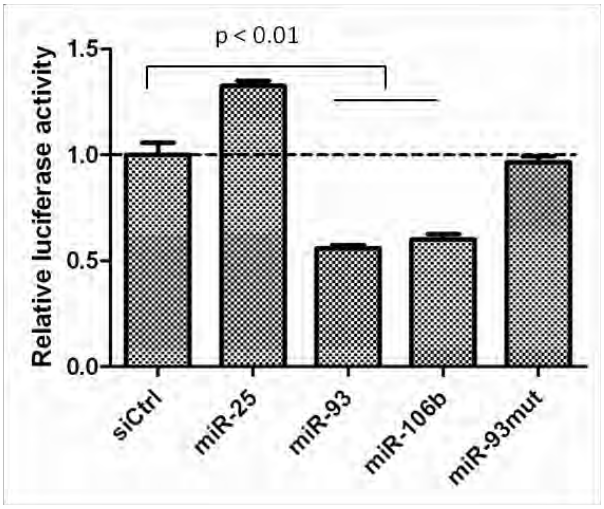


Figure 2.S19. Ectopic expression of p21 inhibits reprogramming

(A) p21 ectopic expression dramatically decreases the number of both GFP+ and AP+ colonies using a 4F transduction protocol. (left) AP+ and GFP+ colony quantification. (right) AP staining at day11 and western blotting for p21 overexpression. p21 virus was introduced at the same time with 4F, and induced cells were stained for alkaline phosphatase activity at day 11. (B) p21 ectopic expression dramatically decreases the number of both GFP+ and AP+ colonies following OSK transduction. (left) AP+ and GFP+ colony quantification. (right) AP staining of samples with OSK and p21 overexpression. p21 virus was introduced at the same time with OSK, and induced cells were stained for alkaline phosphatase activity at day 20.

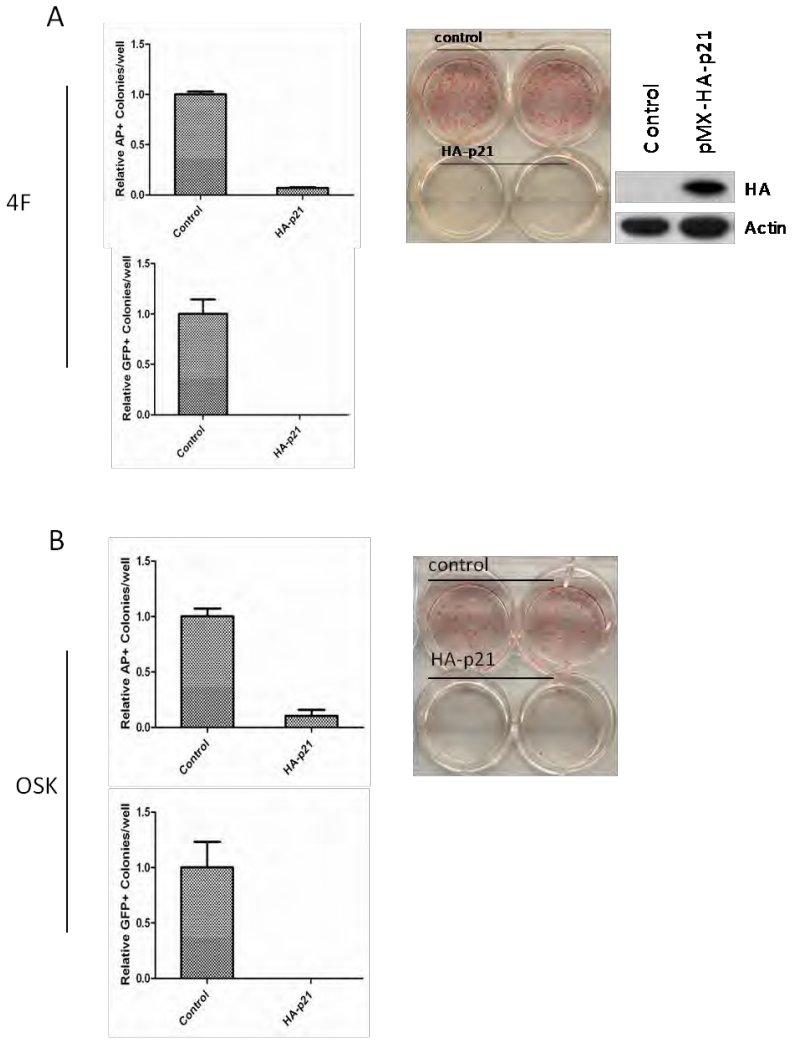


Figure 2.S20. Ectopic expression of Tgfbr2 inhibits reprogramming and compromises the enhancement of miR-106b transfection

(A) TGFBR2 was overexpressed in MEFs by pMX retroviral vector. MEFs were transduced with 4F and pMX-TGFBR2 and harvested for western blotting analysis of TGFBR2. (B) Overexpression of TGFBR2 decreased reprogramming efficiency and compromised miR-106b enhancing effect. Both 4F and 4F+TGFBR2 transduced MEFs were transfected with miR-106b and Oct4-GFP+ colonies were quantified at day12. miR-106b can enhance reprogramming by three fold in 4F only cells while only increases colony number by ~1.5 fold in TGFBR2 overexpressing cells. Error bar represents standard deviation of three separate wells.

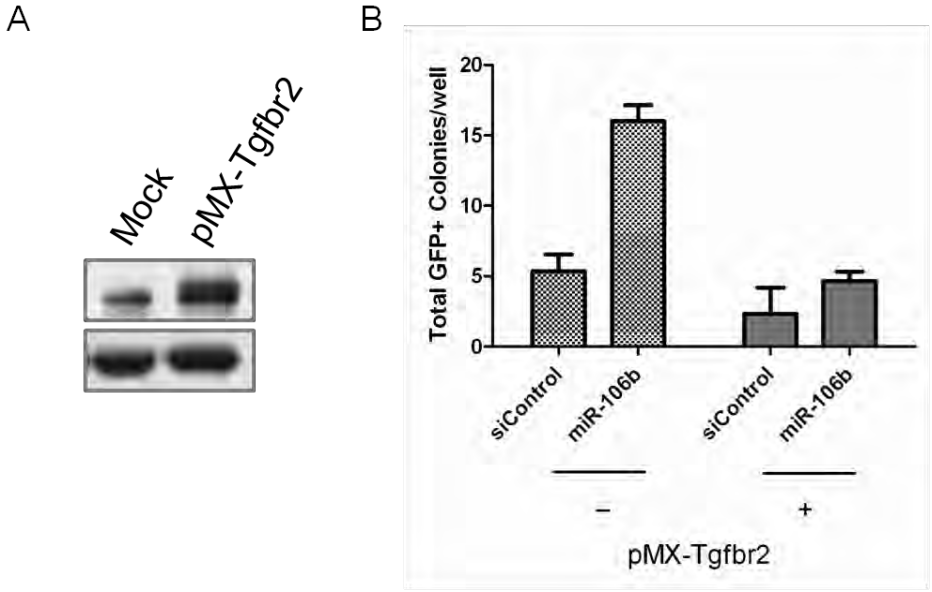


Figure 2.S21. Tgfbr2 is targeted by miR-93 and its family microRNAs

miR mimics were transfected into MEFs and total RNAs were extracted at day2 post transfection. RT-qPCR was used to quantify the relative mRNA level of Tgfbr2 in samples treated with different microRNAs. Only miR-93 and its family microRNAs showed efficient decrease of Tgfbr2 mRNA while other unrelated miRs did not have any effect.

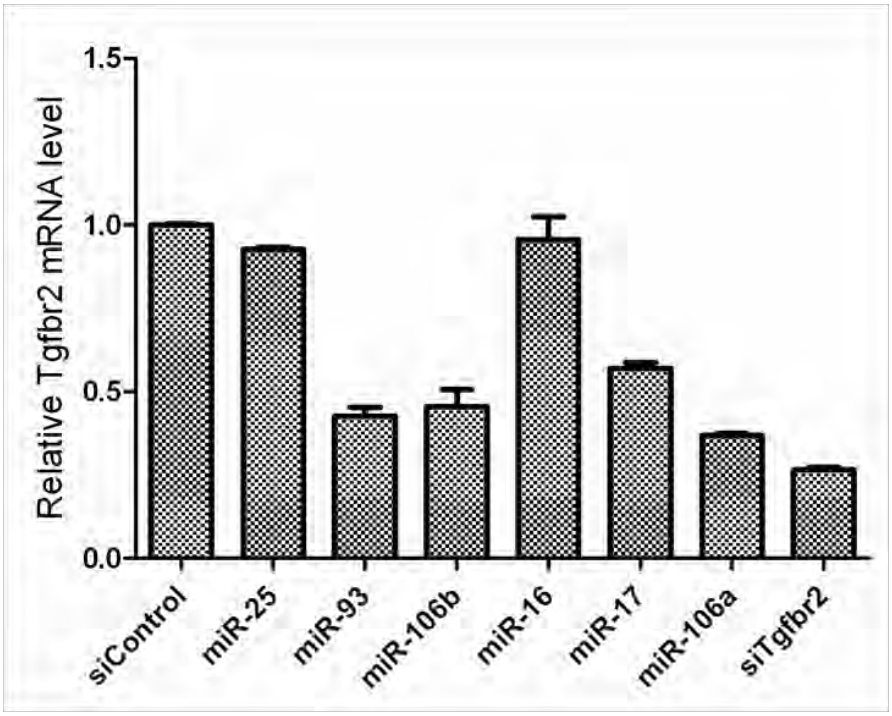
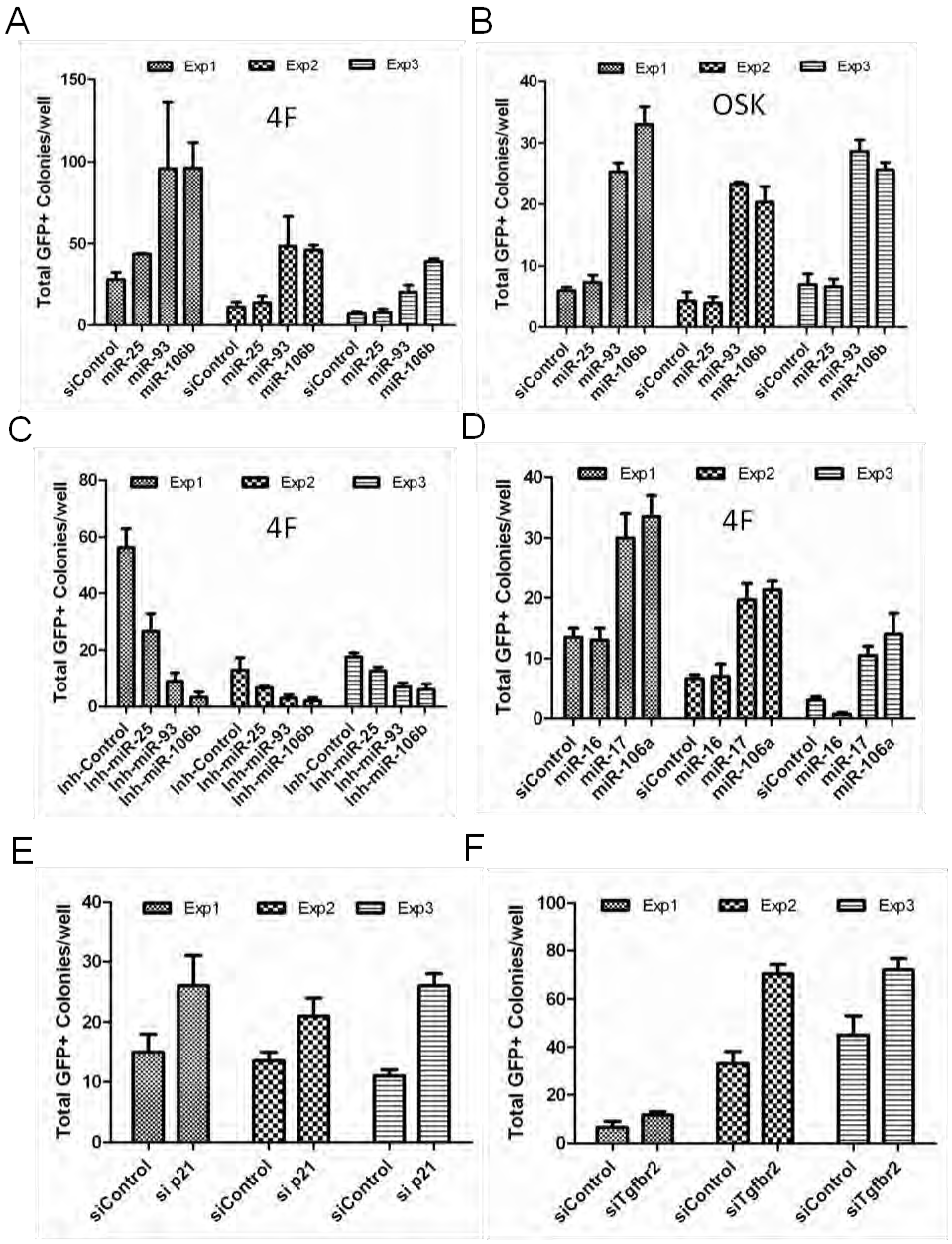


Figure 2.S22. Absolute Oct4-GFP+Colony Quantification

(A) Three representative experiments for Figure 3c. (B) Three representative experiments for Figure 3d. (C) Three representative experiments for Figure 3e. (D) Three representative experiments for Figure 6a. (E) Three representative experiments for Figure 5c. (F) Three representative experiments for Figure 5f.



CHAPTER III: MicroRNA-mediated Regulation of Extracellular Matrix Formation Modulates Somatic Cell Reprogramming

3.1. Abstract

Somatic cells can be reprogrammed to reach an embryonic stem cell-like state by overexpression of defined factors (Takahashi et al, 2007; Takahashi & Yamanaka, 2006; Wernig et al, 2007; Yu et al, 2007). The current reprogramming process is extremely inefficient, suggesting the need for a better understanding of the underlying mechanisms to develop new reprogramming methods and to understand the transition to a pluripotent state. MicroRNAs (miRs) are small non-coding RNAs that primarily regulate target gene expression post-transcriptionally. Here we present a systematic and comprehensive study of microRNAs in mouse embryonic fibroblasts (MEFs) during the early stage of cell fate decisions and reprogramming to a pluripotent state. One microRNA found to be highly induced during reprogramming, miR-135b, targeted the expression of extracellular matrix (ECM) genes including Wisp1 and Igfbp5. Wisp1 was shown to be a key regulator of additional ECM genes that serve as barriers to reprogramming. Regulation of Wisp 1 is likely mediated through biglycan, a glycoprotein highly expressed in MEFs that is silenced in reprogrammed cells. Collectively, this is the first report, to the best of our knowledge, revealing a novel link between microRNA-mediated regulation of ECM formation and somatic cell reprogramming, and demonstrate that microRNAs are powerful tools to dissect the molecular mechanisms of reprogramming.

3.2. Introduction

Since the first report that mouse fibroblasts can be reprogrammed into a pluripotent state reminiscent of embryonic stem cells (termed induced pluripotent stem cells, iPSC)(Takahashi & Yamanaka, 2006), this phenomenon has been confirmed using many different mouse and human cell types (Lowry et al, 2008; Nakagawa et al, 2008; Park et al, 2008b; Takahashi et al, 2007; Wernig et al, 2007; Yu et al, 2007). Currently, the main obstacle for reprogramming to iPSCs is its extremely low efficiency; typically only 0.01%–0.2% of starting cells are successfully reprogrammed into iPSCs (Aoi et al, 2008; Meissner et al, 2007; Nakagawa et al, 2008; Takahashi & Yamanaka, 2006). Great effort has been made to identify small molecules that enhance the reprogramming process or that replace one or more of the four transgenes (Oct3/4, Sox2, Klf4, and c-Myc; OSKM, 4F) commonly used in the reprogramming protocol(Esteban et al, 2010; Ichida et al, 2009; Li et al, 2009b; Lyssiotis et al, 2009; Maherali & Hochedlinger, 2009; Shi et al, 2008a; Shi et al, 2008b). However, the molecular mechanisms by which the four factors are able to reprogram somatic cells remain largely unknown.

Mounting evidences from recent researches have suggested that somatic reprogramming could be a complicate process involving many different processes. Systematic analysis of the promoters targeted by overexpression of the four reprogramming factors has demonstrated that expression of the factor

target genes is similar in iPSCs and mouse embryonic stem (mES) cells, and is altered in some partially reprogrammed cells (Sridharan et al, 2009). p53 pathway is also identified as one primary barrier to reprogramming (Banito et al, 2009; Hong et al, 2009; Kawamura et al, 2009; Li et al, 2009a; Utikal et al, 2009). Chemical screening has also discovered that inhibition of TGFβ signaling significantly enhances reprogramming(Ichida et al, 2009) and that some inhibitors of this pathway can replace the Sox2 transgene in inducing expression of Nanog, a transcription factor crucial for ESC pluripotency(Maherali & Hochedlinger, 2009). In addition, it was also suggested that a mesenchymal-to-epithelial transition (MET) is a key step that takes place at an early stage of reprogramming(Li et al, 2010; Samavarchi-Tehrani et al, 2010). During reprogramming, expression of markers on the initial somatic cell, such as mouse embryonic fibroblasts (MEFs), are downregulated and characteristic mES markers, such as alkaline phosphatase, SSEA1, Nanog, and endogenous Oct4 become expressed(Brambrink et al, 2008; Stadtfeld et al, 2008). Interestingly, the cellular origin of the iPSCs apparently influences their ability to retain an epigenetic —memory” of the originating cell, a property that is gradually lost through continuous passaging of iPSCs(Polo et al, 2010). However, iPSCs do not seem to have a generic epigenetic state that could clearly define fully reprogrammed state(Carey et al, 2011). Despite these progresses, there remains only limited information on the mechanisms by which the four transgenes and other cellular factors reprogram MEFs to an undifferentiated or ES-like state.

Extracellular matrix (ECM) is a multifunctional system that is involved in many stages of mammalian developments (Adams & Watt, 1993; Rozario & DeSimone, 2010; Sanes, 1989) and human disease progressions, including tumor formation (Bissell & Hines, 2011; Kessenbrock et al, 2010). ECM encodes a variety of proteins which could be divided into two groups: proteins with structural role, such as fibrous proteins and glycosaminoglycans, and proteins with regulatory role, including different growth factors (TGFβ, IGFs etc), matricellular proteins (CCN family proteins, IGFBPs, decorin, biglycan etc), enzymes (metalloproteinases etc) and receptors (integrins etc). ECM plays a crucial role in regulating various cellular behaviors and maintaining the identity and normal function of those cells(Bissell & Hines, 2011; Kessenbrock et al, 2010). For embryonic stem cells, recent discoveries have found that ECM components are essential for establishing the proper niche for long term ES cell survival and self-renewal(Bendall et al, 2007; Peerani et al, 2007). Therefore, it is possible that ECM is also involved in regulating somatic reprogramming process. In fact, given the dramatic changes of both cellular morphology and functional characteristics during course of reprogramming, potential iPSCs would need to establish their own niche for supporting their growth and colony formation. Meanwhile, iPSCs also need to exclude the effects brought by secreted ECM proteins from surrounding unreprogrammed cells. However, despite that iPSCs expressed a different set of ECM proteins from starting fibroblasts cells (ref), little is known about the dynamic remodeling of ECMs during reprogramming and

studying it could yield considerate novel information regarding the molecular mechanism of this process.

MicroRNAs are 18–24 nucleotide long, single-stranded RNAs associated with a protein complex termed the RNA-induced silencing complex (RISC). Small RNAs are usually generated from non-coding regions of gene transcripts and function to suppress gene expression by translational repression and mRNA destabilization (Ambros, 2004; Bartel, 2004; Guo et al, 2010; Kim et al, 2009a; Rana, 2007). Individual microRNAs target a relatively limited set of genes, suggesting they could be used as tools to modulate expression of distinct subsets of genes and determine their involvement in the molecular mechanism of reprogramming. Recent work indicates that ES-specific microRNAs can enhance iPSC induction(Judson et al, 2009) and, specifically, that the hES miR-302 can antagonize the senescence response induced by four-factor expression in human fibroblasts (Banito et al, 2009). In addition, our recent findings suggest that the microRNA biogenesis machinery may be required for efficient reprogramming (Li et al, 2011), and microRNAs induced by OSKM are known to regulate several key pathways affecting reprogramming efficiency, including cell cycle control, the p53 pathway, TGFβ signaling, and MET (Choi et al, 2011; Li et al, 2011; Liao et al, 2011; Subramanyam et al, 2011; Yang et al, 2011). Importantly, expression of microRNAs alone can fully reprogram fibroblasts to iPSCs (Anokye-Danso et al, 2011; Miyoshi et al, 2011). These findings clearly suggest that microRNAs play crucial roles during the reprogramming process by

targeting key barrier signaling networks. However, most studies to date have focused on intracellular signaling networks regulated by microRNAs, and the ability of microRNAs to influence critical cellular interactions with the microenvironmental niche during reprogramming has not yet been investigated.

Here, we performed a systematic analysis of expression of microRNAs and their potential target genes at an early stage of reprogramming, and identified a novel link between ECM formation and reprogramming of MEFs. In particular, we found that microRNA-135b is induced to a very high level and modulating its expression significantly affected the reprogramming process. Using genome-wide mRNA array analysis, we show that miR-135b controls expression of Tgfbr2, Igfbp5, and Wisp1, the latter two genes encoding components of the MEF ECM. Wisp1 was found to regulate the secretion of several ECM proteins including TGFBI (TGF-beta induced), IGFBP5 (insulin-like growth factor binding proteins-5), NOV (nephroblastoma overexpressed gene), and DKK2 (dickkopf homolog 2) proteins. Interestingly, the effects of Wisp 1 are mediated through biglycan, a glycoprotein that is highly expressed in MEFs and is incompletely silenced in reprogramming cells. Notably, knockdown or overexpression of biglycan enhanced or suppressed MEF reprogramming, respectively. Collectively, our results have identified a novel role for microRNA-mediated regulation of ECM formation in iPSC generation, and further, demonstrate that microRNAs can be powerful tools to dissect and understand the molecular mechanisms of somatic reprogramming.

3.3. Results

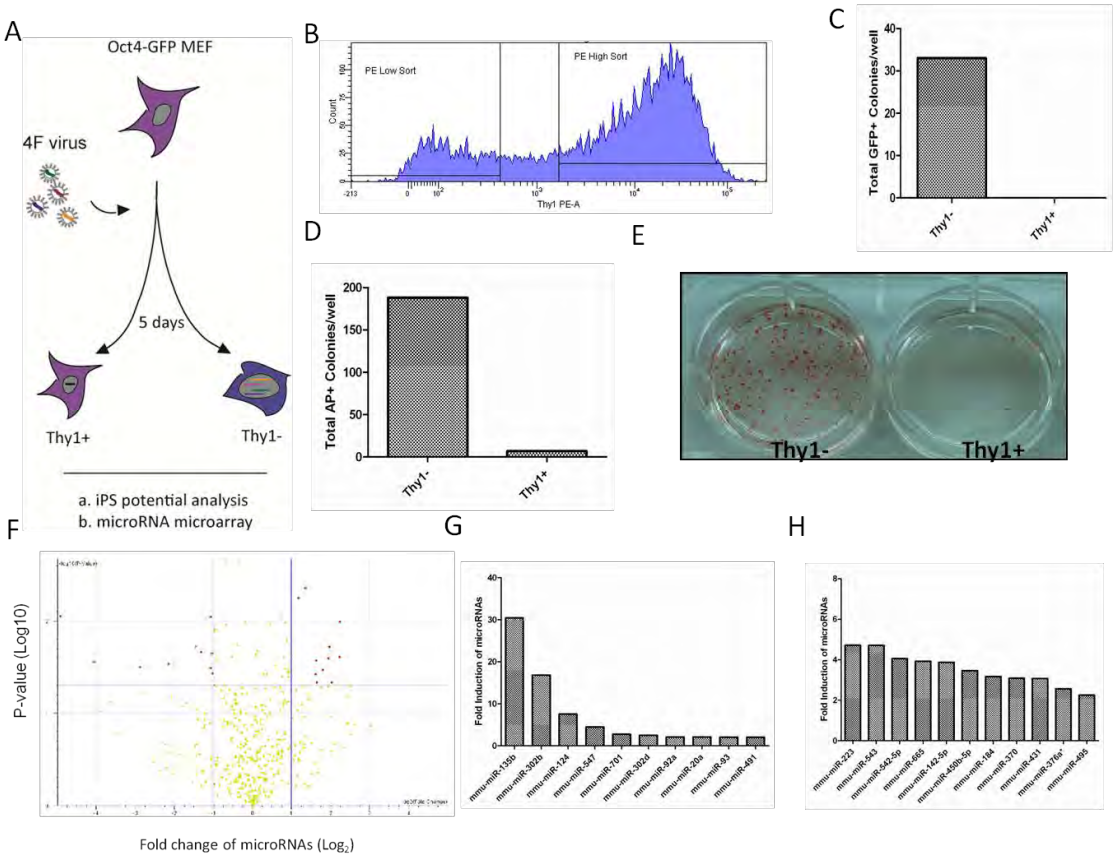
3.3.1. Systematic identification of highly regulated microRNAs during the early stages of reprogramming

We hypothesized that at different reprogramming stages, potential iPSCs may express unique _marker signatures' of microRNAs that regulate how the cells reach a fully reprogrammed stage. Previous findings indicate that reprogramming of MEFs is accompanied by sequential modulation of somatic cell and stem cell markers at different reprogramming stages (Brambrink et al, 2008; Stadtfeld et al, 2008), which can be used to track the process. These markers include the cell surface antigen Thy1, the mES markers alkaline phosphatase (AP) and SSEA1, and the self-renewal genes Nanog and Oct4. Thy1 is highly expressed in MEFs but its expression is repressed at the initiation of reprogramming. Conversely, AP and SSEA1 expression is upregulated, followed by upregulation of Nanog and endogenous Oct4. Thus, MEFs expressing GFP under control of Oct4 are often used as the starting somatic cells because GFP expression then identifies cells that have been fully reprogrammed to the iPSC stage. To identify key microRNAs in reprogramming, we focused on the early reprogramming stage in the first 5 days after transduction of MEFs with the four factors (4F; OSKM). To determine whether the fate of 4F-transduced cells is set at that stage, Oct4-GFP MEFs were infected with 4F virus and then harvested five days later for cell sorting (Figure 3.1.A). PE-conjugated Thy1 antibody was used to isolate pure Thy1+ and

Figure 3.1. Identification of highly regulated microRNAs during the early reprogramming stage

(A) Scheme of experimental design. MEFs were infected with 4F virus for 5 days, and sorted based on expression of the Thy1 surface antigen. Both $Thy1^-$ and Thy1+ cells were collected for microRNA expression profile analysis. (B) Representative gating for day 5 4F-infected MEF sorting. PE-conjugated Thy1 antibody was used to detect Thy1- and Thy1+ populations. (C) iPSCs were enriched in the Thy1- population of 4F-infected MEFs at day 5. Equal numbers of cells (10,000 cells) sorted from 4F-infected MEFs were replated into feeder plates and cultured for 14 days, then GFP^+ colonies were counted. (D) AP staining confirmed that iPSCs generated in (C) were enriched in the Thy1- population. Cells were harvested for AP staining at day 14 post-infection. (E) Representative image of AP+ colonies from replated Thy1- and Thy1+ cells. (F) Induced or repressed microRNAs were identified in Thy1- cells. Both Thy1- and Thy1+ cells were harvested for microRNA expression profiling. Data from the $Thy1^-$ population was compared with the original MEFs and microRNAs showing a 2-fold change and $p<0.05$ were identified using a volcano map. Hits are labeled as red dots. (G) Set of significantly induced microRNAs. MicroRNAs induced by at least 2-fold are shown. (H) Set of significantly repressed microRNAs. MicroRNAs repressed by at least 2-fold are shown.

Figure 3.1. Identification of highly regulated microRNAs during the early reprogramming stage



Thy1- populations, with gates set to exclude cells expressing intermediate Thy1 levels (Figure 3.1.B). Equal numbers (10,000 cells) of Thy1+ and Thy1$^-$ cells were reseeded in 12-well plates on CF1-MEF feeders and their potential for iPSC induction was evaluated based on GFP and marker expression. Potential iPSCs were enriched mainly in the Thy1- population, as determined by counting of colonies expressing GFP or AP (Figure 3.1.C, D). We detected no GFP$^+$ colonies and only a few AP$^+$ colonies in the Thy1+ population at day14 post 4F infection (Figure 3.1.C, E). These results suggest that the fate of 4F-infected MEFs is determined before day 5 post-infection and that potential iPSCs are enriched in the Thy1- population. We therefore collected total RNA from sorted Thy1- cells at day 5 post-transduction to analyze overall microRNA expression changes by microarray. To identify microRNAs whose expression is significantly altered relative to that seen in starting MEFs, we filtered the data by setting a gate of at least a 2-fold change in expression with $p<0.05$ (Figure 3.1.F). We identified a set of microRNAs in the Thy1- $^-$ population that were significantly induced by 4F transduction (Figure 3.1.G). Among them, miR-135b was the most highly induced and showed a statistically significant change in expression (Table 3.2.), and was thus selected for further analysis of its role, and that of its direct gene targets, in the reprogramming process. We observed that other microRNAs, such as miR-93 which belongs to miR-25~106b cluster, miR-92a which belongs to miR-17~92 cluster, and miR-302b which belongs miR-302 cluster, were also highly induced at the early stage of reprogramming, confirming previous findings(Li et al, 2011;

Liao et al, 2011; Subramanyam et al, 2011). Our analysis also revealed a set of microRNAs that were significantly repressed (Figure 3.1.H), suggesting that they may serve as reprogramming barriers. Of these, we chose to evaluate the potential barrier function of miR-223 and miR-495, because they are highly expressed in MEFs.

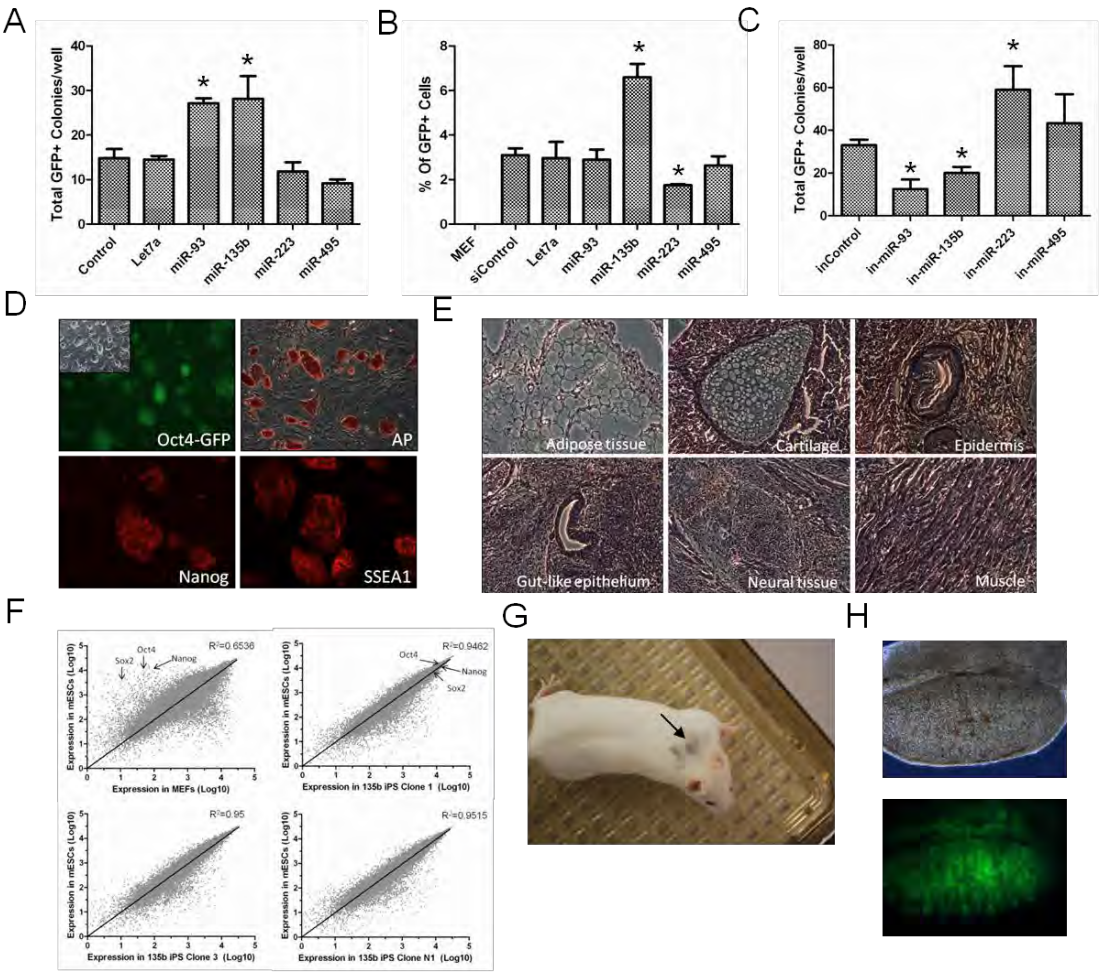
3.3.2. Reprogramming is enhanced by miR-135b and inhibited by miR-223 and miR-495

To determine how miR-135b affects reprogramming, miR-135b microRNA mimic was transfected into Oct4-GFP MEFs infected with 4F virus, and GFP^+ colonies were counted at day 11–12 post-transduction. Transfection of the miR-135b mimic increased the number of Oct4-GFP^+ colonies by ~ 2-fold, as did transfection with miR-93, which was previously characterized as an enhancer of reprogramming (Anokye-Danso et al, 2011) (Figure 3.2.A). In similar experiments, cells were transfected with miR-223 or miR-495 mimics, which had minor inhibitory effects on reprogramming (Figure 3.2.A). This observation is potentially due to the saturation effect of endogenous miRs as these miRs already have high expression in MEFs. We then analyzed the percentage of GFP^+ cells in the miR-transfected reprogrammed cells and found that although both miR-93 and miR-135b increased GFP^+ colony formation, only miR-135b increased the overall percentage of GFP+ cells by ~2 fold (Figure 3.2.B, Figure 3.S1.). In the same assay, miR-223 transfection significantly decreased the GFP+ population (Figure

Figure 3.2. miR-135b enhances reprogramming of MEFs to iPSCs

(A) miR-135b enhances Oct4-GFP+ colony formation. The indicated microRNA mimics were transfected at a final concentration of 50 nM into MEFs on day 0 and again on day 5 after 4F transduction. GFP+ colonies were counted at day 11-12. Data represents two independent experiments with triplicate wells. Let-7a was used as a control. *$p<0.05$. (B) miR-135b increases the percentage of Oct4-GFP+ cells. Cells from the indicated treatments were harvested at day 14 post-infection with 4F and paraformaldehyde-fixed prior to FACS analysis to determine the percentage of GFP+ cells. Data represents two independent experiments with triplicate wells. *$p<0.05$. (C) Blocking of miR-135b compromises reprogramming. MicroRNA inhibitors were transfected into MEFs on days 0 and 5 post-infection with 4F. GFP+ colonies were counted at day 11-12 post-infection. Data represents two independent experiments with triplicate wells. *$p<0.05$. (D) miR-135b iPSCs reach a fully reprogrammed state. miR-135b–transfected iPSCs were fixed with paraformaldehyde and stained for alkaline phosphatase, Nanog, and SSEA1 expression. Endogenous Oct4 expression was monitored by GFP expression. (E) Teratoma formation confirms the pluripotency of miR-135b iPSCs. $1x10^6$ iPSCs were injected into athymic nude mice and tumors were harvested for H&E staining 3–4 weeks later. (F) miR-135b iPSCs show expression profiles similar to mES cells. Total RNA from miR-135b iPSCs was used for mRNA expression profile analysis and compared with original MEFs and with mES cells. The three tested miR-135b iPSC clones (clones 1, 3, and N1) showed similar expression patterns to mES cells, which were quite different from the expression profile of the original starting MEFs. (G) Chimeric mouse from miR-135b iPSC clone 4. (H) miR-135b iPSC could contribute to the germline of recipient embryos (miR-135b iPSC clone 4)

Figure 3.2. miR-135b enhances reprogramming of MEFs to iPSCs



3.2.B), supporting the possibility that it serves as a reprogramming barrier. To confirm our findings, we used microRNA inhibitors. As expected, blocking miR-135b compromised reprogramming efficiency, while inhibiting miR-223 resulted in a significant increase in the number of Oct4-GFP+ colonies (Figure 3.2.C). Overall, these data demonstrate that miR-135b enhances reprogramming, consistent with its high induction by the 4F factors, while miR-223, which our analysis showed to be the most highly repressed microRNA, serves as a barrier.

Because GFP expression by putative iPSC could result from inappropriate reactivation of the Oct4 locus, we asked whether miR-135b–transfected iPSCs reached a fully reprogrammed state, both phenotypically and functionally. Analysis of miR-135b–transfected iPSCs indicated that they expressed appropriate markers, including AP, SSEA1, Nanog, and endogenous Oct4 (Figure 3.2.D). Moreover, these cells had the full capacity to differentiate into three germ layers as indicated by marker analysis (Figure 3.S2.), and to form heterogeneous teratomas when injected into athymic nude mice (Figure 3.2.E). Genome-wide mRNA profiling also confirmed that gene expression in miR-135b–transfected iPSCs resembled mES cells and differed significantly from MEFs (Figure 3.2.F), and these cells contributed to chimeric mice and showed germline transmission (Figure 3.2.G, H) which clearly indicated that a fully reprogrammed state has been achieved in these cells. These data demonstrated that miR-135b transfection in iPSCs did not adversely affect their pluripotency.

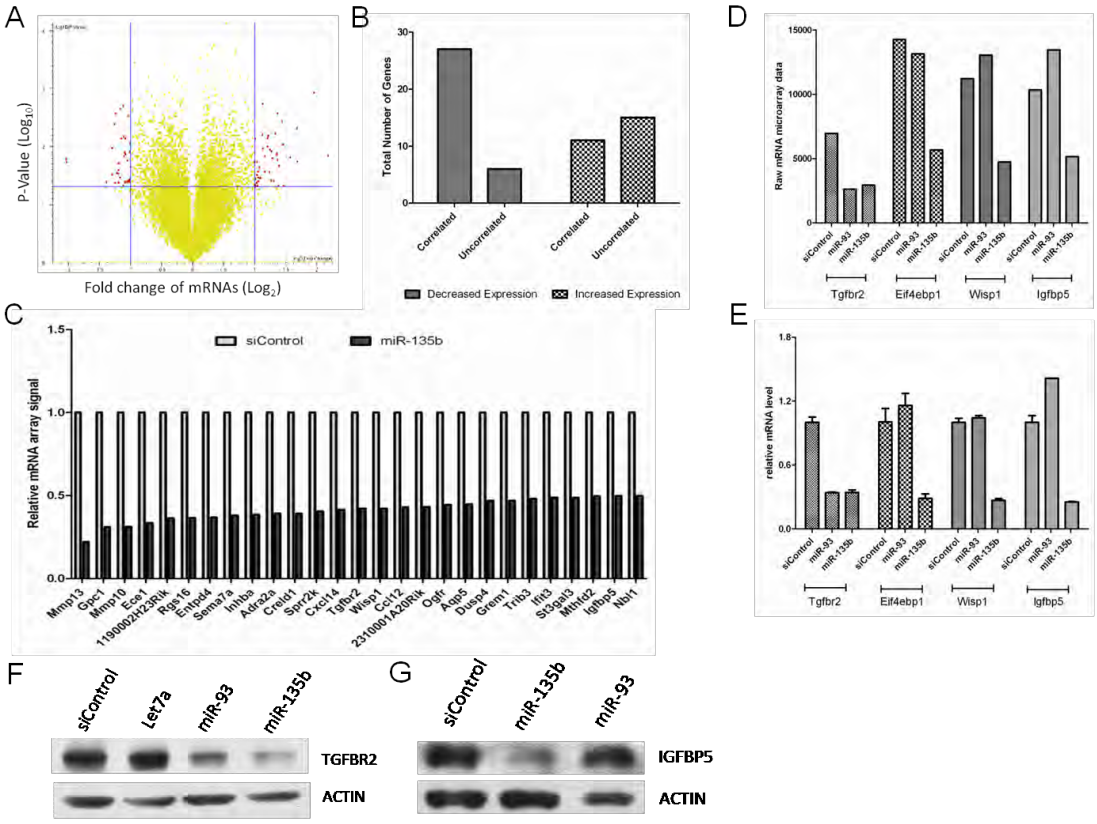
3.3.3. Identification of miR-135b–regulated genes

We next sought to identify genes that are directly regulated by miR-135b. Initially, microRNAs were thought to simply repress mRNA translation. However, recent findings suggest that microRNA-mediated destabilization of mRNA is a major mechanism of repression(Guo et al, 2010). Thus, we performed a genome-wide mRNA expression analysis to detect potential miR-135b targets. miR-135b or control siRNA were transfected into Oct4-GFP MEFs, and total RNAs were harvested 48 hr later for array analysis. The raw data was filtered to detect at least 2-fold changes in gene expression, (either increased or decreased) with $p<0.05$ (Figure 3.3.A). Candidate genes were then compared with published mESC, iPSC, and MEF expression profiles (Sridharan et al, 2009) and segregated into genes induced (group 1) or repressed (group 2) after miR-135b transfection, the latter being considered more likely to contain direct targets. Notably, we found that over 80% of the genes repressed by miR-135b transfection (group 2) were genes that are silenced as MEFs are reprogrammed to iPS/mES cells (correlated) (Figure 3.3.B). This was not observed in genes that were induced by miR-135b transfection (group 1), of which approximately half are normally suppressed during reprogramming (uncorrelated), and the other half are increased (correlated). This data suggests that miR-135b targets a subset of genes that are normally repressed during reprogramming.

Figure 3.3. Genome-wide identification of potential miR-135b target genes

(A) Volcano maps from miR-135b–transfected MEFs. MEFs were transfected with siControl and miR-135b for two days and analyzed by mRNA expression array. Hits (red dots) were gated for at least 2-fold expression change and $p<0.05$. (B) miR-135b–repressed genes are enriched for genes suppressed in ES/iPS cells. miR-135b–regulated genes were separated into two groups (induced or repressed) and then compared with existing iPS/ES/MEF expression profiles. "Correlated genes" indicates that genes changed upon miR-135b transfection showed similar changes from MEFs to iPS/mES cells. "Uncorrelated genes" indicates a group of genes that were changed upon miR-135b transfection but had a different (reversed) change in expression pattern from MEFs to iPS/mES cells. (C) List of correlated miR-135b–repressed genes. (D) Representative miR-135b–regulated genes from microarray. (E) Expression of miR-135b–regulated genes was confirmed by RT-qPCR. MEFs were transfected with microRNA mimics for two days before harvesting for RT-qPCR analysis. Error bar represents two independent experiments with duplicate samples. (F) TGFBR2 protein expression is suppressed by miR-93 and miR-135b. Total proteins were harvested for western blotting analysis at day 2 post-transfection with miR mimic. (G) IGFBP5 protein expression is suppressed by miR-135b. A miR-93-transfected sample was included as a negative control.

Figure 3.3. Genome-wide identification of potential miR-135b target genes.



To identify the targets of miR-135b, the —correlated" genes in group 1 (Figure 3.3.C) were analyzed using both miRanda (Enright et al, 2003) and Targetscan (Lewis et al, 2005). Potential target sites were identified based on seed region matches and overall predicted binding energy. Of 27 genes repressed by miR-135b by at least 2-fold, 14 contained at least one predicted miR-135b target site (Figure 3.S3.A and Table 3.1). Among them, Wisp1, Tgfbr2, and Igfbp5 showed high expression intensity detected by microarray and appeared to have direct miR-135b target sites. Therefore, they were chosen for further validation.

To confirm our mRNA microarray analysis, total RNAs were harvested from miR-135b–transfected Oct4-GFP MEFs in an independent experiment, and RT-qPCR was used to quantify the representative mRNAs. Indeed, we detected decreases in mRNA levels upon miR-135b transfection that were in good agreement with the mRNA array data (Figure 3.3.D, E). Tgfbr2 and Igfbp5 mRNA levels were decreased ~70% upon miR-135b transfection, and western analysis confirmed that this was accompanied by a dramatic decrease in Tgfbr2 and Igfbp5 protein expression (Figure 3.3.F, G). Although expression of Wisp1 mRNA was also markedly reduced by miR-315b expression (Figure 3.3.D, E), no Wisp1 antibodies are currently available, which prevented us from analyzing Wisp1 protein expression. We cloned the 3‘UTR of these potential targets into the pGL3 luciferase reporter vector and co-transfected the reporters plus the pRL-TK plasmid into HeLa cells. Indeed, miR-135b decreased luciferase activity of Tgfbr2

and Wisp1 reporters by ~80%, and the Igfbp5 reporter by ~30% (Figure 3.S3.B). These data strongly suggest that Tgfbr2, Wisp1, and Igfbp5 are direct targets of miR-135b, of which the latter two are key component of extracellular matrix proteins.

3.3.4. Wisp1 has dual roles during reprogramming and is a key regulator of ECM proteins

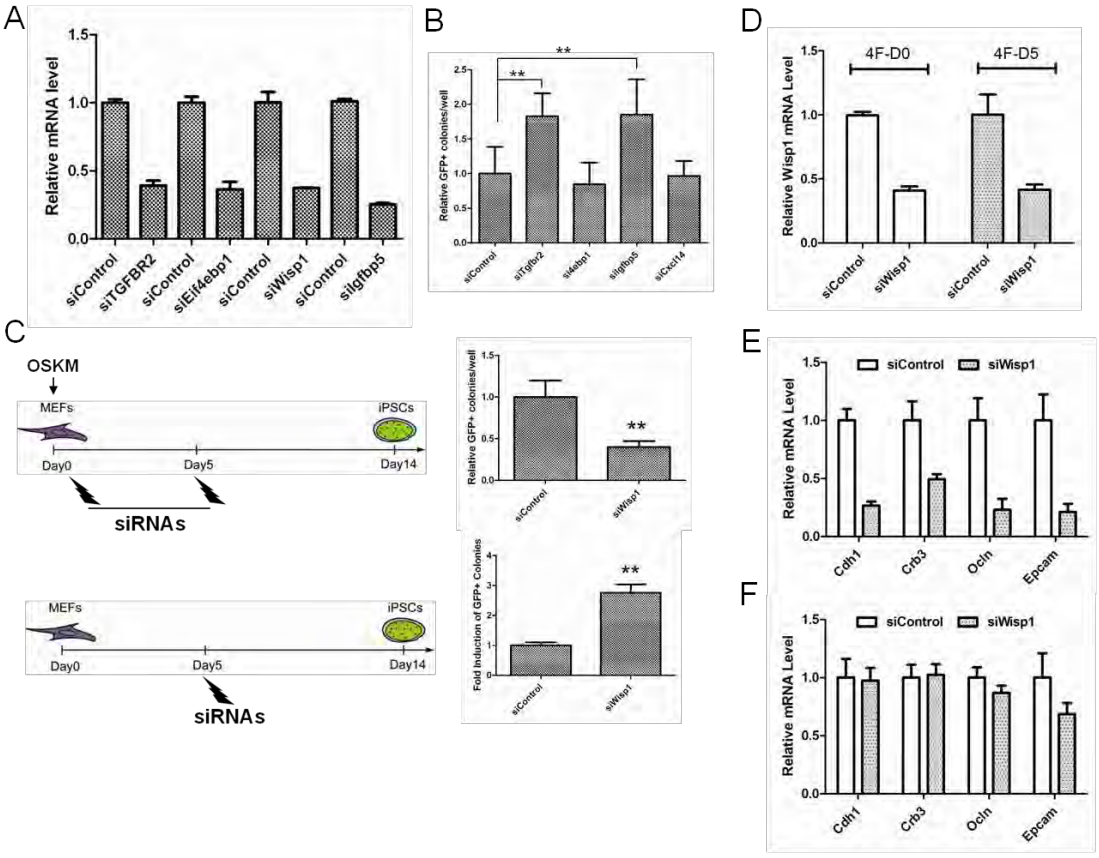
We next asked whether the potential miR-135b targets Tgfbr2, Wisp1, and Igfbp5 function as reprogramming barriers. Tgfbr2 was previously reported to be a reprogramming barrier and a potential target of miR-93 and its family of microRNAs (Li et al, 2011). In addition to Tgfbr2, Wisp1, and Igfbp5, we chose to investigate several other genes that might be indirectly regulated by miR-135b, such as Eif4ebp1 and Cxcl14 as they do not have predicted miR-135b target sites. Before using siRNAs for these experiments, we confirmed by RT-qPCR that each mRNA was efficiently knocked down by at least 60% by its cognate siRNA (Figure 3.4.A).

To determine whether knock-down of the candidate barrier genes increased reprogramming efficiency, we transfected siRNAs into Oct4-MEFs on the same day as 4F transduction (day 0), then again on day 5 post-infection, and counted GFP^+ iPSC colonies on day 11–12. We detected a significant increase in the number of GFP+ colonies after transfection of siRNA targeting Igfbp5 and

Figure 3.4. Wisp1 plays a dual role during reprogramming, while Tgfbr2 and Igfbp5 knockdown enhances reprogramming

(A) Potential target genes are efficiently knocked down by siRNAs. Smartpool siRNAs at a final concentration of 50 nM were used to transfect MEFs. Total RNAs were harvested at day 2 for RT-qPCR to evaluate knockdown efficiency of each siRNA. (B) Knockdown of Tgfbr2 or Igfbp5 enhances Oct4-GFP+ colony formation, while knockdown of Eif4ebp1 and Cxcl14 had no effect. MEFs were transfected with siRNAs on days 0 and 5 at the same time as 4F infection. GFP+ colonies were counted at day 11-12 post-infection. Error bars represent three independent experiments with triplicate wells. The p value was calculated using Student's *t*-test. **p<0.01. (C) Knockdown of Wisp1 shows stage-specific effects on reprogramming. Knockdown of Wisp1 on the same days as 4F transduction (day 0) decreased the reprogramming efficiency by ~70% percent, while knockdown on day 5 enhanced reprogramming by ~3 fold. Error bars represent three independent experiments with triplicate wells. **p<0.01. (D) Wisp1 is efficiently knocked down by siRNAs during both procedures. siWisp1 was transfected at a final concentration of 50 nM on day 0 or day 5. Total RNAs were harvested at day 2 post-transfection for RT-qPCR analysis of Wisp1 expression. (E) Knockdown of Wisp1 at day 0 inhibits mesenchymal-to-epithelial transition (MET). MEFs were infected with 4F and transfected with siRNA on the same day (day 0). Total RNAs were harvested 2 days later. Expression of several MET markers was evaluated. (F) Knockdown of Wisp1 at day 5 does not affect MET. MEFs were transduced with 4F at day 0 and transfected with siRNA at day 5 post-4F infection. Total RNAs were harvested 2 days after transfection and expression of the MET markers was evaluated.

Figure 3.4. Wisp1 plays a dual role during reprogramming, while Tgfbr2 and Igfbp5 knockdown enhances reprogramming



Tgfbr2, consistent with their possible function as barrier genes (Figure 3.4.B). Interestingly, a dramatic decrease in reprogramming efficiency was observed in cells transfected with siWisp1 on days 0 and 5 post-4F infection. However, if siWisp1 was transfected on day 5 only, there was a 3-fold increase in the number of GFP^+ colonies (Figure 3.4.C), suggesting that Wisp1 can play temporally distinct roles during reprogramming. This effect was not due to a difference in siRNA transfection efficiency, because Wisp1 mRNA knockdown was equivalent under both protocols (Figure 3.4.D). To probe this observation further, we next analyzed the effect of Wisp1 siRNA transfection on markers of MET, which is believed to be the initial step of the reprogramming process (Li et al, 2010; Samavarchi-Tehrani et al, 2010). Remarkably, knockdown of Wisp1 on day 0 dramatically decreased mRNA expression of each of the MET markers tested, suggesting a significant delay or suppression of MET by siWisp1 (Figure 3.4.E). In contrast, Wisp1 knockdown on day 5 had little effect on MET marker mRNA levels, except a small and insignificant decrease in Epcam expression (Figure 3.4.F). Thus, these data suggest that Wisp1 may play dual roles during reprogramming.

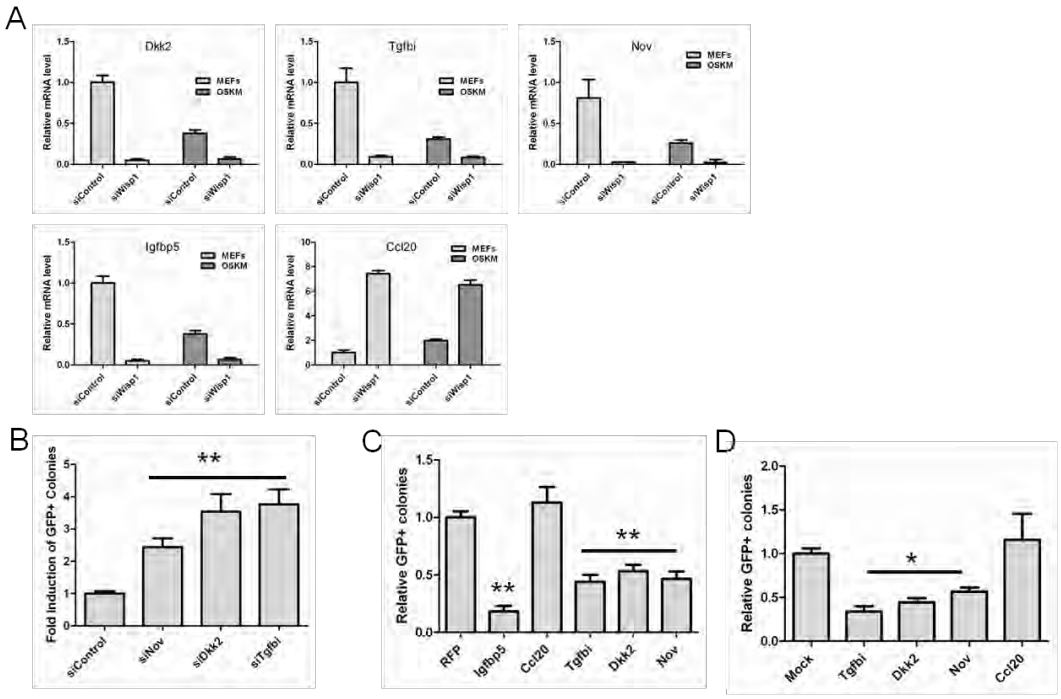
To probe the mechanism by which Wisp1 affects reprogramming, we next investigated the downstream targets of Wisp1. Wisp1 is a member of CCN family proteins, the function of which usually includes two aspects: (1) binding of scaffold of extracellular matrix proteins; (2) binding receptors and transcriptionally regulating signaling events mediated by biological active molecules such as

growth factors and cytokines(Jun & Lau, 2011). We reasoned that since somatic reprogramming is an *in vitro* process, it is more likely that Wisp1 functions through transcriptional regulation of downstream genes. To identify the downstream targets of Wisp1, we utilized mRNA microarrays to search for genes significantly changed upon Wisp1 knockdown in control, non-infected and 4F-transduced MEFs (Table 3.4). The microarray experiments identified a panel of ECM genes, including Dkk2, Igfbp5, Nov, and Tgfbi, that showed profoundly decreased expression upon Wisp1 knockdown, which was confirmed by RT-qPCR (Figure 3.5.A). Moreover, expression of Dkk2, Igfbp5, Nov, and Tgfbi was suppressed by 4F transduction. In addition, Wisp1 knockdown increased expression of Ccl20 (Figure 3.5.A), which was also induced in MEFs by 4F transduction alone. To rule out the possibility of off-target effects of the Wisp1 siRNA, two additional shRNAs were tested. These shRNAs efficiently suppressed Wisp1 expression, and had the same inhibitory effects on expression of Wisp1 target genes. (Figure 3.S4.). To confirm that the miR-135b effects on MEFs was at least partially mediated through Wisp1, we transfected MEFs with an miR-135b mimic, and found decreased expression of Dkk2, Igfbp5, Nov, and Tgfbi (Figure 3.S5.). Thus, Wisp1 may serve as a key regulator of ECM genes in MEFs.

To determine if expression of Wisp1-regulated ECM genes could affect reprogramming, Oct4-GFP MEFs were infected with 4F and on day 5 were transfected with siRNAs targeting Dkk2, Igfbp5, Nov, and Tgfbi. Indeed,

Figure 3.5. Wisp1 is a key regulator of extracellular matrix genes

(A) Wisp1 regulates expression of several ECM genes. Expression of Tgfbi, Igfbp5, Dkk2 , Nov, and Ccl20 were dramatically changed upon Wisp1 knockdown. Uninfected and 4F-infected MEFs were transfected with siWisp1 for 2 days and total RNAs were harvested for RT-qPCR analysis of different ECM genes. Error bars represent two independent experiments with duplicate wells. (B) Knockdown of Nov, Dkk2, and Tgfbi significantly enhances iPSC generation. MEFs were transduced with 4F at day 0 and transfected with siRNAs at day 5 post-infection. GFP+ colonies were quantified at around day 11-13. Error bars represent three independent experiments with triplicate wells. **p<0.01. (C) Overexpression of Wisp1-regulated ECM genes compromises reprogramming. The indicated ECM genes were cloned into pMX retroviral vectors. MEFs were transduced with 4F plus the indicated ECM genes and GFP+ colonies were quantified at around day 11-13. Data was normalized to pMX-RFP–transduced cells. Error bars represent three independent experiments with triplicate wells. **p<0.01. (D) Addition of recombinant ECM proteins compromises reprogramming. Purified recombinant TGFBI, DKK2, NOV, and CCL20 were added at a final concentration of 100 ng/ml to cultures of 4F-MEFs undergoing reprogramming. GFP+ colonies were quantified at day 11-13. Error bars represent two independent experiments with triplicate wells. *p<0.05.



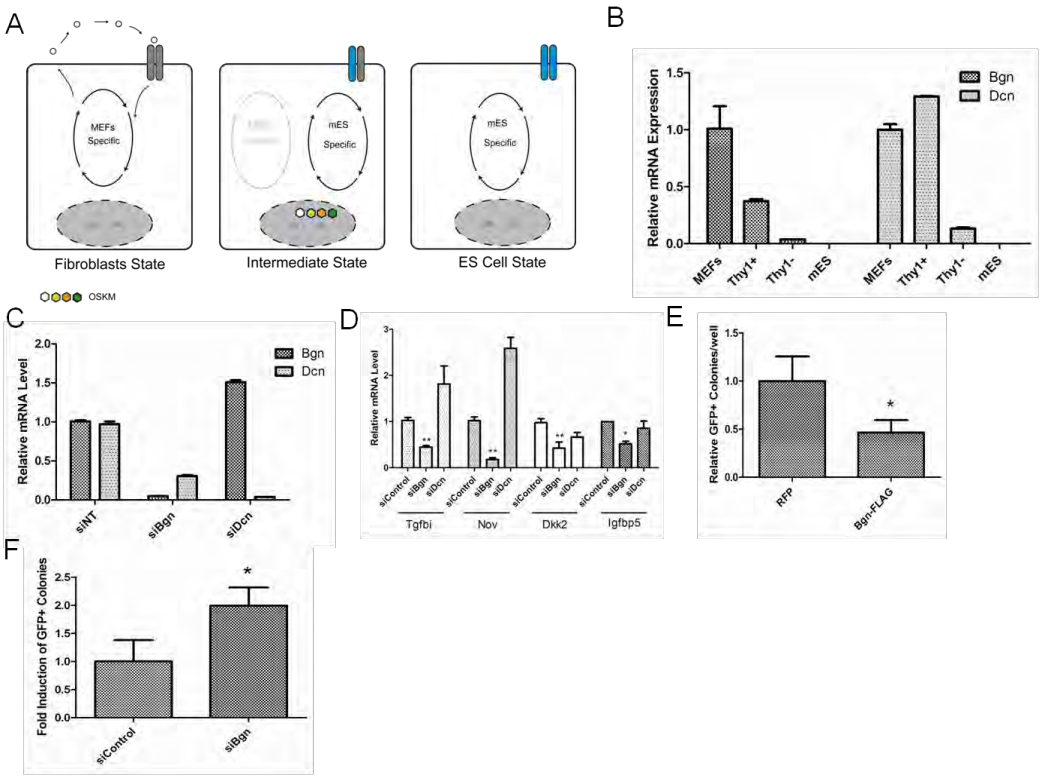
knockdown of each of these genes significantly increased reprogramming efficiency (Figure 3.4.B, Figure 3.5.B). We also detected an increase in mES marker gene expression in the siRNA-transfected cells (Figure 3.S6.). Conversely, overexpression of these genes in MEFs strongly reduced GFP+ colony formation, particularly with Igfbp5, which reduced reprogramming by ~70% (Figure 3.5.C). Interestingly, addition of recombinant DKK2, TGFBI, and NOV proteins to the 4F-transfected MEF cultures from day5 post infection had similar effects on the cells as overexpression of the genes (Figure 3.5.D), demonstrating that the effects of Wisp1 were mediated by secretion of the protein products of its target genes, and confirming that the Wisp1-regulated ECM genes do indeed act as barriers to the reprogramming process.

Based on the results described above, we propose a model of how Wisp1 may display dual roles in MEF reprogramming (Figure 3.6.A). Wisp1 is highly and specifically expressed in MEFs compared with iPSCs (Sridharan et al, 2009), and through its effects on the downstream ECM genes, plays a crucial role in maintaining normal MEF growth. This is supported by our finding that persistent knockdown of Wisp1 in MEFs compromises their proliferation (Figure 3.S7.). Upon 4F transduction and reprogramming, infected MEFs would have two regulatory networks, one established by the four reprogramming factors, and the other being endogenous. The ability of a cell to become fully reprogrammed would depend on whether the 4F-induced network could silence the existing MEF regulatory network. In these cells, although MEF-specific genes such as Wisp1

Figure 3.6. Target gene regulation by Wisp1 through biglycan

(A) Proposed model for Wisp1 dual role during reprogramming. In wild type MEFs (fibroblast state), normal proliferation and function of the cells are dependent on a MEF-specific regulation network, where Wisp1 is one of the most important ECM components and regulates the expression of several other ECM genes. In 4F-transduced MEFs (intermediate state), two systems co-exist; one from the MEF-specific network and the other from the four transcription factors. ECM signals from the MEF-specific network interfere with the cells becoming fully reprogrammed. In fully reprogrammed cells (ES cell state), ECM receptors are no longer expressed, and the cells are thus resistant to interfering signals from surrounding MEFs. (B) Biglycan and decorin are specifically expressed in MEFs. Expression of biglycan and decorin was analyzed by RT-qPCR in sorted cells. (C) Biglycan and decorin are efficiently knocked down by siRNAs. MEFs were transfected with siRNAs for 2 days and total RNAs were harvested for RT-qPCR analysis. (D) Knockdown of biglycan decreases expression of Wisp1-regulated ECM genes. Expression of Wisp1-regulated ECM genes was analyzed in MEFs subjected to knockdown of biglycan or decorin. Error bars represent two independent experiments with duplicate wells. $^{**}p<0.01$. (E) Overexpression of biglycan inhibits reprogramming. Flag-tagged biglycan was cloned into pMX vector and transduced into MEFs together with 4F. GFP+ colonies were quantified at day 11-13. Error bar represents two independent experiments with triplicate wells. $^{*}p<0.05$. (F) Knockdown of biglycan enhances reprogramming. Biglycan siRNAs were transfected into MEFs at day 5 post-4F transduction. GFP+ colonies were quantified at day 11-13. Error bar represents two independent experiments with triplicate wells. $^{*}p<0.05$.

Figure 3.6. Target gene regulation by Wisp1 through biglycan



and its potential receptors are being down-regulated, the remaining receptors could still be stimulated by signals secreted by surrounding unreprogrammed cells. This constant stimulation of original MEF network would compete with 4F-mediated ES regulatory network and resulted in a low efficiency for cells to become fully reprogrammed. Thus, knocking down Wisp1 in these cells could reduce the MEF signaling stimulation, significantly break the balance and push them toward fully reprogrammed state. Once the cells become mES-like cells, MEF ECM genes and receptors are completely shut down and they become resistant to the signals from nearby feeder cells.

3.3.5. Wisp1 may regulate ECM genes through biglycan

To test our model (Figure 3.6.A), we searched the literature for known factors that could interact with Wisp1. If our model is correct, we predict we will see high expression of these genes in the starting population of MEFs, whereas cells undergoing reprogramming will downregulate but not extinguish their expression, and expression will be silenced in fully reprogrammed iPSCs/mES cells. Interestingly, Wisp1 has been reported to bind the proteoglycans decorin and biglycan on the surface of human skin fibroblasts(Desnoyers et al, 2001) and both are highly expressed in MEFs(Sridharan et al, 2009). To determine if decorin and biglycan might be involved in Wisp1 regulation in MEFs, we first examined their gene expression in the starting MEFs, the sorted $Thy1^{+/-}$ cells, and in mES populations. The two genes were highly expressed in MEFs but

undetectable in mES cells (Figure 3.6.B). They were highly expressed in Thy1+ cells and showed strongly reduced but detectable expression in Thy1- cells (Figure 3.6.B), which are enriched in potential iPSCs (Figure 3.1.C-E). We then transfected MEFs with siRNAs targeting these two genes and confirmed the knock-down efficiency by RT-qPCR (Figure 3.6.C). Of interest, knockdown of biglycan also decreased decorin expression, suggesting possible cross regulation of the two genes. Knockdown of biglycan also decreased the expression of the Wisp1 target genes Dkk2, Igfbp5, Nov, and Tgfbi, to a similar level to that seen with Wisp1 knockdown (Figure 3.6.D). Consistent with these observations, overexpression of biglycan strongly suppressed reprogramming, and conversely, knockdown significantly enhanced reprogramming (Figure 3.6.E, F). Therefore, we conclude that biglycan may be an intermediate for Wisp1-mediated regulation of its target ECM genes.

3.4. Discussion

Since the discovery that MEFS can be directly reprogrammed to iPSCs, considerable effort has been made to understand how the four reprogramming transcription factors extinguish endogenous MEF gene expression and gradually re-establish mES-like regulatory networks. Because the process is extremely inefficient, understanding the critical barriers to reprogramming is essential to allow development of novel technologies and compounds to improve the efficiency. Here, we show that microRNAs can be used as a powerful tool to

dissect the molecular mechanisms that elicit successful reprogramming. We analyzed a Thy1– cell population enriched in potential iPSCs to identify its microRNA expression profile during the early stages of reprogramming. From these experiments, we identified sets of microRNAs that were induced or repressed during the process, and showed that manipulating their expression with miR mimics or inhibitors dramatically altered the efficiency of iPSC induction. Among the microRNAs analyzed, miR-135b was the most highly induced by the four factors, and was shown to enhance iPSC generation. Moreover, by mining genome-wide mRNA expression data for potential miR-315b target genes, we showed that Wisp1 and its downstream ECM genes could compromise the efficiency of the reprogramming process. Therefore, our approach has not only identified a novel ECM network that is involved in modulating the reprogramming process, but we have also shown that using microRNAs as probes could be an efficient method to study the molecular mechanisms of reprogramming.

Wisp1 was first described as a Wnt1-inducible protein (Pennica et al, 1998). It belongs to the CCN gene family that encodes six 30–40 KDa secreted proteins (Berschneider & Konigshoff, 2011; Chen & Lau, 2009). CCN proteins have four conserved structural domains with sequences homologous to insulin-like growth factor binding proteins (IGFBPs), von Willebrand factor type C repeat (VWC), thrombospondin type I repeat (TSP), and carboxyl-terminal (CT) domain (See the marked sentence). These domains determine the function of CCN member proteins during development and in human diseases. Although Wisp1

has been linked to oncogenic transformation (Pennica et al, 1998; Xu et al, 2000), proliferation and cell survival (Venkatachalam et al, 2009; Venkatesan et al, 2010), and epithelial-to-mesenchymal transition (Konigshoff et al, 2009), little is known about its downstream genes or how it regulates their expression. In this study, we identified several ECM components that were regulated by Wisp1, likely through its interaction with biglycan. These include Tgfbi, Dkk2, Igfbp5, and Nov. These findings provide some new insights into Wisp1 function. For example, TGFBI is a known downstream gene induced by TGFβ signaling and has profound tumor suppressive effects (Ahmed et al, 2007; Zhang et al, 2009). The TGFβ signaling pathway has itself been identified as a barrier for somatic reprogramming (Ichida et al, 2009; Maherali & Hochedlinger, 2009). Our finding thus indicates there may be crosstalk between Wisp1 and TGFβ signaling in regulating expression of the ECM protein TGFBI. Knockdown of Wisp1 decreases Tgfbi expression, which might compromise TGFβ signaling and allow cells to become fully reprogrammed. Two other Wisp1 target genes we identified are DKK2 and IGFBP5. DKK2 is known as a Wnt signaling antagonist (Kawano & Kypta, 2003) and IGFBP5 could regulate IGF signaling by binding to IGF-1/2 (Beattie et al, 2006). We found that knockdown of Wisp1 decreased expression of Dkk2 and Igfbp5, which would derepress Wnt and IGF signaling. Consistent with this, previous studies have indicated that Wnt signaling could promote somatic reprogramming (Marson et al, 2008). It was recently shown that IGFBP5 overexpression induces cell senescence in a p53-dependent manner (Kim et al,

2007). This protein is highly expressed in fibroblasts, and its expression is further increased upon senescence (Yoon et al, 2004). Thus, decreased expression of IGFBP5 and DKK2 is likely to be beneficial to iPSC generation.

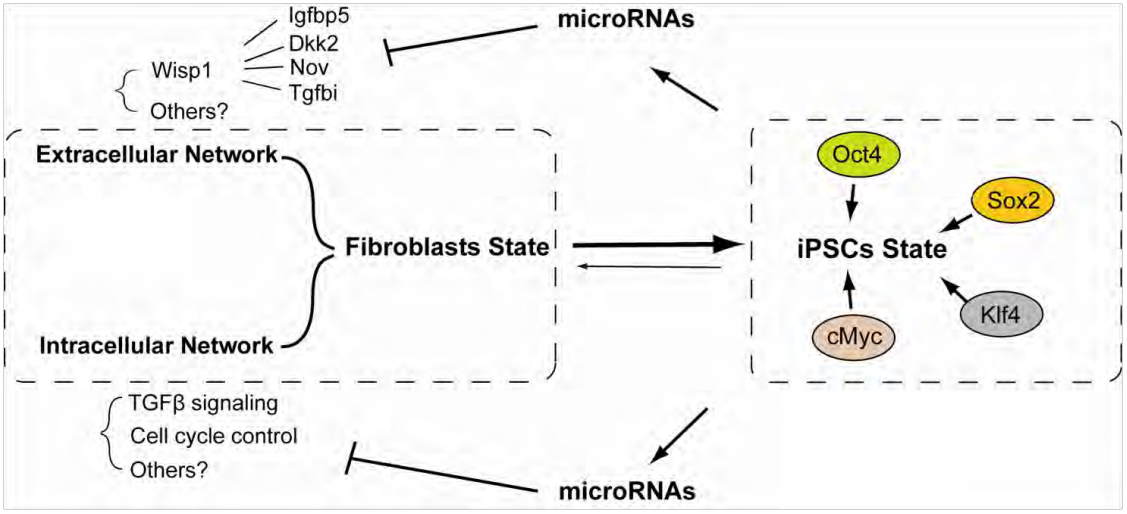
Over the past few years much progress has been made in understanding the molecular mechanisms of somatic reprogramming and several important barrier pathways have been discovered. However, these efforts have mainly focused on intracellular signaling networks, and the effect of the extracellular environment on reprogramming has not been fully explored. Interestingly, recent work has indicated that more than 92% of the monoclonal pre-B cells could reach a fully reprogrammed state when cultured as monoclonal for extended periods(Hanna et al, 2009), which could be only less than 0.1% with mixed starting population and total number of iPSC colonies usually reached plateau at later time points post 4F(Sridharan et al, 2009). This clearly suggests that secreted ECM components could affect the ability of neighboring cells to become fully reprogrammed. In our study, biglycan, a surface glycoprotein that binds Wisp1, is expressed in MEFs but decreases in reprogramming cells, as shown in $Thy1^-$ cells that are enriched with potential iPSCs. These cells will still be stimulated by Wisp1 and presumably other ECM proteins secreted by surrounding feeder MEF cells or unreprogrammed cells, as they still express the receptors such as biglycan, although at much lower level compared with original MEFs. These stimulations would prevent the cells from shutting down MEF-specific regulation network and compete with four factors-mediated regulatory

network to determine the fate of target cells. Meanwhile, our discovery that microRNAs induced by the four factors can regulate ECM genes reveals some new insights into how the four factors manage to reprogram a small percentage of cells. Down-regulation of MEF-specific ECM proteins seems to be part of the entire reprogramming process and is mediated at least in part by 4F-mediated induction of microRNAs such as miR-135b. Together with previous findings, it is clear that microRNAs are important regulators of reprogramming, both through intracellular and extracellular mechanisms (Figure 3.7.).

In summary, we have identified a novel microRNA-mediated pathway of ECM gene regulation that is involved in iPSC generation. Our results indicate that 4F-induced miR-135b expression in turn regulates expression of Wisp1 and Igfbp5. Wisp1 is a key regulator of several ECM proteins, which may be mediated through Wisp1 interaction with biglycan. Our findings not only identify a novel role for ECM components in somatic reprogramming, but also demonstrate that microRNAs can be powerful tools to dissect the molecular mechanisms of iPSC generation.

Figure 3.7. Model for roles of microRNAs during the reprogramming process

MicroRNAs induced by the four factors could have both intracellular and extracellular roles. Intracellularly, some microRNAs may target signaling pathways that are barriers for iPSC generation, such as TGFβ signaling, the p53-p21 pathway, and cell cycle control. Meanwhile, some microRNAs may regulate expression of ECM genes to establish a growth environment that promotes the fully reprogrammed state. Both groups of microRNAs work collaboratively to help 4F to reprogram MEFs to iPSCs.



Materials and Methods

Cell culture, vectors, and virus transduction

Oct4-GFP MEFs were derived from mouse embryos harboring an IRES-EGFP fusion cassette downstream of the stop codon of *pou5f1* (Jackson lab, Stock#008214) at E13.5. MEFs were cultured in DMEM (Invitrogen, 11995-065) with 10% FBS (Invitrogen) plus glutamine and nonessential amino acids (NEAA). Only MEFs at passage 0 to 4 were used for iPSC induction. pMXs-Oct4, Sox2, Klf4, and cMyc were purchased from Addgene. Tgfbi, Dkk2, Igfbp5, Nov, and biglycan overexpression vectors were constructed by inserting cDNA coding sequences into the pMX vector. To generate retrovirus, PLAT-E cells were seeded in 10 cm plates. The next day, the cells were transfected with 9 µg of each vector using Lipofectamine (Invitrogen, 18324-012) and PLUS (Invitrogen, 11514-015). Viruses were harvested and combined 2 days later. For iPSC induction, MEFs were seeded in 12-well plates and the next day were transduced with "four factor" (4F) virus with 4 µg/ml Polybrene. One day later, the medium was changed to fresh MEF medium, and 3 days later it was changed to mES culture medium supplemented with LIF (Millipore, ESG1107). GFP^+ colonies were picked at day 14 post-transduction, and expanded clones were cultured in DMEM with 15% FBS (Hyclone) plus LIF, thioglycerol, glutamine, and NEAA. Irradiated CF1 MEFs served as feeder cells to culture mES and derived iPSC clones.

Recombinant proteins were obtained from commercial sources as follows: mouse Dkk2 (R&D systems, 2435DK/CF), human NOV/CCN3 (R&D systems, 1640NV), human TGFBI (Prospec, #PRO-568), CCL20 (R&D systems, 760-M3).

MicroRNAs, siRNAs, and MEF transfection

microRNA mimics and inhibitory siRNAs were purchased from Dharmacon. To transfect MEFs, microRNA mimics were diluted in Opti-MEM (Invitrogen, 11058-021) to the desired final concentration. Lipofectamine 2000 (Invitrogen, 11668-019) (2 μl/well) was added and the mixture was incubated for 20 min at RT. For 12-well plate transfections, 80 μl of the miR mixture was added to each well with 320 μl of Opti-MEM. Three hours later, 0.8 ml of the virus mixture (for iPSC) or fresh medium was added to each well, and the medium was changed to fresh MEF medium the next day.

Western blotting

Total cell lysates were prepared using M-PER buffer (PIERCE, 78503), incubated on ice for 20 min, and cleared by centrifuging at 13,000 rpm for 10 min. Equal amounts of lysate were loaded onto 10% SDS-PAGE gels. Proteins were transferred to PVDF membranes (Bio-Rad, 1620177) using the semi-dry system (Bio-Rad) and then blocked with 5% milk in Tris-buffered saline–Tween 20

(TBST: 50mM Tris, 150mM NaCl, 0.05% Tween20) for at least 1 hr at room temp or overnight at 4°C. The following antibodies were used: anti-mNanog (R&D Systems, AF2729), anti-h/mSSEA1 (R&D Systems, MAB2156), anti-TGFBR2 (Cell Signaling, #3713), anti-IGFBP5 (R&D Systems, AF578), anti-actin (Thermo, MS1295P0), anti-AFP (Abcam, ab7751), anti-beta III tubulin (R&D Systems, MAB1368), and anti-alpha actinin (Sigma, A7811).

mRNA and microRNA RT and quantitative PCR

Total RNAs were extracted using Trizol (Invitrogen), and then 1 μg total RNA was used for RT using Superscript II (Invitrogen). Quantitative PCR was performed using a Roche LightCycler480 II and the SYBR Green mixture from Abgene (Ab-4166). Mouse Ago2, Dicer, Drosha, Gapdh, and p21 primers are defined in Supplemental Table 2. Other primers were described previously (Takahashi & Yamanaka, 2006). For microRNA quantitative analysis, total RNA was extracted using the method described above. Between ~1.5 and 3 μg of total RNA was used for microRNA reverse transcription using QuantiMir kit following the manufacturer‘s protocol (System BioSciences, RA420A-1). RT products were then used for quantitative PCR using the mature microRNA sequence as a forward primer and the universal primer provided with the kit.

Immunostaining

Cells were washed twice with PBS and fixed with 4% paraformaldehyde at room temperature for 20 min. Fixed cells were permeabilized with 0.1% Triton X-100 for 5 min, and then blocked in 5% BSA in PBS containing 0.1% Triton X-100 for 1 hr at room temperature. Primary antibody was diluted at 1:100 to 1:400 in 2.5% BSA PBS containing 0.1% Triton X-100, according to the manufacturer‘s protocol. Cells were stained with primary antibody for 1 hr and then washed three times with PBS. Secondary antibody was diluted 1:400 and cells were stained for 45 min at room temperature.

Embryoid body formation and differentiation assay

iPSCs were trypsinized to a single cell suspension, and the hanging drop method was used to generate embryoid bodies (EB). For each drop, 4000 iPSCs in 20 µl EB differentiation medium were used. EBs were cultured in hanging drops for 3 days before being reseeded onto gelatin-coated plates. After reseeding, cells were cultured until day 14, when apparent beating areas could be identified.

Teratoma formation

To generate teratomas, iPSCs were trypsinized and resuspended at a concentration of $1x10^7$ cells/ml. Athymic nude mice were anesthetized with

Avertin, and 150 µl of iPSCs were injected into each mouse. Tumors were monitored every week for ~3–4 weeks. Tumors were then harvested and fixed in Z-Fix solution for 24 hrs at room temperature, before paraffin embedding, sectioning, and H&E staining. To further evaluate pluripotency of derived iPSC clones, iPSCs were injected into C57BL/6J-Tyr$^{(C\text{-}2J)/J}$ (albino) blastocysts. Generally, each blastocyst received 12–18 iPSCs. ICR recipient females were used for embryo transfer.

Acknowledgments

We thank Greg Hannon, Craig Mello, Mark Mercola, Evan Snyder, and members of the Rana laboratory for helpful discussions and support. We are grateful to the following shared resource facilities at the Sanford-Burnham Institute: Genomics, Informatics and Data Management Core for miRNA and mRNA array experiments and data analysis; Animal Facility for the generation of teratomas and chimeric mice; and the Histology and Molecular Pathology Core for characterization of various tissues.

Figure 3.S1. miR-135b enhances the overall percentage of Oct4-GFP+ cells during reprogramming

MEFs were transfected with the indicated microRNA mimics 3 hrs before infection with 4F, and cells were trypsinized on day 14 for FACS analysis. Single cells were collected by filtering through a cell strainer. Non-transduced MEFs served as negative controls.

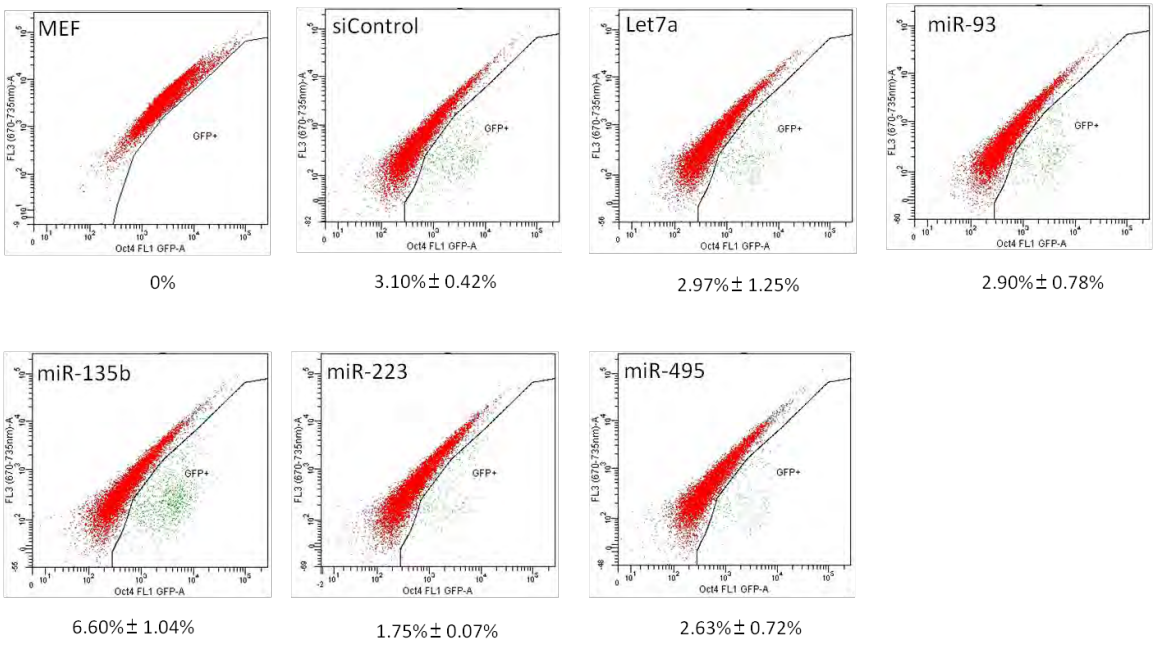


Figure 3.S2. miR-135b iPSCs show full differentiation capacity

Lineage markers are expressed in differentiated EBs from miR-135b–induced iPSCs. EBs were formed using the hanging drop method for two days and replated onto gelatin-coated plates until day 12-14. Cells were then fixed and stained for AFP (endoderm), tubulin III (ectoderm), and α-actin (mesoderm) expression. DAPI was used for nuclear staining.

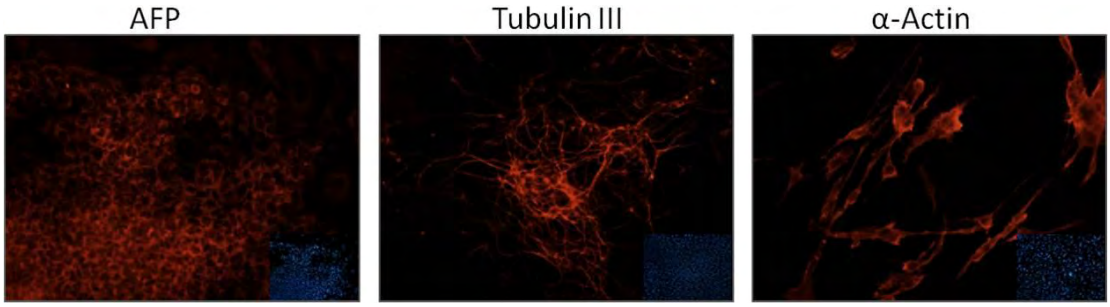
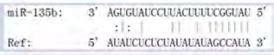
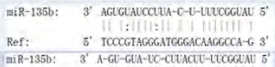
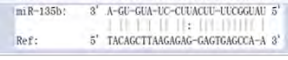


Figure 3.S3. Tgfbr2, Wisp1, and Igfbp5 are directly regulated by miR-135b

(A) List of predicted miR-135b target sites identified by both miRanda software and Targetscan in Tgfbr2, Wisp1, and Igfbp5 3'UTRs. (B) Dual luciferase assay supported direct regulation by miR-135b. The full length Tgfbr2 3'UTR, a Wisp1 fragment, and the Igfbp5 3'UTR were cloned into pGL3 luciferase reporters and transfected into HeLa cells together with pRL-TK. Relative luciferase activity was calculated by the GL/RL signal and normalized to siControl-transfected cells. p values were calculated using Student's *t*-test from at least two independent experiments with duplicate wells.

A

Gene Names	# of Predicted Sites (miRanda/TargetScan)	Representative Sites	Predicted Energy
Tgfbr2	0/1		N/A
Wisp1	3/1		-24.7kcal/mol
Igfbp5	1/0		-16.8kcal/mol

B

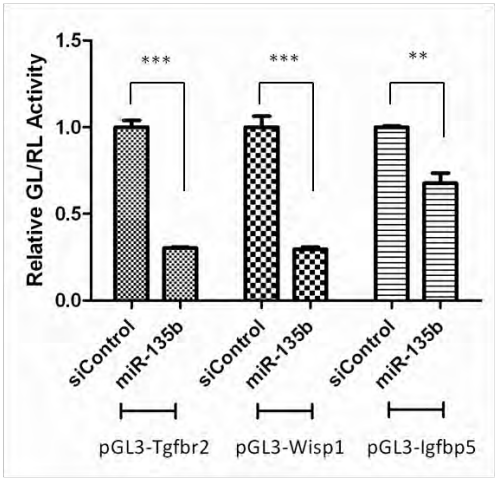


Figure 3.S4. Wisp1 regulates expression of several ECM genes

Wisp1 was knocked down in MEFs by two shRNAs. Expression of representative ECM genes was examined 4 days post-infection. Expression of Tgfbi, Nov, and Dkk2 were strongly decreased upon Wisp1 knockdown, similar to results from siRNAs transfection (Figure 3.5.A)

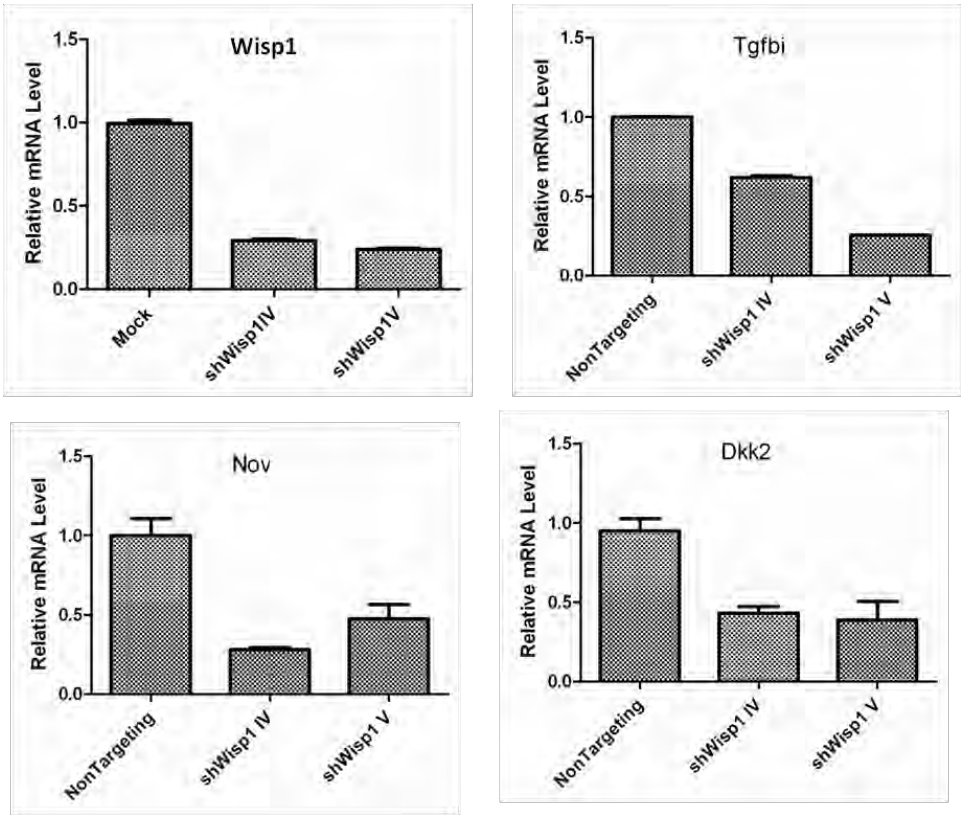


Figure 3.S5. Wisp1 ECM target genes are regulated by miR-135b in MEFs

MEFs were transfected with miR-135b mimic at a final concentration of 50 nM for 4 days. Total RNAs were harvested for RT-qPCR analysis of the indicated Wisp1-regulated ECM genes. Error bar represents experiment with duplicate wells.

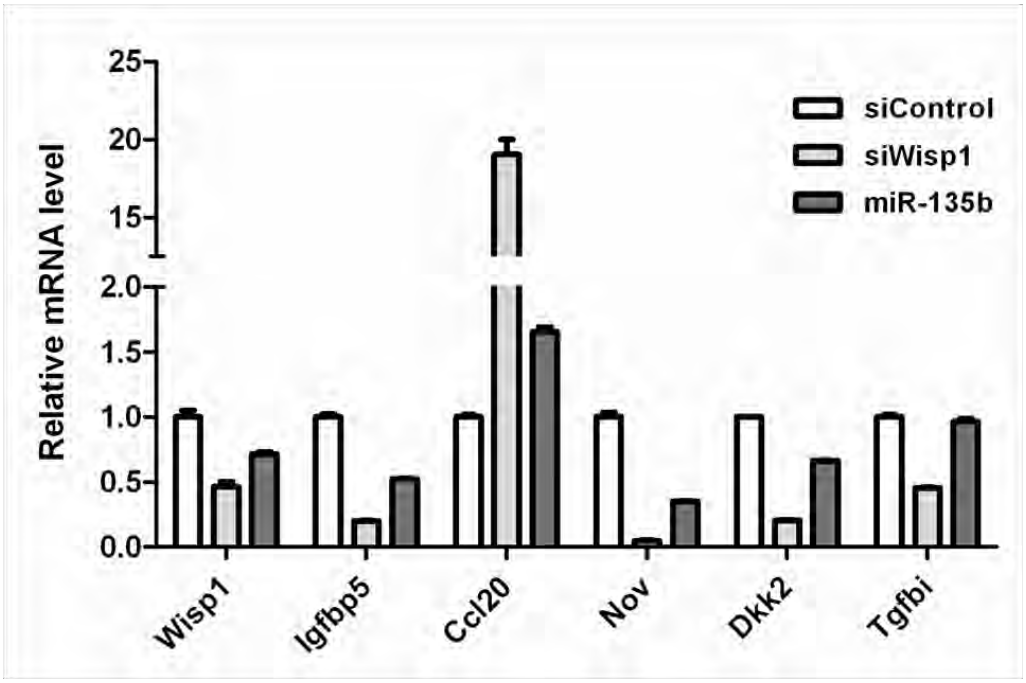


Figure 3.S6. Knockdown of Wisp1 target genes enhances iPSC marker expression

Nov, Dkk2, and Tgfbi were knocked down in 4F-transduced MEFs at day 5 post-transduction. Cells from each well were harvested at around day 14 and total RNAs were extracted for RT-qPCR analysis of the representative mES markers, E-Ras, Nanog, and Tet1.

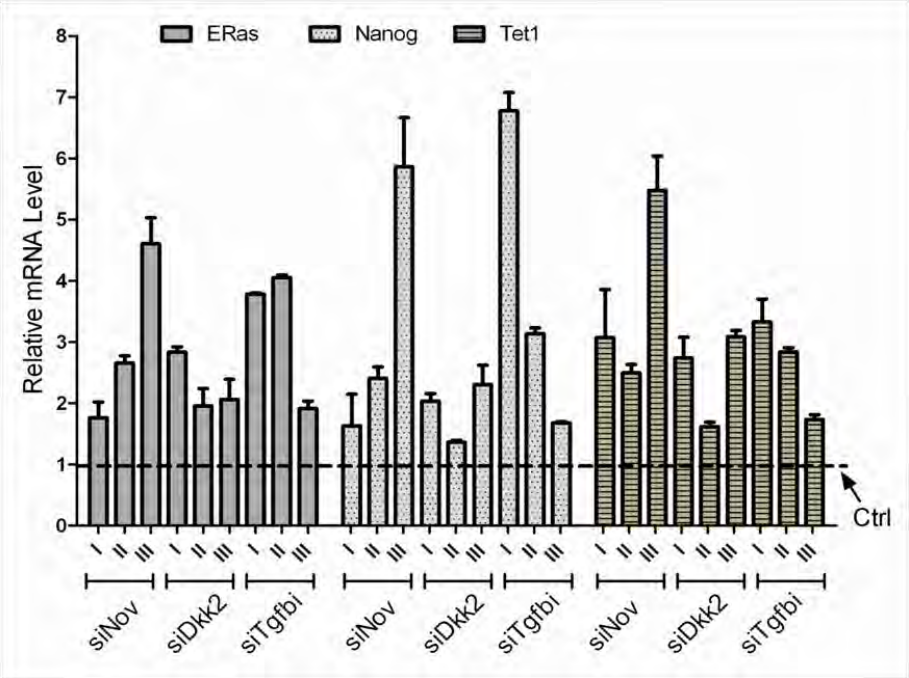


Figure 3.S7. Knockdown of Wisp1 compromises proliferation of normal MEFs

(A) Wisp1 was efficiently knocked down by shRNAs. Two shRNAs targeting mouse Wisp1 were transduced into MEFs. Knockdown efficiency was evaluated at day 4 post-transduction. (B) Consistent knockdown of Wisp1 compromised proliferation of MEFs. MEFs were transduced with shRNAs and then reseeded into 96-well plates. Proliferation of MEFs was measured every four days using Celltiter 96 One Solution assay (Promega, G3582).

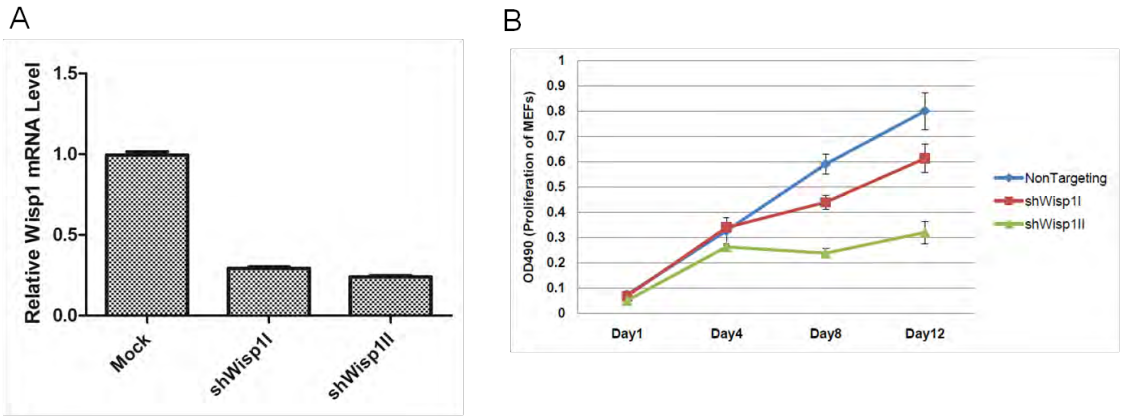


Table 3.1. miR-135b target site analysis

Genes showing significantly repressed expression upon miR-135b transfection were analyzed with miRanda and TargetScan to identify potential miR-135b target sites in their 3'UTR regions. Sites with good seed match and significant predicted energy are listed.

Gene Names	No. Of Predicted Sites	Represetative Sites	Predicted Energy
Mmp13	2/0	miR-135b: 3' AG-U-GU-A-UCC-UUACUUUUC-GGUAU 5' Ref: 5' TTAAATATTTAGGCATTGAAAAGACCATA 3'	-20.3kcal/mol
Gpc1	1/1	miR-135b: 3' AGUGUAUCCUUA-CUUUUCGGU-AU 5' Ref: 5' TCAC-T-GGGTTAGAGAGGCTATTG 3'	-20.8kcal/mol
Mmp10	0/0	N/A	N/A
Ece1	1/1	miR-135b: 3' AGUGUAUCCUUA-C-U-UUUCGGUAU 5' Ref: 5' TC-CA-GGGAATGGCAGAAAGTCATC 3'	-18.79kcal/mol
1190002H23Rik	1/na	miR-135b: 3' AGUGUAUCCUU-ACUUUUCGGUAU 5' Ref: 5' TTATATA-TAGTTG-GAGGTCATA 3'	-16.7kcal/mol
Rgs16	2/2	miR-135b: 3' AGU-GUA-UCCUUACUUUUCGGU-A-U 5' Ref: 5' TGAGC-TAAGGGA-GAGAAGTCATTGA 3'	-19.2kcal/mol
Entpd4	1/1	miR-135b: 3' AGUGUAUCCU-UACUU--UUCGGUA-U 5' Ref: 5' TC-C-TGGGACA-GAACCAAGCCGTGA 3'	-17.2kcal/mol
Sema7a	0/0	N/A	N/A
Inhba	0/1	miR-135b: 3' AGUGUAUCCUUACUUUUCGGUAU 5' Ref: 5' AUAAUAAUGUCUAGAAAGCCAUG 3'	N/A
Adra2a	0/1	miR-135b: 3' AGUGUAUCCUUACUUUUCGGUAU 5' Ref: 5' UCAGCCCCGUGUAUAAAGCCAUU 3'	N/A
Creld1	0/0	N/A	N/A
Sprr2k	0/na	N/A	N/A
Cxcl14	0/0	N/A	N/A
Tgfbr2	0/1	miR-135b: 3' AGUGUAUCCUUACUUUUCGGUAU 5' Ref: 5' AUAUCUCUCUAUAUAUAGCCAUA 3'	N/A
Wisp1	3/1	miR-135b: 3' AGUGUAUCCUUA-C-U-UUUCGGUAU 5' Ref: 5' TCCCGTAGGGATGGGACAAGGCCA-G 3'	-24.7kcal/mol

Ccl12	0/na	N/A	N/A
2310001A 20Rik	1/na	miR-135b: 3’ AGUGUAUCCUUA-CUU-UUCGGUAU 5’ Ref: 5’ TCACATAGGTCTGGAAGGAGCCATG 3’	-28.6kcal/mol
Ogfr	0/0	N/A	N/A
Aqp5	1/na	miR-135b: 3’ AGUGUAUCCUUAC-UUUUC-GGUA-U 5’ Ref: 5’ TGGGAT-GGGA-GCAGAAGCCCATGA 3’	-18.7kcal/mol
Dusp4	0/0	N/A	N/A
Grem1	0/0	N/A	N/A
Trib3	0/0	N/A	N/A
Ifit3	0/0	N/A	N/A
St3gal3	1/0	miR-135b: 3’ A-G-UGUAUCCUU-ACUUUUCGGUAU 5’ Ref: 5’ TCCCACAGGGGAACCCCAGAGCCATC 3’	-20.3kcal/mol
Mthfd2	0/0	N/A	N/A
Igfbp5	1/0/1*	miR-135b: 3’ A-GU-GUA-UC-CUUACUU-UUCGGUAU 5’ Ref: 5’ TACAGCTTAAGAGAG-GAGTGAGCCA-A 3’	-16.8kcal/mol
Nbl1	0/0	N/A	N/A

Target sites are shown as ~~-~~miRanda predicted/ Targetscan predicted”. * RNAhybrid identified another seed-match site.

Table 3.2. Original microRNA expression profile data

List of microRNAs significantly (2-fold, $p<0.05$) altered at reprogramming day 5 in Thy1- cells.

Group ()*	MEF_D0	Thy1+ D5	Thy1- D5
Systematic	Norm	Norm	Norm
ILMN_3166979	23609.834	11127.709	5009.208
ILMN_3167887	2198.5	1020.5833	467
ILMN_3167409	186.45834	177.125	45.958336
ILMN_3167846	2968.25	3031.4583	756.5416
ILMN_3168509	9271.25	5771.125	2394.125
ILMN_3167397	1008.0833	618.4167	291.58334
ILMN_3167698	332	295.5	104.75
ILMN_3167215	1054.2916	709.2291	341.58334
ILMN_3167383	3672.5	3614.625	1195.1666
ILMN_3167419	10785	11485.334	4213.4585
ILMN_3167052	9908.291	7695.708	4398.333
ILMN_3167553	1670.4167	1320.9583	3398.25
ILMN_3168517	10821.084	18965.5	22198.043
ILMN_3166986	5989.583	6450.375	12607.75
ILMN_3167510	17315.918	33186.332	36422.082
ILMN_3167938	94.625	197.79167	236.04166
ILMN_3167918	5277.917	7928.625	14486.333
ILMN_3168117	166.47916	1955.6667	743.75
ILMN_3168303	696.3333	6182.8335	5252
ILMN_3167279	222.91666	757.625	3749.1665
ILMN_3167874	446.2083	8373.084	13559.25

Table 3.3. mRNA expression profile upon miR-135b transfection

Significantly altered mRNAs upon miR-135b mimic transfection are listed.

Genes	siControl	miR-135b	135b/ctrl	Genes	siControl	miR-135b	135b/ctrl
Mmp13	2439.7168	538.7083	0.22080772	Igfbp5	10339.18	5151.6	0.49826
Ssr2	9958.9	2575.1333	0.25857608	Nbl1	4588.85	2290.642	0.499176
Gpc1	17654.266	5494.533	0.31122976	Ndufa4	25525.38	12770.29	0.500298
Mmp10	898.45	280.84583	0.31258927	Ly6a	20063.09	40192.02	2.003281
Ece1	3008.7	1005.7916	0.33429441	BC031353	960.2375	1930.3	2.010232
1190002H23Rik	885.3833	321.30835	0.36290311	Btbd1	1216.267	2449.933	2.014306
Rgs16	7778.225	2841.0166	0.36525256	Kif1b	1098.592	2218.404	2.019316
Entpd4	4623.3667	1703.7334	0.36850493	Adamts12	1180.025	2405.308	2.038353
Sema7a	980.56665	373.225	0.38062176	Tmem43	2145.792	4415.1	2.057562
Inhba	12432.358	4787.7	0.38509991	4933439C20Rik	1743.217	3600.117	2.065215
Adra2a	1711	668.8833	0.39093121	Ncald	178.2	368.5583	2.068229
Creld1	2141.55	837.1625	0.39091429	Nnmt	2737.992	5777.225	2.110023
Eif4ebp1	14258.191	5669.408	0.39762464	Notch3	243.0708	513.575	2.112862
Sprr2k	6371.658	2581.3	0.40512218	Cxadr	208.1375	440.5708	2.11673
Cxcl14	5082.5	2109.0833	0.41496966	Ptprv	540.675	1150.179	2.127302
Tgfbr2	6950.125	2935.2915	0.4223365	Mxd4	588.1625	1261.933	2.145552
Wisp1	11202.6	4737.1167	0.42285868	Gabarap	7238.642	15543.54	2.147301
Ccl12	1894.2	812.8458	0.42912354	Tle6	489.025	1058.233	2.163966
2310001A20Rik	1022.275	441.84998	0.43222223	Fas	751.0667	1629.85	2.170047
Ogfr	2861.7666	1272.846	0.44477631	Ephx1	6257.483	14002.01	2.237642
Aqp5	1304.9917	584.11664	0.4476018	Ddit4l	1087.962	2482.025	2.281352
Nptx2	925.5	421.3333	0.45524938	Nipa1	718.1333	1642.921	2.287766
Dusp4	1473.9667	691.94995	0.46944748	Ctxn	457.9667	1069.146	2.334549
Grem1	1314.4417	616.475	0.4690014	Tcea3	1136.992	2677.825	2.355184
1200015F23Rik	898.875	432.25833	0.48088814	Il7	292.7917	692.8709	2.36643
Nras	7671.4497	3686.5417	0.48055346	Asah1	10981.79	26243.63	2.38974
Trib3	2113.9917	1017.7708	0.48144503	Hmox1	889.9375	2178.9	2.448374
Ifit3	2212.2583	1080.0833	0.48822658	1700007K13Rik	630.1209	2569.5	4.077789
St3gal3	3312.25	1609.75	0.48599894	Sncg	691.1	2831.779	4.097496
Mthfd2	1443.7	716.5416	0.49632306				

Table 3.4. mRNA microarray data upon Wisp1 knockdown

Significantly altered mRNAs upon siWisp1 transfection are listed.

Probe Set ID	Gene Accession	Gene Symbol	MEF-siControl-1	MEF-siControl-2	MEF-siWISP1_1	MEF-siWISP1_2	AVG siCtrl	AVG siWisp1	siWisp1/siCtrl
	ENSMUST0000009781								
10357135	5	2900080B14Rik	0.00414595	0.142908	1.851892	4.25005	8.073661298	2.950971	40.08184972
10541605	NM_020001	Clec4n	87.13544	72.50385	514.3284	657.1372	79.819545	585.7328	7.338212715
10451846	NM_178796	A530064D06Rik	97.63457	83.87197	409.7281	402.8122	90.75327	406.27015	4.476644753
10487588	NM_010554	Il1a	59.19801	59.11544	228.3206	276.6121	59.156725	252.46635	4.267754004
10389222	NM_009139	Ccl6	599.854	576.1665	2543.921	2132.172	588.01125	2338.0465	3.976193483
10582592	NM_009606	Acta1	1924.714	1910.103	8591.48	8974.367	1917.4085	8782.9235	3.537547424
10492964	NM_009690	Cd5l	34.14756	50.81131	142.0664	134.4922	42.479455	138.2793	3.255205725
10347688	NM_016960	Ccl20	444.7801	557.894	1526.542	1556.135	501.08705	1541.3385	3.075988491
10415425	NM_001024714	Cma2	194.6143	233.9767	554.6512	716.4804	214.2955	635.5658	2.965791629
10375137	NM_031169	Kcnmb1	376.4274	401.6824	1048.762	1149.925	389.0549	1099.3435	2.825677045
10541614	NM_010819	Clec4d	483.6474	543.2029	1401.896	1400.961	513.42515	1401.4385	2.729586776
10437885	NM_013807	Myh11	296.9653	262.4589	738.3807	772.7568	279.7121	755.56875	2.701237272
10559467	NM_011089	Pira2	205.6219	222.582	538.6416	614.342	214.10195	576.4918	2.692604154
10403671	NM_012054	Aeah	121.583	153.4939	375.9355	363.3268	137.52845	369.63105	2.687689715
10347335	NM_013612	Slc11a1	883.8206	868.6085	1047.515	951.2921	876.21455	999.40355	2.656472351
		OTTMUSG0000000097							
10379727	NM_001081957	1	1235.559	1375.978	3267.534	3598.148	1305.7685	3432.841	2.628981324
10489852	NM_026385	Lce1h	388.4616	490.0786	1126.301	1156.005	439.2701	1141.153	2.597839006
10543239	NM_031198	Tcfec	385.9265	354.348	840.411	1001.148	360.13725	920.7795	2.556746074
10544273	NM_001038604	Clec5a	129.1076	121.6859	285.2968	344.5444	125.39675	314.9205	2.511392839
10493866	NM_011477	Spr2b	270.4891	271.6027	693.8135	662.226	271.0459	678.01975	2.501494212
10389214	NM_011336	Ccl9	1399.83	1613.804	3525.689	3823.239	1506.817	3674.464	2.438553587
10567590	NM_030691	Igsf6	489.8768	472.0884	1139.302	1198.944	476.4826	1169.073	2.452561021
10446965	NM_207046	Rasgrp3	429.6595	480.7079	1068.112	1131.505	455.1837	1099.8085	2.413989121
10531415	NM_021274	Cxcl10	453.0296	631.5493	1218.222	1380.99	542.28945	1299.606	2.396517211
10559454	NM_011089	Pira2	345.2755	357.8511	787.1019	879.8149	351.5633	833.4584	2.370720721
10376765	NM_007436	Aldh3a1	3.235134	35.81816	50.87209	39.79919	19.426847	45.33564	2.333683214
10527838	NM_009683	Alox5ap	654.0897	609.5988	1352.605	1538.227	631.84315	1460.416	2.311358444
10420254	NM_008572	Mcpt8	538.1488	582.7675	1206.307	1339.947	560.45815	1273.122	2.271573712
10548552	NM_008462	Klra2	58.40049	41.42778	94.62579	123.3385	49.914125	108.982145	2.228930145
10420041	NM_010782	Mcpt9	165.6362	199.6875	327.2544	480.0104	182.66185	403.6324	2.209724691
10531420	NM_019494	Cxcl11	132.152	137.726	300.6798	294.8491	134.939	297.76445	2.206659676
10345085	NM_001077353	Gsta3	185.1671	176.529	378.9051	410.4502	180.84805	394.67765	2.18237161
10547664	NM_019948	Cldn4-e	98.03539	112.7028	240.6388	216.1872	105.369995	228.413	2.167741879
10388556	NM_013804	Hey2	113.158	121.1043	230.4421	276.7768	117.13115	253.60945	2.165175105
10564211	NM_013870	Snrpn	552.9745	519.1485	1139.826	1179.947	536.0605	1159.8865	2.163723125
10534927	NM_153510	Pira	167.8992	200.3085	379.2683	410.3368	184.10285	394.80255	2.144467345
10448124	NM_013521	Fpr1-	106.0472	113.65	232.7581	241.4954	110.8636	237.12675	2.138905376
10510280	NM_008726	Nppb	175.3245	199.0905	383.5877	410.8336	187.2075	397.21065	2.121786758
10572897	NM_010442	Hmox1	6046.282	2955.981	6259.595	6428.228	3091.1215	6343.9115	2.11384694
	NM_011468 //								
10493850	NM_011468	Spr2a // Spr2a	491.8218	596.2354	1094.969	967.1816	498.9286	1031.0753	2.066578865
10401831	NM_001013824	EG435537	12.05905	11.44185	31.79687	16.58742	11.75045	24.192145	2.058827109
10404132	NM_001111110	Cmah	225.5872	254.0497	473.2353	513.6889	239.81845	493.4621	2.057648609
10557962	NM_001082960	Itgam	545.4023	636.3674	1204.543	1227.073	590.88485	1215.808	2.057605564
10583056	NM_008605	Mmp12	972.1359	1038.697	2050.397	2089.776	1005.41645	2069.0865	2.058988258
10474488	NM_001005568	Olfr1281	1.076104	21.47083	21.78144	24.19744	11.273367	22.98944	2.039270078
10570432	AF357390	Rpl27a	1406.751	1440.196	1887.047	3847.022	1424.4735	2867.0345	2.012697674
10461628	NM_146990	Olfr1494	26.09767	37.20738	51.81674	78.93707	32.052525	65.376905	2.002200596
10519857	NM_010427	Hgf	780.8367	821.1526	391.6112	408.6849	800.99465	400.14805	0.499563948
10362418	NM_029726	Trdn	18.61873	11.71195	6.760596	8.21148	15.165355	7.486038	0.494267011
10409633	NM_008086	Gas1	5909.345	6136.092	2057.145	2664.018	6022.7135	2360.5815	0.491589373
10508659	NM_053116	Wnt16	599.4598	649.7448	259.1581	303.0444	624.6023	281.10125	0.450090548
10548899	NM_181988	Rerg	821.7764	748.3213	381.0078	388.0913	785.04885	384.54955	0.489841568
10496125	NM_020265	Dkk2	2173.677	2189.542	1080.232	1047.231	2171.6095	1063.7315	0.489835534
	NM_001011734 //								
10473521	NM_001011734	Olfr1116 // Olfr1116	24.19383	13.11542	9.637003	8.438077	18.654625	9.03754	0.484466453
10545372	NM_153778	Atoh8	689.5418	670.2474	333.7052	315.0749	679.8945	324.39005	0.484240504
10540298	NM_007697	Chl1	346.6982	352.2888	169.4216	148.6097	349.4935	159.01565	0.483601098
10405587	NM_009369	Tgfbi	5161.856	4932.289	2493.344	2399.927	5047.1125	2446.9955	0.480999138
10440256	NM_010140	Epha3	1278.706	1343.538	625.5641	621.6735	1311.123	623.6188	0.475637145
10439009	NM_007470	Apod	1369.713	1662.219	729.6307	718.3503	1515.966	723.9905	0.474447334
10587816	NM_178738	Prss35	346.8098	371.4227	168.1128	169.2805	359.31825	168.69685	0.469754989
10348451	NM_007722	Cxcr7	848.3486	917.3533	432.4385	393.0077	883.10105	412.7231	0.467356595
10413012	AK034234	Fut11	249.4395	91.20593	55.16494	103.7449	170.322715	79.45492	0.466496321
10487890	BC099428	Etv3	299.1889	311.6796	150.3044	133.8258	305.42425	141.965	0.46481247
10607124	NM_001114385	Chrdl1	495.2882	485.2388	241.299	214.3399	490.7635	227.81945	0.464214331
10382231	NM_008425	Kcng2	313.9214	318.7788	141.5998	159.1168	316.3501	145.8583	0.461069239
10357870	NM_054077	Prelp	378.9084	333.2182	162.9578	160.1762	353.5633	161.5669	0.456967395
10538802	NM_172399	A930038C07Rik	1679.919	1513.927	786.936	804.4774	1746.923	795.7067	0.455490425
	ENSMUST0000009707								
10498383	5	Igsf10	3895.838	4004.488	1724.1	1788.919	3850.183	1746.5095	0.453619626
10450038	NM_020581	Angptl4	3711.44	3384.789	1849.014	1351.207	3548.1145	1600.1105	0.450974877
10536811	NM_019697	Kcnd2	401.7512	432.9271	170.3589	205.1546	417.33915	187.75575	0.449890095
10486575	NM_172873	Frmd5	550.8121	557.9852	249.3508	247.8025	554.14865	248.57665	0.448575952
10597817	NM_031161	Cck	1638.339	1638.386	611.8253	754.4223	1638.3535	683.1238	0.444061854
	ENSMUST0000007848								
10566543	2	Dchs1	727.931	707.0428	308.5058	326.8718	717.4869	317.7388	0.442849619
10502805	NM_008966	Ptgfr	405.9804	428.9369	202.8004	167.094	417.45865	184.9472	0.442791639
10353859	NM_009961	Rgs2	1087.129	1176.586	479.9415	516.2652	1131.8575	498.10335	0.440076025
10488722	NM_010278	Gtra1	3918.313	4030.264	1768.492	1728.284	3974.2885	1748.388	0.439924762
10382495	NM_009925	Col10a1	896.3862	907.6484	384.3805	406.4207	903.0163	395.4006	0.437866515
10388317	NM_134005	Enpp3	1307.094	1296.795	539.0471	598.126	1301.945	568.58655	0.436720868
10395409	NM_008584	Meox2	377.6159	439.1997	171.2692	184.9001	408.4078	178.08465	0.436046153
10522503	NM_011058	Pdgfra	3590.019	3451.862	1589.233	1469.671	3520.9405	1529.452	0.434387348
10602896	NM_176712	Gpr84	993.9915	1092.597	444.3554	458.507	1043.29425	451.4312	0.43269787
10575893	NM_173016	Vat1l	381.1485	377.8017	172.6703	154.2159	379.4751	163.4431	0.430708365
10428388	NM_172815	Rspo2	1427.674	1550.445	620.7136	680.9503	1489.0596	650.83195	0.430380204
10389752	NM_175878	Lrrtm2	769.2408	840.5013	320.8965	367.4495	804.87095	344.17	0.427608923
10428767	NM_009701	Aqp5	423.3056	417.7951	170.0578	186.2587	420.54535	178.15675	0.423837427
	ENSMUST0000009707								
10498386	5	Igsf10	2976.198	3240.756	1287.082	1321.116	3108.477	1304.099	0.419526669
10423049	BC096586	Prlr	130.788	111.9024	53.24813	45.57176	121.3452	49.409845	0.407176757
10566272	NM_008221	Hbb-y	32.03662	20.66527	10.49784	10.76144	26.350945	10.62964	0.403387431
10408557	NM_025429	Serpinb1a	264.4742	289.0148	98.26661	123.9195	276.2445	111.093055	0.402154812
10353252	NM_001033288	Gm106	233.7257	219.0332	86.9255	93.29115	226.37945	90.108325	0.398041099
10458828	NM_033037	Cdo1	2089.156	2247.747	826.0869	896.5013	2168.4515	861.2941	0.397188547
10459866	NM_028122	Slc14a1	943.0928	968.3106	368.2968	385.6486	955.7017	367.0727	0.384067106
10359307	NM_177833	Tnn	1096.158	1118.896	404.2644	425.2074	1107.527	414.7359	0.374470239
10518526	NM_001039554	Angptl7	674.3639	642.2687	231.3114	244.3028	658.3153	237.8071	0.36123587
10455721	NM_026408	Sncaip	1154.966	1133.055	373.0126	443.5481	1144.0105	408.28035	0.356885142
10487987	NM_177303	Lrrn4	444.326	443.2074	156.4888	159.9246	443.7667	157.9067	0.356035502
10607848	NM_019397	Egfl6	462.1656	469.6539	155.7824	175.2723	465.90975	165.52785	0.355278785
10424119	NM_010930	Nov	2515.598	2415.481	911.5641	820.9291	2465.5395	866.2466	0.351343028
10436304	NM_001014423	Abi3bp	1088.328	1198.543	402.5055	398.7149	1140.9385	400.7102	0.351211042
10422164	NM_007904	Ednrb	1493.266	1476.403	493.9314	514.1288	1484.8345	504.0301	0.33945204
10373638	NM_146553	Olfr806	1.730693	1.975389	1.249091	0.006305264	1.853051	0.627698132	0.338741301
10492355	NM_008604	Mme	1124.89	1251.758	385.0107	415.9889	1188.324	400.4998	0.337029127
10526441	NM_175309	Upk3b	978.3041	924.0715	322.615	274.4681	951.1878	298.54155	0.313861837
10365983	NM_008524	Lum	3533.349	2765.936	782.4329	872.8568	3149.6425	827.64475	0.312360913
10441093	NM_133659	Erg	1058.393	897.4709	287.2943	282.8025	977.93195	284.8484	0.291071787
10570957	NM_013634	Sfrp1	4208.514	4610.861	1269.831	1103.955	4410.1875	1186.893	0.269125292
10511363	NM_001002927	Perk	4872.767	4688.165	1236.203	1173.404	4680.466	1204.8035	0.257411014
10599001	NM_007429	Agtr2	1468.541	1581.244	340.0576	375.7199	1524.8925	357.88875	0.234697692
	ENSMUST0000011923								
10598085	5	ATP6	5640.785	610.3068	639.0136	701.9263	3125.5469	670.46965	0.214512843
10439500	NM_178924	Upk1b	1241.1	1223.741	279.6418	339.0854	1232.4205	259.3636	0.210450573
10372503	NM_010195	Lgr5	1275.458	1244.944	286.5887	249.1412	1310.201	242.85495	0.185357018

CHAPTER IV: A KINASE INHIBITOR SCREEN IDENTIFIES SMALL MOLECULE ENHANCERS OF REPROGRAMMING AND IPSC GENERATION

4.1 Abstract

Somatic cells can be reprogrammed to iPS cells, a process that suffers from low efficiency and whose molecular mechanisms remain poorly understood. We report an inhibitor screen identifying kinases that enhance or present a barrier to reprogramming. Overall, inhibitors of p38, IPTK and aurora kinases potently enhanced iPSCs generation. iPS cells derived from inhibitor-treated samples were capable of reaching a fully reprogrammed state. Knockdown of target kinases by siRNAs confirmed that these genes function as barriers. We show that Aurora A kinase, which functions in centrosome activity and spindle assembly, is highly induced during reprogramming and inhibits Akt-mediated inactivation of GSK3β, resulting in compromised reprogramming efficiency. Together, our results not only identify new compounds that enhance iPSC generation but provide heretofore unreported insight into the function of Aurora A kinase in the reprogramming process.

4.2 Introduction

Since the discovery of techniques to create cells closely resembling embryonic stem cells, various types of mouse and human somatic cells have been reprogrammed to establish induced pluripotent stem cells (iPSCs) (Meissner et al, 2007; Takahashi et al, 2007; Takahashi & Yamanaka, 2006; Wernig et al, 2007; Yu et al, 2007). These cells have acquired full capacity to differentiate into different lineages (Takahashi et al, 2007; Wernig et al, 2007; Yu et al, 2007). Resultant differentiated cells reportedly function *in vitro* and *in vivo* and serve to correct various diseases in mouse models (Hanna et al, 2007). Moreover, iPSCs have been generated from tissues of patients with different disease conditions and could be a valuable source to study those pathologies or for drug screening in vitro (Itzhaki et al, 2011; Lee et al, 2009; Liu et al, 2011; Marchetto et al, 2010). Nonetheless, the reprogramming process suffers from extreme low efficiency (Aoi et al, 2008; Meissner et al, 2007; Nakagawa et al, 2008; Takahashi & Yamanaka, 2006) .

Currently, there is a need to both better understand molecular mechanisms underlying reprogramming and develop more efficient methods to generate iPSCs. Elegant approaches have been applied to identify pathways regulating reprogramming. For example, mRNA profiling of somatic cells, iPSCs generated from those cells and intermediate populations that emerge during reprogramming indicates that cells can become —trapped" in a partially

reprogrammed state and that treatment with DNA methyl transferase inhibitors enables them to become fully reprogrammed (Mikkelsen et al, 2008). Genome-wide analysis of promoter binding of specific transcription factors supports the idea that DNA-binding and gene activation are altered in partially reprogrammed iPSCs (Sridharan et al, 2009). Moreover, several groups have shown that p53 pathways, which are activated following overexpression of oncogenic reprogramming factors, act as a major reprogramming barrier (Hong et al, 2009; Kawamura et al, 2009; Li et al, 2009a; Utikal et al, 2009). Recent studies show that TGFβ signaling also inhibits reprogramming (Ichida et al, 2009; Maherali & Hochedlinger, 2009) and perturbs the mesenchymal-to-epithelial transition (MET) (Ashford et al, 2010; Garamszegi et al, 2010), a process that enhances reprogramming and is regulated by microRNAs (Li et al, 2011). However, there remains little information about how terminally differentiated cells are reprogrammed to an ES-like state by four transcriptional factors.

Much effort has gone into identifying factors that enhance iPSC derivation. In addition to small molecules that can reportedly replace some reprogramming factors (Ichida et al, 2009; Lyssiotis et al, 2009; Shi et al, 2008a; Yuan et al, 2011), some compounds are known to enhance overall reprogramming efficiency in the presence of the classic four factors (4F), namely, Tgfbr inhibitors, AZA, vitamin C and VPA (Esteban et al, 2010; Huangfu et al, 2008; Maherali & Hochedlinger, 2009; Mikkelsen et al, 2008). Although some investigators report that VPA treatment dramatically enhances iPSC generation, more recent reports

have reexamined the effects of the compound and found them to be modest (Anokye-Danso et al, 2011; Warren et al, 2010; Yusa et al, 2009). Therefore, currently only a limited number of compounds are available to enhance iPSC generation.

Kinases promote phosphorylation of targets by transferring phosphate groups from high-energy donors such as ATP. Kinases regulate many key processes such as cell cycle events and metabolic switching (Lens et al, 2010; Levine & Puzio-Kuter, 2010). However, few kinases have been shown to function in the reprogramming process (Feng et al, 2009). Given their critical function in numerous signaling pathways, we hypothesized that unidentified kinases may modulate the reprogramming process and that iPSC generation might be significantly enhanced by manipulating their activity.

Here we report an inhibitor screen to identify both barrier and essential kinases that function in reprogramming. We found that essential kinases were enriched within cell cycle and proliferation regulators, while three kinases, p38, ITPK and Aurora A kinase, were identified as new barrier genes. Inhibiting their function by small molecules significantly enhanced iPSC generation. iPSCs derived from inhibitor treatments reached a fully reprogrammed state and differentiated into different lineages *in vitro* and *in vivo*. The specificity of these factors was confirmed through analysis using other inhibitors and RNAi knockdown. Moreover, we found that Aurora A kinase functions to inhibit Akt-

mediated GSKβ phosphorylation, which to our knowledge is the first report of a novel function for this kinase in addition to its characterized role in centrosome formation and spindle assembly. GSK3β inhibition has been reported to enhance somatic cell reprogramming efficiency (Silva et al, 2008). Overall, our data provide new insights into mechanisms underlying reprogramming and identify inhibitors that could significantly enhance iPSC generation.

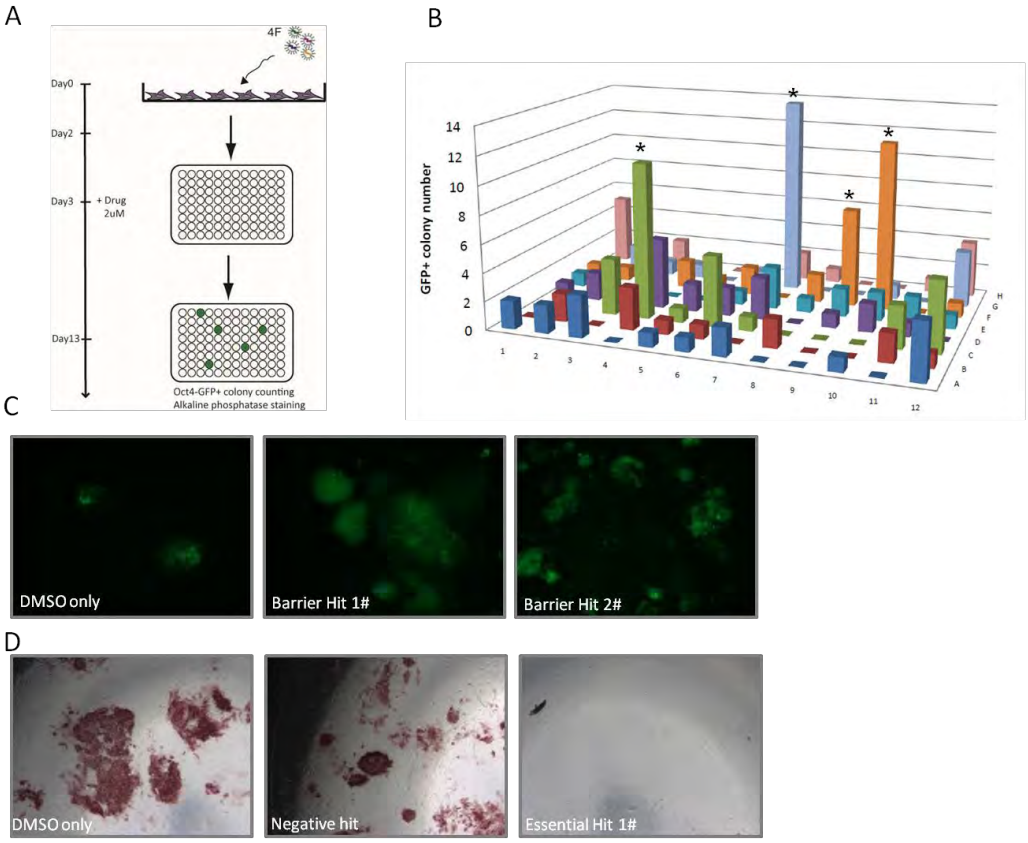
4.3 Results

4.3.1 A kinase inhibitor library screen identifies small molecule activators or inhibitors of iPSC generation

To define signaling mechanisms underlying reprogramming, we undertook a double-blind screen of 244 well-characterized cell-permeable protein kinase inhibitors to identify kinases that enhance or inhibit the process. Mouse embryonic fibroblasts from a transgenic line in which GFP expression is driven by the endogenous *Oct4* promoter were chosen as the starting material. Fully reprogrammed cells switch on endogenous *Oct4* expression, making resultant iPS colonies GFP-positive and enabling us to quantify reprogramming efficiency. To minimize well-to-well variation, MEFs were first transduced with 4F factors in bulk (Figure 4.1.) and then reseeded at 3000 cells/well into gelatin-coated 96-well plates before inhibitor treatment. Starting at day 3 post-transduction, inhibitors were added at a final concentration of 2μM and media were refreshed every other day with mES culture medium plus inhibitor. At day 13 post-transduction,

Figure 4.1. A kinase inhibitor library screen identifies essential and barrier kinases

(A) Design of the screen. MEFs were transduced with the four mouse reprogramming factors (4F) for two days and reseeded into 96-well plates. Drugs were added at a final concentration of 2μM on day 3. Medium was changed every other day until day 13, and cells were then harvested for colony counting and AP staining. (B) Representative plate showing quantification of Oct4-GFP+ colonies. GFP+ colonies were counted directly under a fluorescence microscope, and data was compared with DMSO-treated controls. * indicates identified hits. Columns 1 and 12 indicate control (DMSO) wells. Potential candidates were determined by both GFP+ colony number, morphology, and AP-positivity. (C) Representative barrier kinase hits. Oct4-GFP+ colony numbers were dramatically increased following some drug treatments. GFP+ colonies were quantified and images were taken at day 14 post 4F transduction. (d) Representative hits of essential kinases. Essential hits were identified by loss of AP staining and lack of signs of cell death



plates were fixed in 4% paraformaldehyde and GFP-positive colonies were directly quantified microscopically.

Two columns of wells from each plate (columns 1 and 12) were treated with DMSO (vehicle) only and served as controls. On average, we observed 2-3 GFP positive colonies per well in control samples, which was around 0.07% overall reprogramming efficiency and comparable to other reports (Aoi et al, 2008; Meissner et al, 2007; Nakagawa et al, 2008; Takahashi & Yamanaka, 2006). To identify inhibitors that significantly enhance reprogramming efficiency, we set a minimum of a 2.5-fold increase in GFP-positive colony number as a filter. Using these criteria, we identified eleven inhibitors as potential activators of reprogramming or —barrier hits" (Figure 4.1.B, C, Table 4.1). Since kinases may also be required for iPSC generation, we undertook alkaline phosphatase staining in order to identify potential —essential hits". Since genes encoding targets of essential hits could function at various reprogramming stages, and most cells did not attain a fully reprogrammed state, we used an extremely stringent criterion: only wells devoid of any AP staining and with no obvious decrease in cell number were scored as essential hits (Figure 4.1.D). Based on these standards, nine kinase inhibitors were identified as essential hits, and further analysis revealed that among them (Table 4.2.) were four direct inhibitors of cell cycle dependent kinases (Cdks), indicating that cell cycle control is critical for reprogramming. We tested four of the remaining essential hits (Figure 4.S1.A) and found that three inhibited MEF proliferation to various extents with or without

four-factor transduction (Figure 4.S1. B,C) and that reprogramming efficiency was positively correlated with the extent of that inhibition (Figure 4.S1.D). Overall, these findings suggest that compromised reprogramming efficiencies seen following inhibition of essential kinases are correlated with inhibition of proliferation.

4.3.2 Inhibitors of TGFβ, IP3K, P38 and Aurora kinase significantly enhance reprogramming

To confirm that the 11 compounds (Figure 4.2.A, Table 4.1.) identified in the primary screen targeting potential barrier genes enhance iPSC generation, we undertook a secondary screen using larger wells and two different drug concentrations (1μM and 2μM). For the eleven barrier candidates (Figure 4.2.A), these analyses confirmed that inhibitors B4, B8 and B10 consistently and significantly enhanced reprogramming (Figure 4.2.B) and were even more potent at the lower 1μM concentration in a secondary screen (Figure 4.2.B). Additionally, B8 and B10 enhanced iPSC generation even in non-permissive conditions in which 4F expression was too low to reprogram vehicle-treated MEFs (Figure 4.S2.A). Two other groups recently identified the inhibitor B4 as enhancing reprogramming and/or capable of replacing Sox2 in the 4F cocktail (Ichida et al, 2009; Maherali & Hochedlinger, 2009). Interestingly, inhibitor B6 enhanced reprogramming more robustly at 1μM than at 2μM. Dose/response analyses confirmed that B6, B8 and B10 act as potent enhancers at 0.5μM (Figure 4.S2.B).

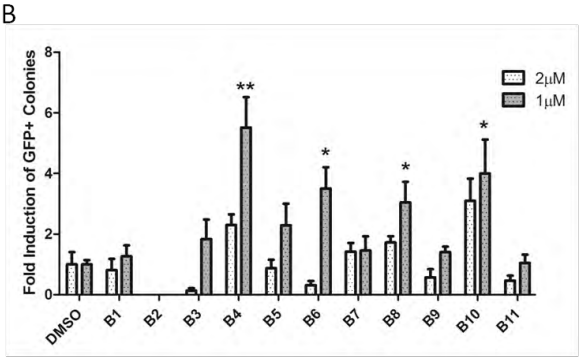
Figure 4.2. Inhibitors of TGFβ, p38, IP3K and Aurora Kinase greatly enhance iPSC generation

(a) Candidate hits for secondary screening conducted in 12-well plates of 11 barrier (B) hits. (b) Compounds B4, B6, B8 and B10 significantly enhanced reprogramming. Drugs were added at day3 at a concentration of 2μM and 1μM and Oct4-GFP+ colonies were quantified at day13 after transduction. Notably, B6 compound showed enhancement only at lower dose (1μM). Data represents three independent experiments. * $p < 0.05$. ** $p<0.01$ (Student t-test). (c) Chemical structures of identified inhibitors.

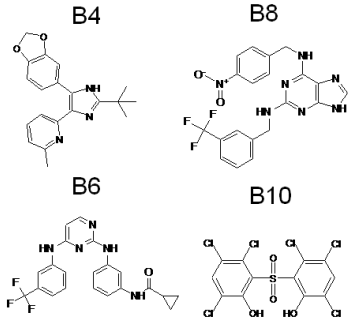
A

ID	Inhibitor	ID	Inhibitor
B1	PP3	B7	KN-62
B2	Alsterpaullone	B8	IP3K
B3	Arcy A	B9	Spring Kinase
B4	TGFb-RI	B10	P38 kinase Inh.
B5	ML-7	B11	Syk Inhibitor III
B6	Aurora Kinase Inh.		

B



C

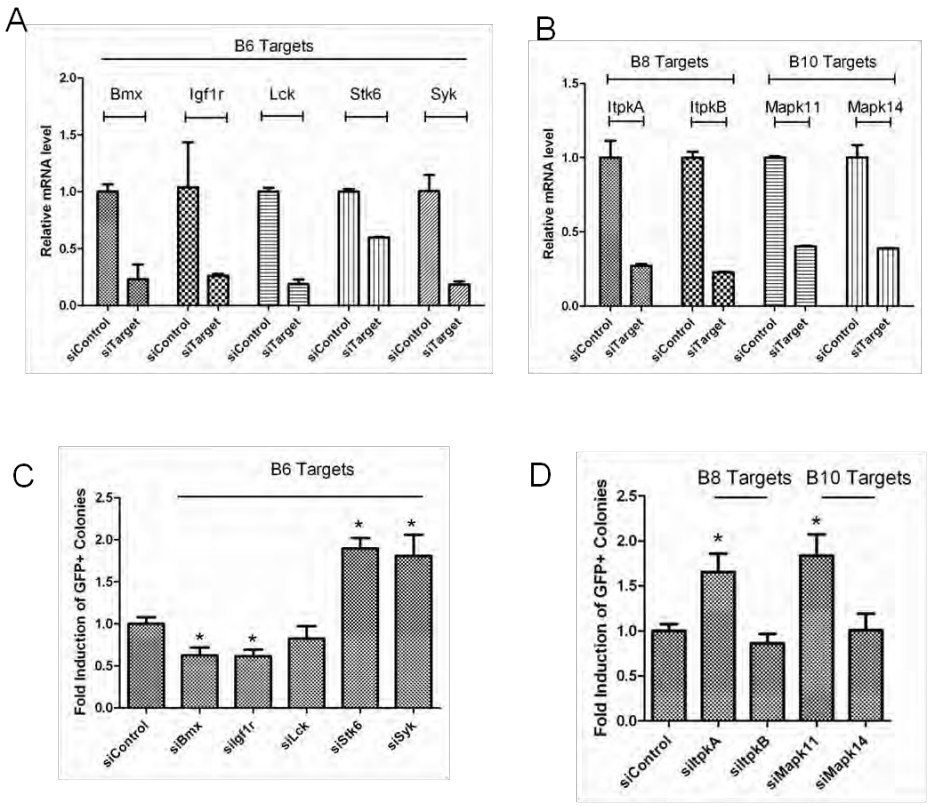


Moreover, treatment of both uninfected and 4F-transduced MEFs with these three inhibitors did not significantly promote proliferation (Figure 4.S3.A, B). Combining B6, B8 and B10 at 1μM each resulted in a synergistic rather than an additive effect of the three compounds (Figure 4.S4.). Since p53 has been identified as a major barrier to reprogramming (Banito et al, 2009; Hong et al, 2009; Kawamura et al, 2009; Li et al, 2009a; Utikal et al, 2009), we asked whether B6, B8 or B10 inhibitors enhanced reprogramming when p53 was down-regulated by shRNA. As expected, p53 knockdown in 4F-transduced MEFs greatly enhanced iPSC generation, but that enhancement was equivalent in inhibitor-treated and -untreated conditions (Figure 4.S5.).

Since B6, B8 and B10 could target multiple kinases at a given concentration, we validated drug specificity by RNAi experiments. To do so, MEFs and mES cells were transfected with siRNAs targeting potential targets of each kinase and knockdown efficiencies were evaluated by RT-qPCR. Indeed, all siRNAs tested efficiently knocked down target genes in MEFs (Figure 4.3.A,B) and in mES cells (Figure 4.3B and data not shown). We next transfected MEFs with these siRNAs and then transduced cells with 4F virus 3 hours post transfection. GFP+ colonies were counted at approximately day 12. Indeed, we found that Mapk11 (p38beta) (a target of inhibitor B10), ItpkA (a target of inhibitor B8), and Stk6 and Syk (targets of inhibitor B6) act as barrier genes: knockdown of any one of these genes during reprogramming resulted in significant increases

Figure 4.3. Inhibitor-targeted kinases are confirmed as barrier genes

(A,B) siRNA knockdown of potential target mRNAs. Potential targets of B6, B8 and B10 inhibitors were knocked down by siRNAs in MEFs or mES (ItpkA only). 50nM siRNAs were transfected and total RNAs was harvested at day2 for RT-qPCR analysis. Data represents analysis of duplicate wells. (C) Knockdown of B6 inhibitor targets enhances reprogramming. Potential targets of B6 were knocked down by siRNAs, and reprogramming efficiency was quantified by counting Oct4-GFP+ colonies. Data from three independent experiments were normalized to siControl-transfected samples. * $p < 0.01$. (D) Knockdown of B8 and B10 targets enhances reprogramming. Data from two independent experiments were normalized to siControl-transfected cells.



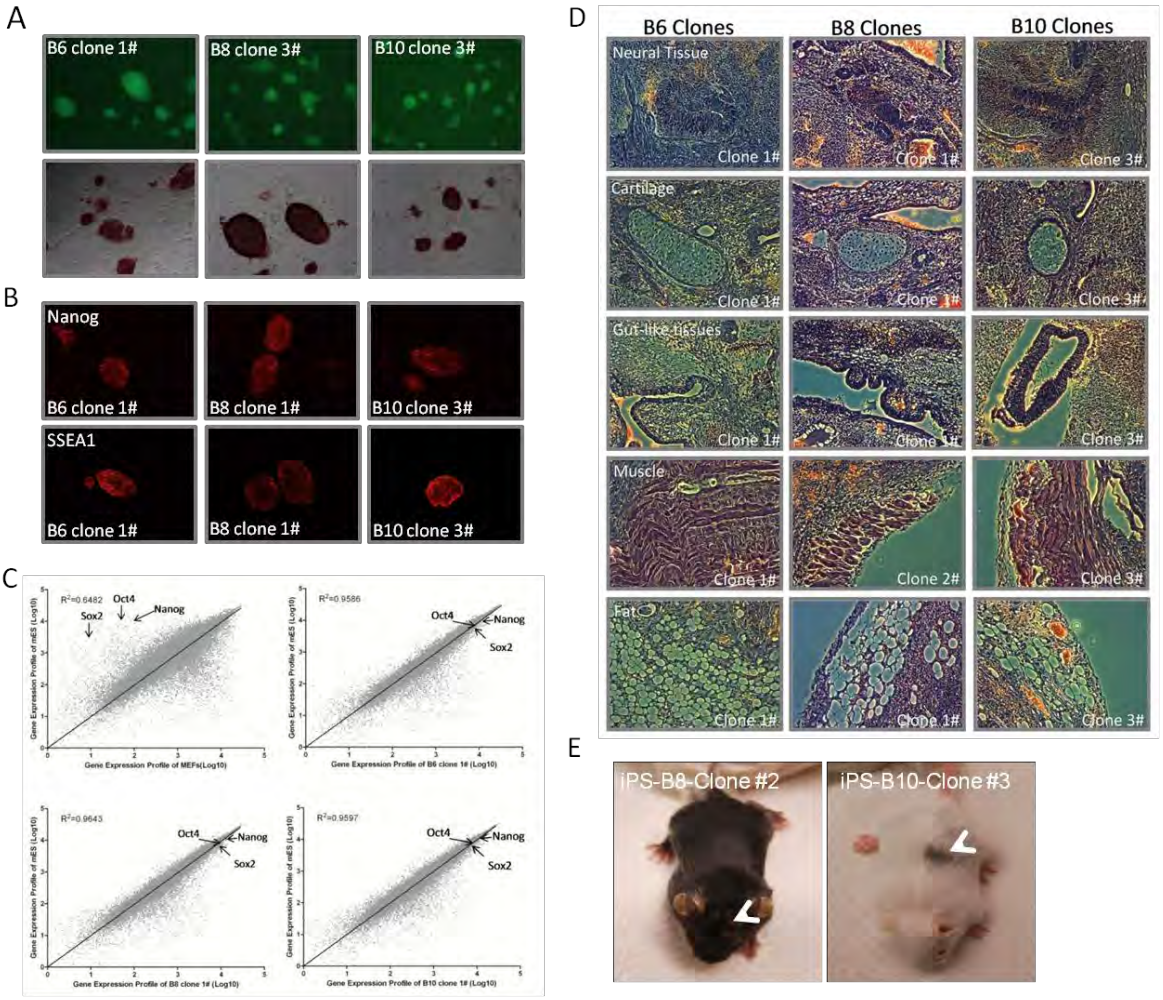
in iPSC generation (Figure 4.3.C, D). Interestingly, knockdown of some B6 targets, such as Bmx, Igf1R and Lck, compromised reprogramming, which may explain in part why B6 both inhibits and enhances reprogramming, depending on concentration. Together, these data confirm that inhibitors B4, B6, B8 and B10 (Figure 4.2.D) are potent enhancers of iPSC generation and that effects of inhibitor treatment are target-specific.

4.3.3 Inhibitor-treated iPSCs reach a fully reprogrammed state

Although B6, B8 and B10 promote reprogramming, it is possible that treatment with these inhibitors turns on endogenous *Oct4* expression but cells do not reach a fully reprogrammed state. To exclude this possibility, iPSCs derived from cells treated with respective inhibitors were analyzed for ES cell marker expression and pluripotency. All GFP+ clones (Figure 4.4.A) also stained positively with alkaline phosphatase (Figure 4.4.A). Immunostaining for other mES self-renewal markers confirmed that these cells expressed Nanog and the mES-specific surface protein SSEA1 (Figure 4.4.B). Moreover, genome-wide mRNA expression profiles verified that these cells showed a gene expression pattern highly similar to mES cells, one that differed significantly from starting MEFs (Figure 4.4.C). To determine whether inhibitor-treated cells acquire the full capacity to differentiate into different lineages, we used *in vitro* embryoid body formation assay to assess pluripotency. Clones tested readily differentiated into

Figure 4.4. Inhibitor-treated iPSCs reach a fully reprogrammed state

(a) Inhibitor-treated iPS cells can be successfully derived. iPS clones from inhibitor-treated samples were picked and expanded. Cells show endogenous Oct4 expression (upper panels) and alkaline phosphatase positivity (lower panels). (B) Inhibitor-treated iPSCs are Nanog- and SSEA1-positive. iPSCs were seeded on irMEF plates and cultured for 3 days, fixed in 4% paraformaldehyde and stained for Nanog and SSEA1 expression (see Supplemental Methods). (C) Genome wide mRNA expression profile of derived iPSCs. Kinase inhibitor-treated iPS clones show an expression profile resembling that of control mouse ES (CCE) cells (R^2>0.95). (D) Inhibitor-treated iPSCs form teratomas in nude mice. iPS clones from each treatment were injected into athymic nude mice and tumors harvested after ~3 weeks. (E) Inhibitor-treated iPSCs can contribute to chimeric mice. Arrow refers to the representative contribution of injected iPSCs to the chimera mice.



three major lineages, including beating cardiomyocytes (Supplemental videos), and stained positively for AFP (endoderm), tubulin III (ectoderm) or cardiac actin (mesoderm) (Figure 4.S6.). As a more stringent test of pluripotency, we injected these iPSCs into athymic nude mice and found that all tested clones generated heterogeneous teratomas within 3-4 weeks (Figure 4.4.D). When iPSCs were injected into the cavity of recipient blastocysts, they successfully integrated with cells of the inner cell mass the next day (Figure 4.7.A,B) and contributed to living chimeric mice (Figure 4.4.E). Cells also contributed to the germline of E13.5 embryos, suggesting that these cells were germline-competent (Figure 4.S7.C). Together, these data strongly suggest that iPSCs derived from inhibitor-treated cells are fully reprogrammed and can differentiate into all lineages *in vitro* and *in vivo*.

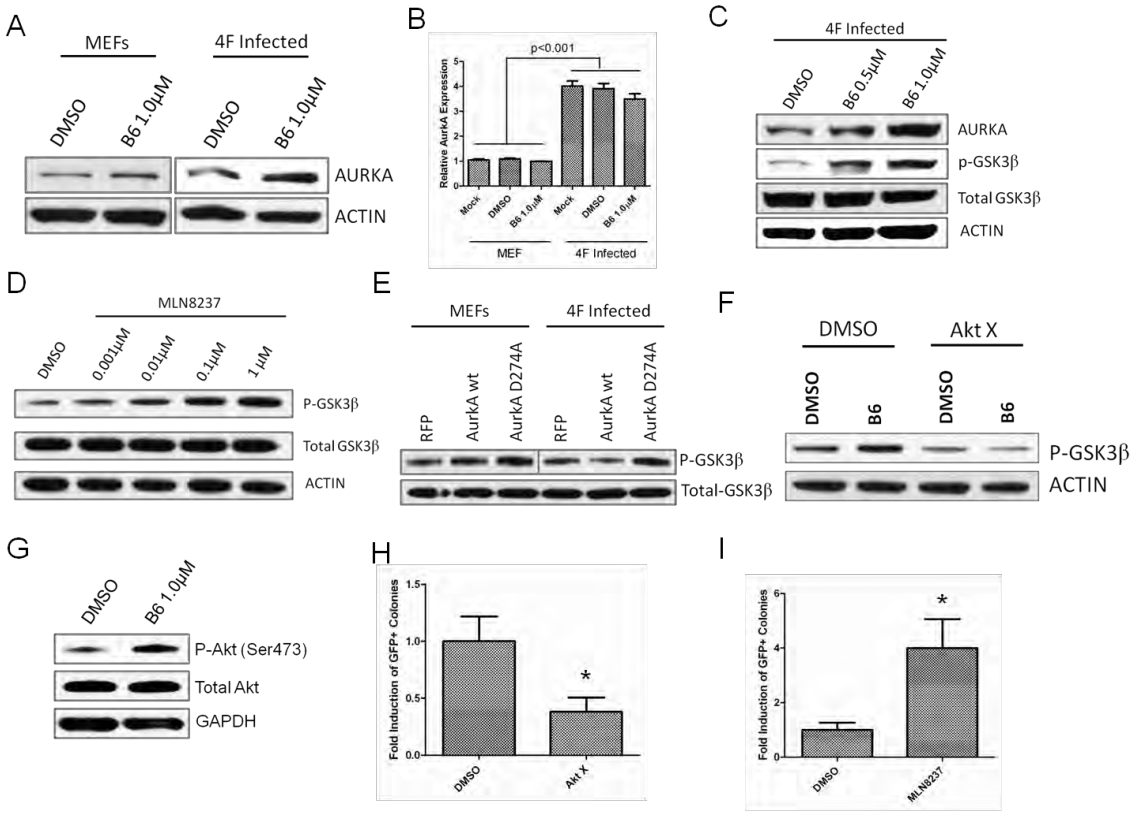
4.3.4 AurkA inhibition by B6 enhances Akt-mediated GSK3β inactivation

To identify the mechanism underlying enhanced reprogramming mediated by a kinase inhibitor, the activity of a target of inhibitor B6, Aurora A Kinase (AurkA), was further analyzed due to its well known function in cell cycle progression, spindle formation and tumor development (Lens et al, 2010). We first determined whether treatment with inhibitor B6 altered levels of AurkA protein. B6 treatment of both wild type MEFs and 4F-infected MEFs resulted in increased AurkA protein levels relative to untreated cells (Figure 4.5.A). In addition, a significant increase in AurkA protein levels was seen in untreated 4F-infected cells relative

Figure 4.5. Aurora A kinase inhibition by B6 promotes Akt mediated inactivation of GSK3β

(A) B6 treatment increases aurora A kinase protein levels. Both 4F-infected (day3) and mock MEFs were treated with 1μM B6 for 2 days and cells were harvested for western blotting of Aurora A kinase. Actin served as loading control. (B) Aurora A kinase mRNA levels are signficantly increased by 4F expression during reprogramming. MEFs were infected with 4F for 3 days and treated under mock conditions or with DMSO or 1μM B6 for 2 days prior to RT-qPCR analysis of total RNA. B6 treatment did not alter induction of Aurora A kinase by 4F. (C) Inhibition of Aurora A kinase promotes increased phosphorylation of GSK3β kinase. 4F-infected MEFs were treated with different doses of B6 inhibitor starting at day 3 post-infection for 48 hrs before being harvested for western blotting analysis. (D) Inhibition of Aurora A kinase by MLN8237 promotes GSK3β phosphorylation dose-dependently. The experiment was the same as (C). Actin served as the loading control. (E) Expression of a dominant negative form of Aurora A kinase promotes GSK3β phosphorylation (Ser9). MEFs were infected with 4F and expression vectors for RFP, wild-type (wt) Aurora A kinase or the D274A kinase-dead mutant of human Aurora A kinase. Expression of wt AurkA inhibited GSK3β phosphorylation (Ser9), while overexpression of the mutant promoted the process. Exposure time was almost doubled for 4F-infected samples. (F) Phosphorylation of GSK3β (Ser9) is mediated by Akt. 4F-infected MEFs were treated with 1μM each of Akt inhibitor (Akt X) and B6. GSK3β phosphorylation (Ser9) (Ser9) was diminished likely due to Akt inhibition. (G) Inhibition of AurkA by B6 induces Akt phosphorylation (Ser473). MEFs were infected with 4F for 3 days and then treated with 1μM B6 compound for 2 more days before harvesting for western blotting analysis. (H)Akt X treatment compromises reprogramming. Akt X inhibitor was added at a final concentration of 1μM. Error bar represents standard deviation of results derived from triplicate wells. *p <0.05. (I) AurkA inhibition by MLN8237 enhances reprogramming. 4F-infected MEFs at day 3 were treated with 10nM MLN8237 for 10 days and GFP+ colonies counted to determine reprogramming efficiency. MLN8237 enhanced iPSC generation similarly to effects seen with 1μM B6. Data is derived from two experiments using triplicate wells. * $p<0.05$

Figure 4.5. Aurora A kinase inhibition by B6 promotes Akt mediated inactivation of GSK3β



to untreated MEFs. Further experiments suggested that this increase could be due to enhanced transcription, since AurkA mRNAs were induced by 3~4 fold in mock 4F-infected cells relative to mock MEFs (Figure 4.5.B), and mRNA levels were not significantly altered by B6 treatment (Figure 4.5.B). These results agree with previous expression profiling studies of MEFs expressed in mES or iPSCs compared with MEFs. During reprogramming with 4F we also observed decreased levels of phosphorylation of GSK3β (data not shown), indicating that GSK3β is activated. Interestingly, recent studies indicate that GSK3β inhibition by small molecules enhances iPSC generation from neural stem cells (Silva et al, 2008). Therefore, we asked whether AurkA inhibition by B6 altered GSK3β phosphorylation. Indeed, we detected a significant increase in phospho-GSK3β in B6-treated cells (Figure 4.5.C), while GSK3β total protein levels were unchanged. As a test of specificity, we assessed reprogramming in the presence of a different AurkA inhibitor MLN8237, which has a potent inhibitory effect and has been tested in myeloma cell lines (Gorgun et al, 2010). AurkA inhibition by MLN8237 promoted a dose-dependent increase of phospho-GSK3β, while total GSK3β levels remained unchanged (Figure 4.5.D). Increased AurkA protein levels were also detected (Figure 4.S8.). Overall, these results indicate that AurkA inhibition promotes phosphorylation and subsequent inactivation of GSK3β, an effect that likely enhances reprogramming.

AurkA reportedly has a kinase-independent function (Otto et al, 2009). We therefore asked whether potential inactivation of GSK3β by AurkA proteins

requires AurkA kinase activity. To answer this question, we overexpressed AurkA in 4F-infected MEFs using two retroviral constructs: one encoding wild-type mouse AurkA and the other encoding a human AurkA kinase dead mutant D274A (Otto et al, 2009) (Figure 4.S9., as AurkA is highly conserved between humans and mouse (Figure 4.S10.)). If GSK3β inactivation seen following B6 treatment is AurkA kinase-independent, overexpression of either construct should decrease GSK3β phosphorylation. If that effect is kinase-dependent, overexpression of the kinase-dead mutant should have a dominant-negative effect, similar to effects seen following AurkA inhibition by small molecules. Indeed, a significant increase in levels of phospho-GSK3β was detected in both 4F- infected and untreated MEFs following expression of the kinase-dead mutant (Figure 4.5.E). Moreover, overexpression of wild-type AurkA promoted a decrease in phospho-GSK3β without altering total protein levels (Figure 4.5.E). Furthermore, AurkA knockdown by siRNAs in MEFs enhanced GSK3β phosphorylation (Figure 4.S11.). Meanwhile, overexpression of GSK3β in MEFs largely abolished the enhancing effect of B6 (Figure 4.S12.). Collectively, these findings suggest that inhibition of AurkA kinase activity promotes GSK3β inactivation.

Next, we asked which kinase potentially functions in GSK3β phosphorylation following B6 treatment. Since Akt kinases are well-characterized mediators of GSK3β phosphorylation, we asked whether Akt inhibitors abolished B6‘s effect on phospho-GSK3β induction. Indeed, treatment of 4F-infected MEFs

by small molecule Akt inhibitor decreased the effect of B6 on GSK3β phosphorylation (Figure 4.5.F) and inhibited reprogramming (Figure 4.5.H). Moreover, an increase of phosphorylated Akt (Ser473) was also detected in B6 treated cells (Figure 4.5.G), suggesting that increased GSK3β phosphorylation seen following AurkA inhibition is mediated by Akt. We also tested expression levels of several genes reportedly upstream of Akt such as Pdk1, Src, Pten and p85α but did not observe significant changes when MEFs were first transduced with 4F for three days and then treated with DMSO or B6 (Figure 4.S13.), indicating that B6 treatment does not alter transcription of Akt regulators.

As a test of specificity, we also determined whether MLN8237 enhanced iPSC generation. Treatment of MEF cells with MLN8237 at the low concentration of 10nM enhanced reprogramming approximately 4-fold, and the effect was dose-dependent (Figure 4.5.I, Figure 4.S14.A). Gene expression analysis of MLN8237-treated samples also confirmed increases in mES-specific gene expression (Figure 4.S14.B)

Since AurkA functions in control of spindle formation and the cell cycle (Lens et al, 2010), we determined whether the cell cycle of 4F-infected MEFs was significantly altered by treatment with AurkA inhibitors. Treatment with various concentrations of either B6 or MLN8237 promoted little change in cell cycle progression (Figure 4.S15.). However, at a higher concentration of 100nM, MLN8237 treatment increased the number of cells in G2. However, a significant

increase in the number of GFP+ colonies accompanied by induction of an mES-specific gene expression profile was still detected at this concentration of MLN8237 (Figure 4.S14.A,B), and the size of resultant colonies resembled that of DMSO-treated cells (data not shown), indicating that iPSC formation was not affected by MLN8237 at this dose, in contrast to fibroblast cells which would have undergone G2-arrest under such dose.

AurkA is highly expressed in mES and iPS cells compared with MEFs (Sridharan et al, 2009), suggesting that it functions to maintain mES self-renewal or pluripotency. To determine whether inhibition of AurkA altered mES self-renewal or differentiation, we treated mES cells with B6 at both 0.5μM and 1μM and cultured cells in LIF+ and LIF- conditions for 4 days. LIF withdrawal in both B6-treated and DMSO control cells promoted mES cell differentiation, as indicated by loss of colonies, based on morphology and AP staining (Figure 4.S16.A). RT-qPCR of self-renewal markers confirmed that differentiation was occurring, as those markers were down-regulated in LIF-minus cells (Figure 4.S16.B). However, we did not see a significant effect on either mES self-renewal (in the presence of LIF) or differentiation (following LIF withdrawal), other than a very small increase in Oct4 expression in mES cells, suggesting that B6 treatment has little effect on iPS cells once they have reached the fully reprogrammed state.

4.4 Discussion

Since the discovery of reprogramming of fibroblasts to iPS cells, much effort has been put into overcoming the extreme low efficiency of the process. A few small molecules have been shown to replace some 4F reprogramming factors in large-scale random screens (Feng et al, 2009), while only a handful—most of which are chromatin remodeling reagents—have been shown to enhance iPSC generation in 4F-infected cells. To date, TGFβ receptor inhibitors are the only kinase inhibitors shown to be capable of directly enhancing reprogramming and replacing Sox2 and cMyc by inducing Nanog expression (Ichida et al, 2009), an observation that led to the discovery that the MET is a key event during early reprogramming stages (Ashford et al, 2010; Garamszegi et al, 2010). Thus, identifying kinases functioning in the course of reprogramming could provide not only targets that could be modulated but also provide novel insight into how reprogramming works.

Here we report a kinase inhibitor library screen aimed at identifying additional kinases important for reprogramming. We found that inhibition of P38, IP3K and AurkA significantly enhanced reprogramming efficiency, indicating that these kinases could function as barriers to the process. Modulation of activities of these kinases possibly in combination with other currently available methods could substantially increase reprogramming efficiency. Interestingly, knock-down of p53 seemed to override the enhancing effects of these kinase inhibitors (Figure 4.S5.B). One potential reason could be that establishment of fast ES-like cell cycle could be most fundamental requirement to reach fully reprogrammed

state. Kinase inhibitor treatments may all lead to enhanced cell cycle progression among small percentage of the cells, which could not be easily detected in mixed population. Meanwhile, knockdown of p53 leads to a major release of cell cycle arrest caused by oncogene overexpression, which then largely tampered the enhancing effects of these inhibitors.

Our experiments also identified novel functional aspects of AurkA, whose kinase activity may inhibit Akt-mediated phosphorylation of GSK3β, which needs to be inactivated to promote iPSC generation (Silva et al, 2008). AurkA kinase is well characterized for its role in modulating centrosome function and spindle assembly(Lens et al, 2010). Aberrant expression of AurkA, either overexpressed or reduced, also reportedly leads to tumor development (Lens et al, 2010). We show that during reprogramming of MEFs to iPS cells, AurkA is highly induced even at an early stage (~day 5 post transduction) (Figure 4.5.B), an event correlated with reduction of phospho-GSK3β in these cells (Figure 4.5.E). Modulating AurkA kinase activity could thus affect GSK3β activity and alter reprogramming efficiency. Meanwhile, treatment with AurkA inhibitors could increase levels of AurkA protein. Recent studies indicate that AurkA may have a kinase-independent function, such as stabilizing N-MYC protein by direct binding to block MYC ubiquitination (Otto et al, 2009). N-MYC is also specifically expressed in mES or iPS cells, and recent work confirms that levels of endogenous N-MYC increase in reprogramming (Sridharan et al, 2009).

Interestingly, N-MYC degradation also requires sequential phosphorylation by cyclin B/Cdk1 and GSK3 (Otto et al, 2009).

Our screen also identified p38 and IP3K as barrier kinases. p38 reportedly regulates diverse processes, including the stress response, chromosome remodeling and the cell cycle (Wagner & Nebreda, 2009). Interestingly, p38 has been shown to have tumor suppressor function, and one target regulated by p38 is p53 (Wagner & Nebreda, 2009). This observation could explain why p53 knockdown abolished the enhancing effect of a p38 inhibitor (Figure 4.S5.). Meanwhile, p38 could also negatively regulate cell cycle progression (Wagner & Nebreda, 2009). Although we did not detect growth effects following inhibitor treatment of 4F-infected cells, we cannot rule out the possibility that a small percentage of cells gain a proliferative advantage following inhibition of p38, since very few cells reach a fully reprogrammed state. By contrast, IP3K is the least studied protein identified here as a barrier kinase. Gene expression profiles indicate that MEFs express low levels of IP3K and its expression is induced in partially reprogrammed iPS cells and iPS/mES cells (Sridharan et al, 2009). IP3K functions primarily in calcium-dependent signal transduction (Xia & Yang, 2005) and its relationship to reprogramming requires further investigation.

We also identified kinases functioning as potential enhancers of reprogramming. Specifically, we found that knockdown of the insulin like growth factor (IGF) receptor Igf1r compromised reprogramming (Figure 3c). Interestingly,

IGF signaling reportedly activates the PI3K pathway, which could also activate Akt function (Zoncu et al, 2011). We have also found that knockdown of negative regulators of IGF signaling enhances reprogramming (data not shown). These findings suggest an important role for Akt function in iPSC generation. Although we found that the kinase inhibitors could enhance iPSCs generation and the reprogrammed cells reached fully pluripotent state, but the exact nature of cells population to increase iPSCs was not obvious. For example, expression of four factors (OSKM) can lead a cell population to the path of pluripotency but not all of them reach fully iPSC state because a large number of unstable or partially reprogrammed cells never go over the barrier to fully reprogrammed state. Do the kinase inhibitors accelerate the process where unstable cells become fully reprogrammed or increase the starting pool of initiator cells for iPSC generation? We do not know the answers to these questions at this stage and future work would reveal these aspects of reprogramming stages.

Overall, our findings provide new insights into how somatic cells are reprogrammed into iPS cells and have identified new barrier genes that could serve as targets to design specific chemical inhibitors. Our study encourages further efforts to screen for small molecules that could prove useful in iPS cell-based therapies.

Materials and Methods

Cell culture, vectors and virus transduction

Oct4-GFP MEFs were derived from mice carrying an IRES-EGFP fusion cassette downstream of the stop codon of *pou5f1* (Jackson lab, Stock#008214) at E13.5. MEFs were cultured in DMEM (Invitrogen, 11995-065) with 10% FBS (Invitrogen) plus glutamine and NEAA. Only MEFs at passage of 0 to 4 were used for reprogramming. pMX-Oct4, Sox2, Klf4 and cMyc were purchased from Addgene. Mouse AurkA was cloned into pMX. The human AurkA D274A mutant retroviral vector was purchased from Addgene. To generate retrovirus, PLAT-E cells were seeded in 10cm plates, and 9ug of each factor were transfected the next day using Lipofectamine (Invitrogen, 18324-012) with PLUS (Invitrogen, 11514-015). Viruses were harvested and combined 2 days later. For reprogramming, MEFs were seeded in 12-well plates and transduced with 4F virus the next day with 4ug/ml Polybrene. One day later, the medium was changed to fresh MEF medium, and 3 days later it was changed to mES culture medium supplemented with LIF (Millipore, ESG1107). GFP+ colonies were picked at day 14 post-transduction, and expanded clones were cultured in DMEM with 15% FBS (Hyclone) plus LIF, thioglycerol, glutamine and NEAA. Irradiated CF1 MEFs served as feeder layers to culture mES cells and derived iPS clones.

Kinase library screening

A kinase library of 244 compounds was obtained from the chemical screening facility at the Sanford-Burnham Institute. The library was purchased from Calbiochem (Library 1: 80 compounds. Cat# 539744-1EA; Library 2: 80 compounds. Cat# 539745-1EA; Library 3: 84 compounds. Cat# 539746-1EA). All compounds are well-characterized protein kinase inhibitors.

Compounds were diluted to 2mM in 96-well plates. 4F-transduced cells were seeded into gelatin coated plates (4000 cells/well). Inhibitors were added every other day until day 13. Cells were then fixed with 4% paraformaldehyde for 20 min at room temperature and number of Oct4-GFP+ colonies was directly counted under a microscope. Cells were then stained with Vector red alkaline phosphatase substrate kit I (Vector laboratories, SK5100).

siRNA transfection of MEFs

siRNAs were purchased from Dharmacon and diluted in Opti-MEM (Invitrogen, 11058-021) to the desired final concentration. Lipofectamine 2000 (Invitrogen, 11668-019) was added to the mix at 2ul/well in 12-well plates, which were incubated 20 min at RT. For 12-well transfections, 80ul of the siRNA/lipid mixture was added to each well with 320ul Opti-MEM. Three hours later, 0.8ml of the virus mixture (for iPS) or fresh medium was added to each well and the medium was changed to fresh MEF medium the next day. siRNAs were transfected twice during reprogramming (at days 0 and 5 post-4F infection).

Western blotting

Total cell lysates were prepared by incubating cells in MPER buffer (PIERCE, 78503) on ice for 20 min, and then cleared by centrifugation at 13,000 rpm for 10 min. An equal volume of lysates was loaded onto 10%SDS-PAGE gels, and proteins were transferred to PVDF membranes (Bio-Rad, 1620177) using the semi-dry system (Bio-Rad). Membranes were blocked with 5% milk in TBST for at least 1hr at room temp or overnight at 4°C. Antibodies used include: anti-mNanog (R&D, AF2729), anti-h/mSSEA1 (R&D, MAB2156), anti-Actin (Thermo, MS1295P0), anti-AFP (Abcam, ab7751), anti-Beta III tubulin (R&D systems, MAB1368), and anti-alpha actinin (Sigma, A7811), anti-mAurkA (Bethyl lab, A300-072A), anti-hAurkA (Bethyl lab, A300-071A), total-GSK3β (Cell signaling technology, 9315S), phospho-GSK3β (Ser9) (Cell Signaling Technology, 9323S), total Akt (Cell signaling technology, 9272S), phosphor-Akt (Ser473) (Cell signaling technology, 9271S).

mRNA quantitative PCR

Total RNAs were extracted using Trizol (Invitrogen). After extraction, 1ug total RNA was used for RT using Superscript II (Invitrogen). Quantitative PCR was performed using a Roche LightCycler480 II and the Sybr green mixture from Abgene (Ab-4166). Gene primers are listed in Supplementary Table 3. Other primers were previously described (Takahashi & Yamanaka, 2006).

Immunostaining

Cells were washed twice with PBS and fixed with 4% paraformaldehyde at room temperature for 20min. Fixed cells were permeabilized with 0.1% Triton X-100 for 5min. Cells were then blocked in 5% BSA in PBS containing 0.1% Triton X-100 for 1hr at room temperature. Primary antibodies were diluted from 1:100 to 1:400 in 2.5% BSA PBS containing 0.1% Triton X-100, according to the manufacturer‘s suggestion. Cells were stained with primary antibody for 1hr and then washed three times with PBS. Secondary antibodies were diluted 1:400 and cells were stained for 45min at room temperature.

EB formation and differentiation assay

iPS cells were trypsinized into a single cell suspension and the hanging drop method was used to generate embryoid bodies (EB). For each drop, 4000 iPS cells in 20ul EB differentiation medium were used. EBs were cultured in hanging drops for 3 days before being reseeded onto gelatin-coated plates. After reseeding, cells were further cultured until day 14, when beating areas could be identified.

Teratoma formation and chimera generation

iPS cells were trypsinized and resuspended at $1x10^7$ cells/ml. Athymic nude mice were anesthetized with Avertin, and then approximately 150 ul of the cell suspension was injected into each mouse. Mice were checked for tumors every week for 3 to 4 weeks. Tumors were harvested and fixed in zinc formalin solution for 24hrs at room temp before paraffin embedding and H&E staining. To test the capacity of derived iPSC clones to contribute to chimeras, iPS cells were injected into C57BL/6J-$Tyr^{(C-2J)/J}$ (albino) blastocysts. Generally, each blastocyst received 12-18 iPS cells. ICR recipient females were used for embryo transfer. Donor iPSCs confer agouti or black coat color.

mRNA microarray analysis

Total RNAs from derived iPSCs were harvested and extracted by Trizol method. mRNA microarray analysis was carried out by the microarray facility at the Sanford-Burnham Institute. A scatter plot was used to compare genome wide mRNA expression profiles between iPSCs, MEFs and mES cells.

Cell proliferation assay

3000 MEFs were seeded in each well in 96-well plates and transduced with 4F virus for three days. Cells were then treated with inhibitors at 0.5μM and harvested every other day. Cells were incubated with mES medium containing

Celltiter 96 Aqueous one solution (Promega, G3580) for 1hr in tissue culture incubator. Absorbance at 490nm was then measured for each well using a plate reader, and data was used to generate relative proliferation curves using the signal from day 3 post-transduction as a reference.

Cell cycle analysis

MEFs or 4F-infected MEFs were treated with inhibitors for two days and then harvested and trypsinized before fixing in 75% ethanol overnight. Cells were centrifuged at 1000 rpm for 5min, washed once with PBS, and treated with PI staining solution for at least 30min at room temperature before flow cytometry analysis. ~20000 events per sample were collected per analysis and cell cycle data was modeled using ModFit.

Acknowledgments

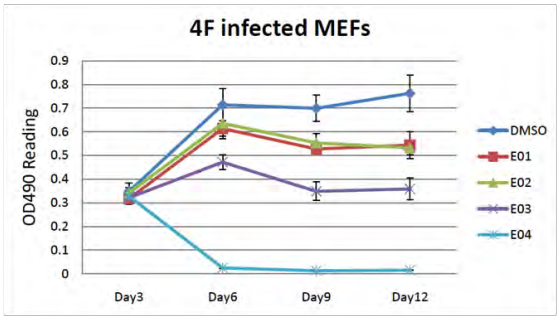
We thank Mark Mercola, Evan Snyder, and Rana laboratory members for helpful discussions and support. We are grateful for the following shared resource facilities at the Sanford-Burnham Institute: the Conrad Prebys Center for Chemical Genomics for providing kinome inhibitors; the genomics and informatics and data management core for mRNA array experiments and data analysis; the animal facility for generation of teratomas and chimera mice; and the histology and molecular pathology core for characterization of various tissues. This work was supported in part by an NIH grant to TMR.

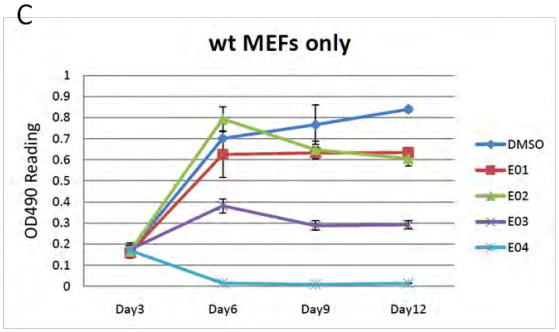
Figure 4.S1. Essential hits can block the reprogramming process

(A) List of essential hits that are not Cdk inhibitors. (B) Proliferation of wild-type MEFs is altered by inhibitor treatment. MEFs were treated with drugs at 2μM and proliferation was assayed by using the Celltiter96 aqueous method (Promega). Y-axis represents the absorbance at 490nm to detect formazan, which is converted from MTS tetrazolium by living cells. A higher reading indicates increased cell number. (C) Proliferation of 4F-transduced MEFs was also altered by drug treatment. (D) Reprogramming efficiency was decreased by drug treatment. Oct4-GFP+ colonies were quantified at day13. *p<0.05. **p<0.01 (Student t-test).

A

ID	Inhibitor
E1	AGL2043
E2	Rho Kinase Inhibitor
E3	Reversine
E4	PI3-K





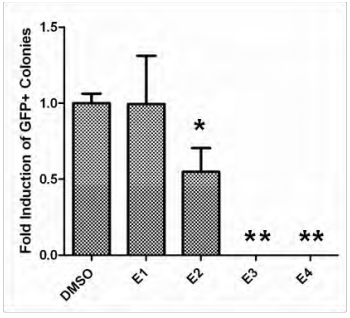


Figure 4.S2. Dose/response analysis of kinase inhibitor hits

(A) Inhibitor B8 and B10 enhanced iPSC generation in non-permissive conditions in which 4F expression was too low to reprogram vehicle-treated MEFs. Cells were infected with 4F and drugs were added at days 3 post-infection. GFP+ colonies were counted at day 12-13. Non permissive condition refers to occasions where 4F expression level does not reach to the threshold for successful reprogramming thus no GFP+ colonies could be detected. Data represents experiments using duplicate wells for each treatment. (B) 4F-transduced MEFs were treated with indicated concentrations of kinase inhibitors B6, B8 or B10 or with DMSO control starting at day3 post 4F transduction, and Oct4-GFP+ colonies were counted at day 13 after transduction. Data represents three independent wells.

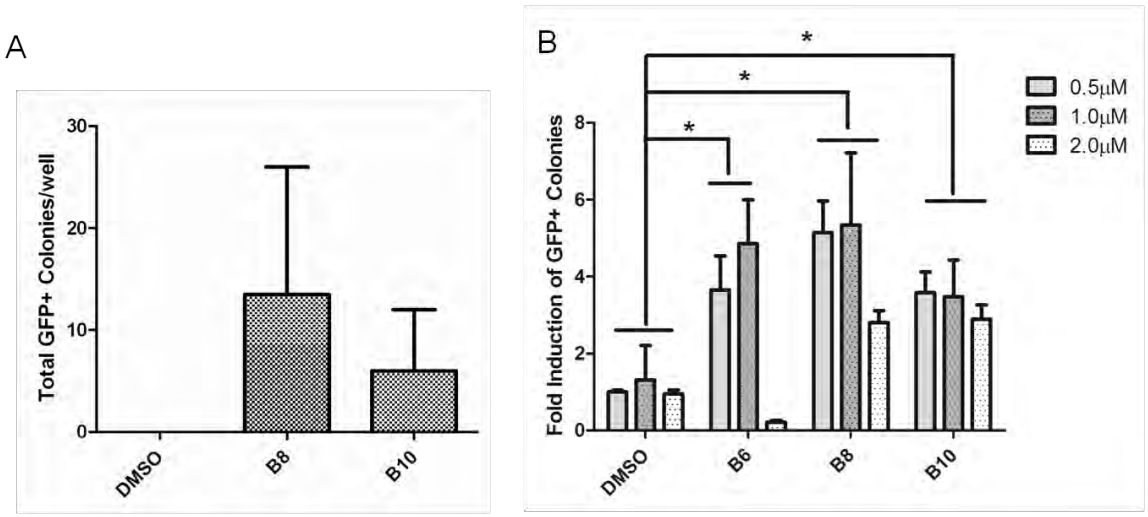
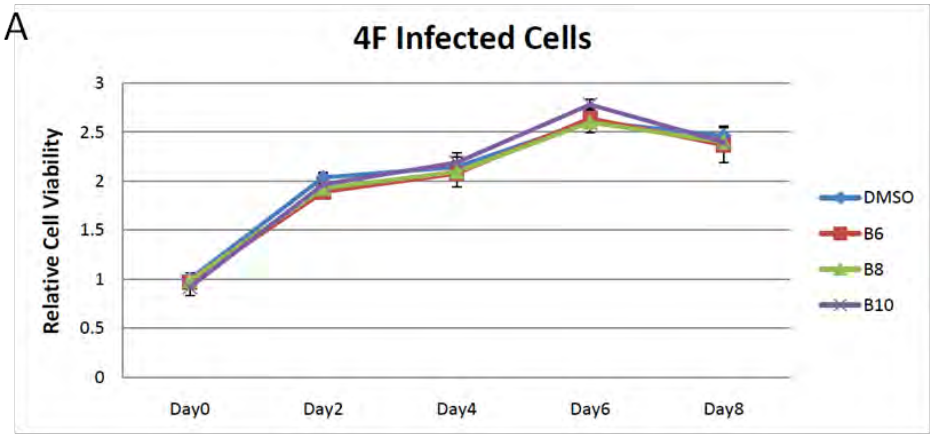


Figure 4.S3. Identified barrier hits do not alter proliferation of 4F-infected MEFs

(A) 4F-infected MEFs were seeded in 96-well plates and 0.5μM of three inhibitors was added at day 3 after transduction (Day 0 in dataset). Proliferation of cells was analyzed every other day using the Celltiter96 aqueous method (Promega). (B) Proliferation of uninfected MEFs is not altered by inhibitor treatment. The experimental procedure is the same as in (A).



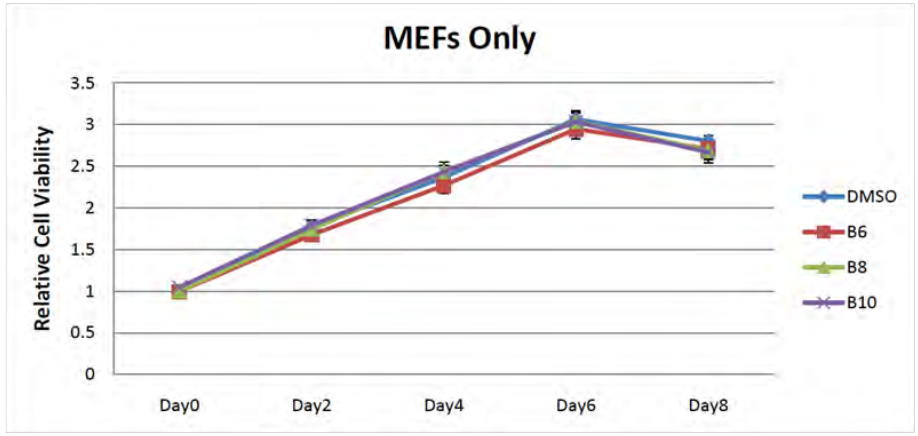


Figure 4.S4. A combination of three inhibitors enhances iPSC generation

A combination of B6, B8 and B10 (Combined) enhances reprogramming. The three inhibitors were used at 1μM each. Data is derived from experiments using triplicate wells. A GSK3β inhibitor (CHIR99021) was also included to evaluate its effect on reprogramming. Error bar represents two experiments with three independent wells. **p <0.01. *p<0.05 (Student t-test).

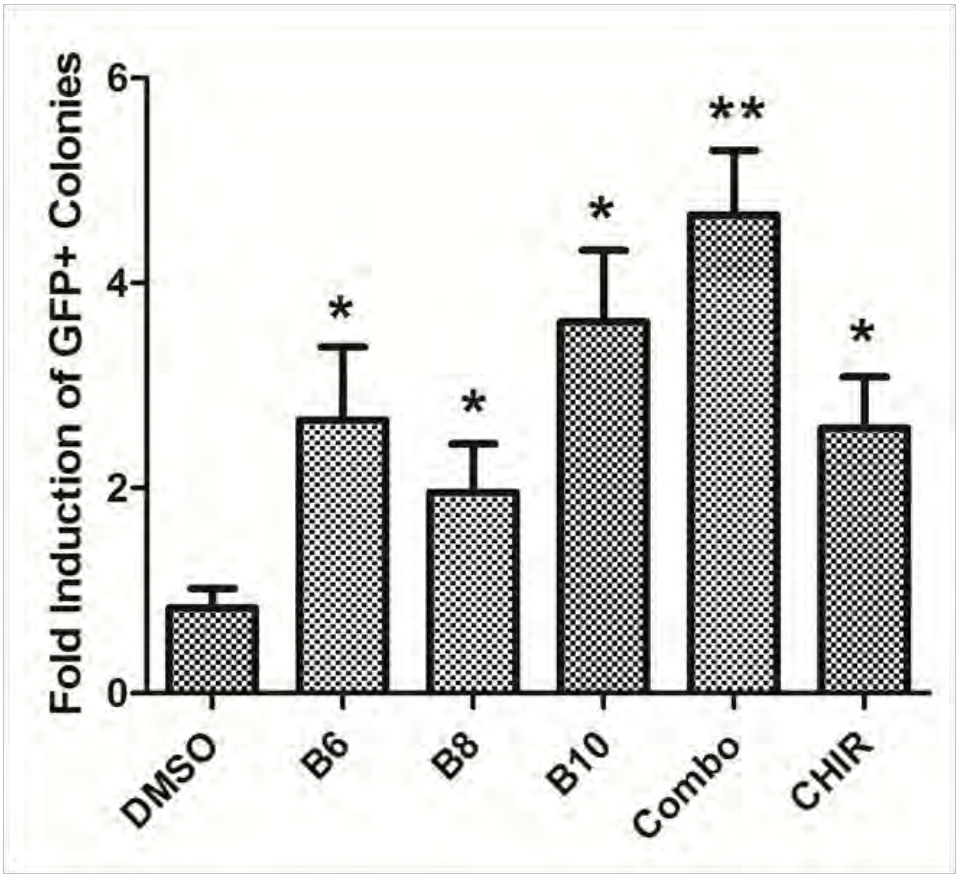


Figure 4.S5. Kinase inhibitors' effect on iPSC generation when p53 is silenced by RNAi

(A) p53 was efficiently knocked down by shRNA. MEFs were transduced by pLKO-shp53 and nontargeting control and harvested for RT-qPCR analysis at day4 post transduction. Error bar represents experiment with duplicate wells. (B) Enhancements of inhibitors in reprogramming were largely abolished in p53-knocked down samples. MEFs were infected with 4F plus either non-targeting or p53 shRNA. Inhibitor treatments were started at day 3 post-infection. Media were changed every other day, and GFP+ colonies were counted at day 12 post-infection.

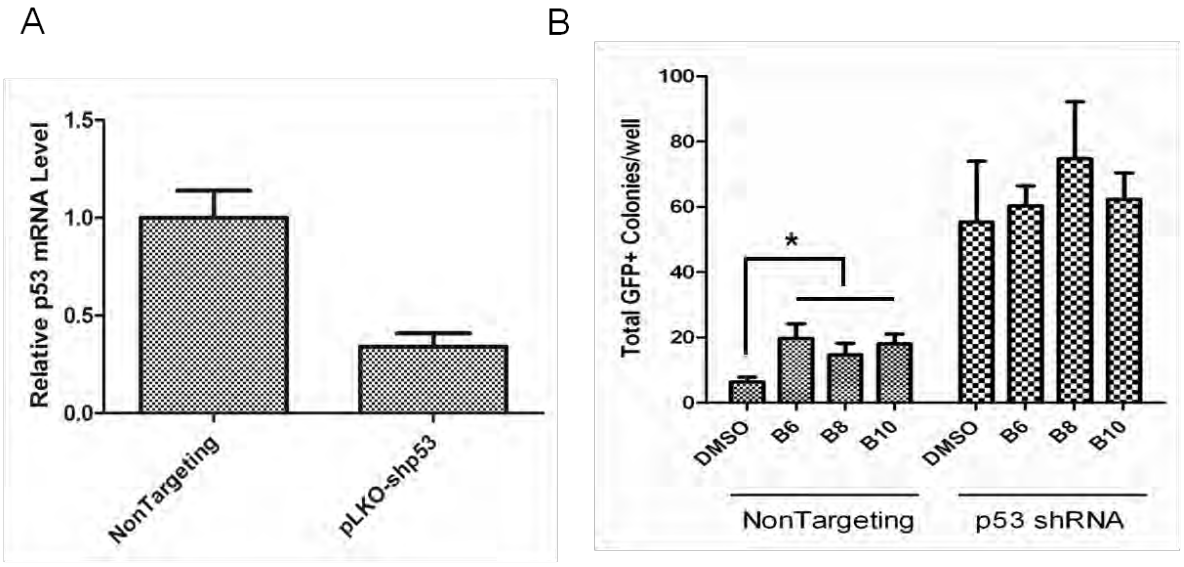


Figure 4.S6. Inhibitor-treated iPSCs can differentiate into tissues representing all three germ layers

iPS clones were picked and expanded from inhibitor-treated cells and used for embryoid body (EB) formation. EBs were formed by the hanging drop method with ~2000 cells/20ul drop. After three days, EBs were transferred to gelatin-coated plates and cultured until day 13, when beating areas were apparent (see Supplemental Videos). Cells were fixed with paraformaldehyde and stained with indicated antibodies. AFP, tubulin III and cardiac actin mark endoderm, ectoderm, and mesoderm, respectively.

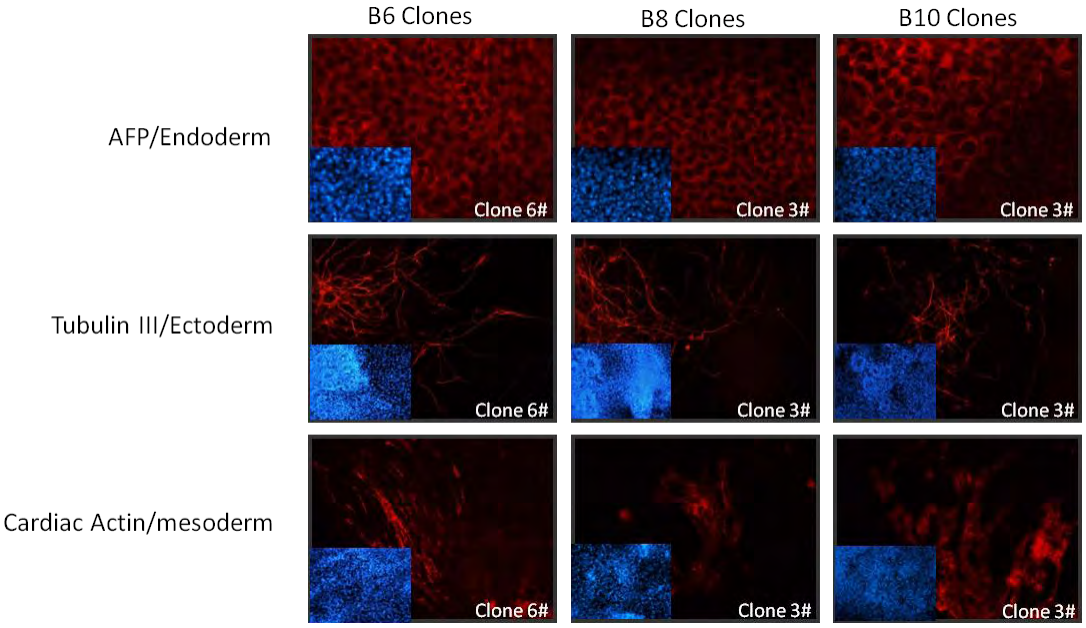


Figure 4.S7. Inhibitor-treated iPSCs contribute to the germline of chimeric embryos

(A) B10-treated iPSCs were injected into the cavity of recipient blastocysts at day 0. Image showed GFP+ iPSCs right after injection into the recipient embryo. (B) iPSCs integrated into inner cell mass of recipient blastocysts by day 1. Arrows refer to GFP+ iPSCs. (C) iPSCs contribute to the germline of chimeric embryos. Injected embryos were harvested at E13.5 and genital ridge tissues were dissected and analyzed for GFP-positivity.

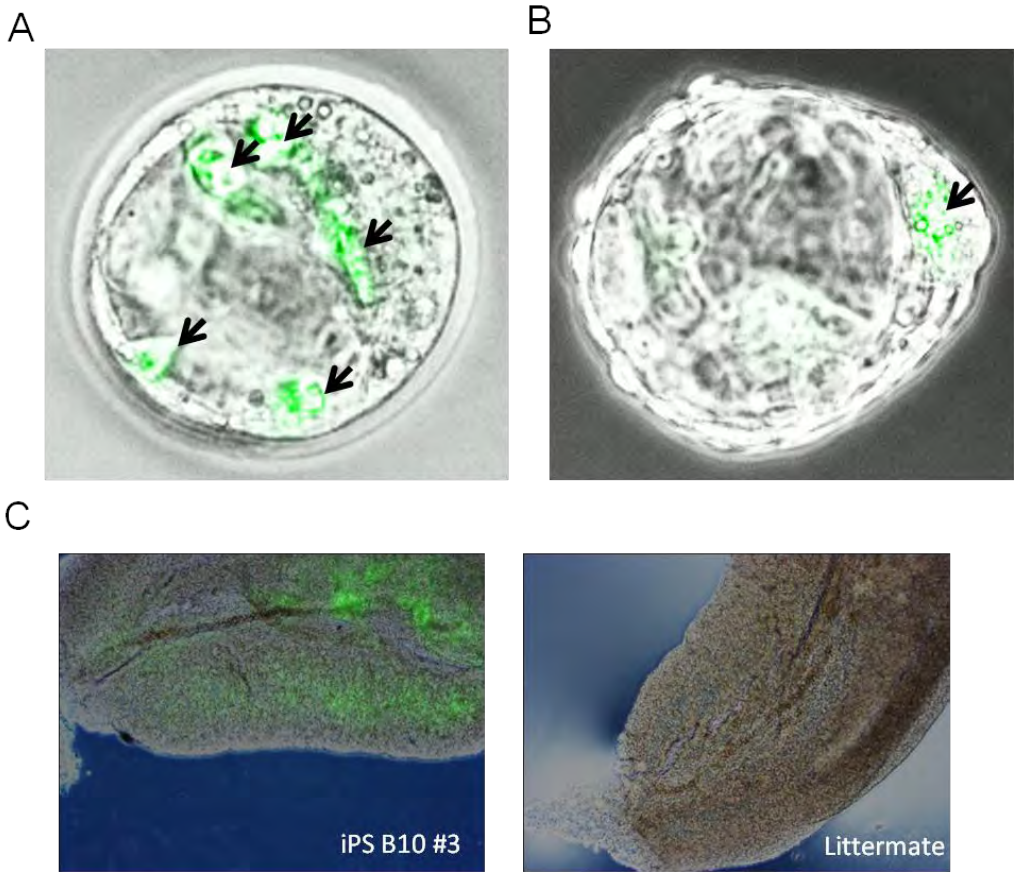


Figure 4.S8. MLN8237 increases levels of AurkA protein

4F-infected MEFs were treated with MLN8237 at various concentrations at day 3 post 4F infection. Cells were harvested 48 hrs after drug treatment for western blotting analysis of AurkA protein. Actin serves as loading control.

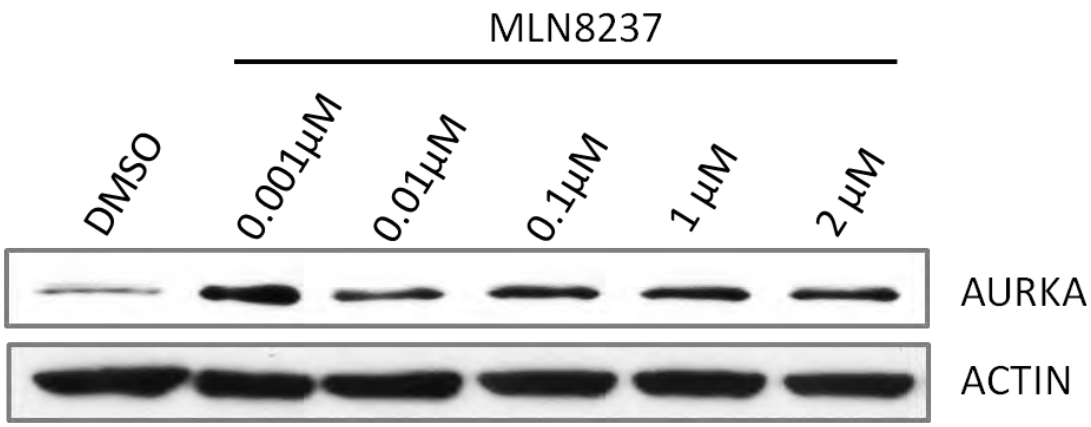


Figure 4.S9. Overexpression of wt and kinase-dead AurkA in MEFs

MEFs or 4F-infected MEFs were infected with either wild-type (wt) mouse AurkA virus or virus expressing a kinase-dead mutant human AurkA D274A virus at day 0. Cells were harvested at day 5 post-infection for analysis of AurkA proteins. mAURKA refers to wild type mouse AurkA protein and hAURKA refers to kinase-dead mutant of human AurkA protein. Specific antibodies to each protein were used to detect overexpressed proteins.

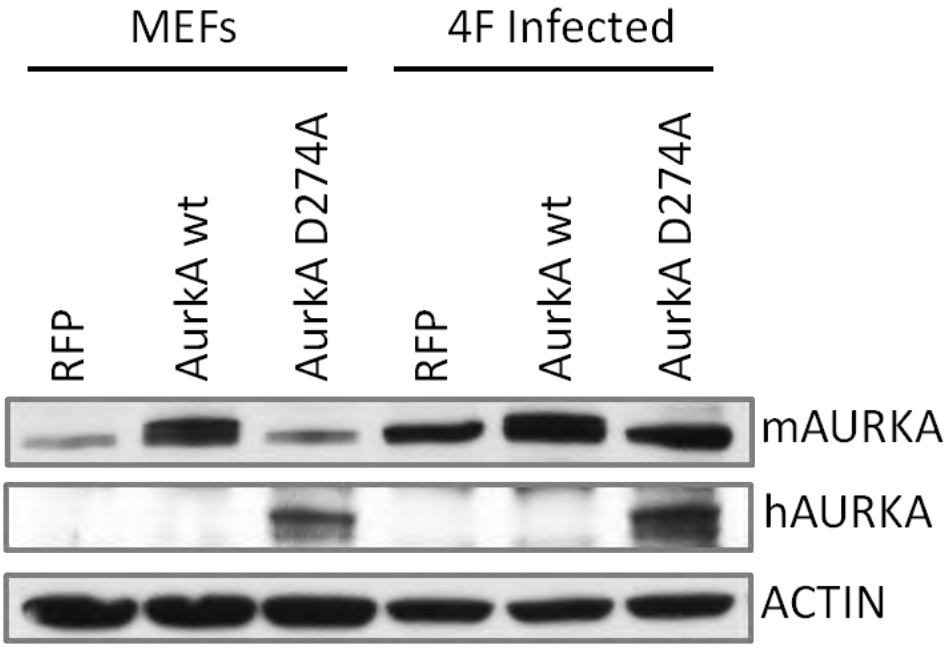


Figure 4.S10. Conservation of mouse and human AurkA

Amino acid alignment of mouse and human proteins shows ~84% identity in amino acid sequence.

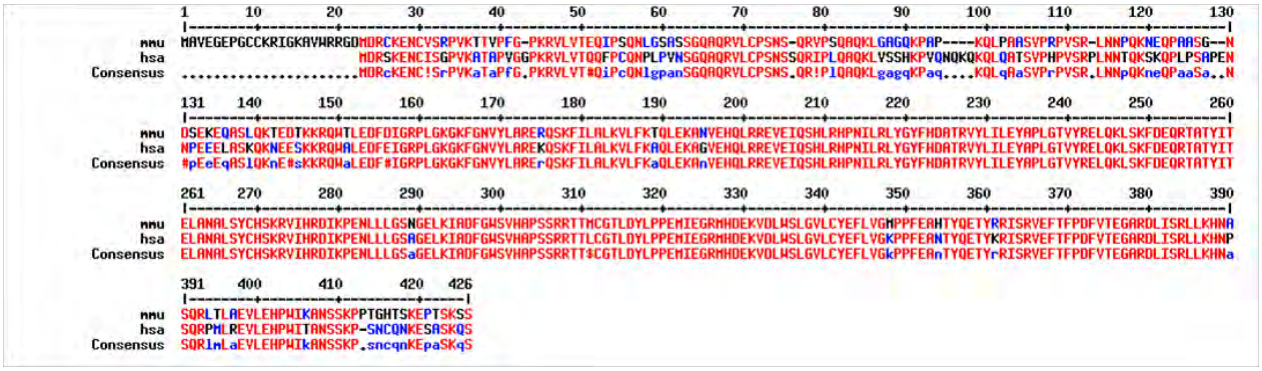


Figure 4.S11. AurkA knockdown promotes GSK3β inactivation

MEFs were transfected with AurkA and control siRNAs at a final concentration of 50nM for two days. Cells were then harvested for western blotting analysis of AurkA, total and phosphorylated GSK3β (Ser9), and actin, as a loading control.

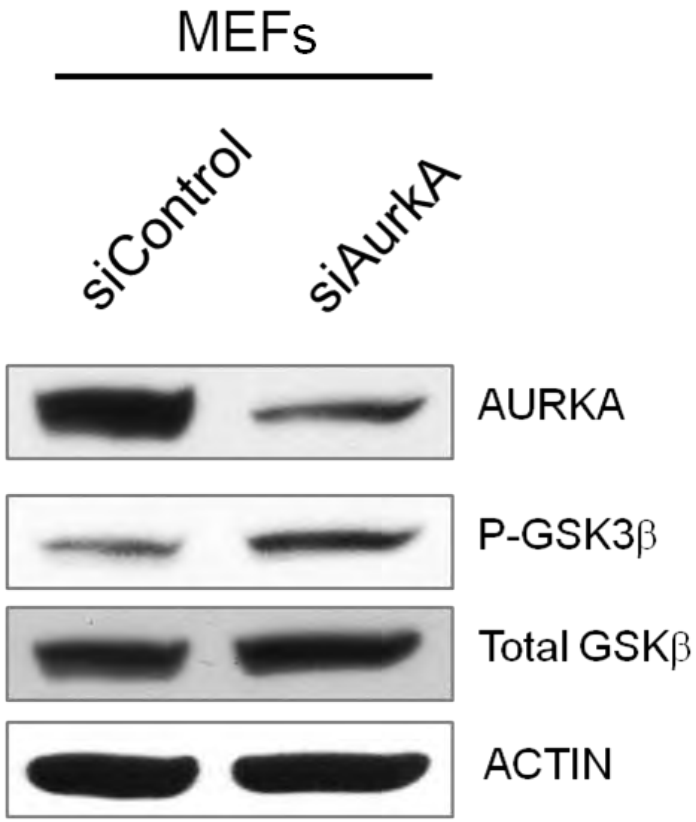


Figure 4.S12. Overexpression of Gsk3β largely blocks B6's effect on reprogramming

(A) Gsk3β was overexpressed in MEFs by pMXs retroviral vector. MEFs were infected with pMX-Gsk3β virus for four days before harvesting for RT-qPCR analysis. pMX-RFP was used as control. Error bar represents experiment with duplicate wells. (B) Overxpression of Gsk3β blocked B6's effect on reprogramming. Mouse Gsk3β was cloned into pMX retroviral vector and transduced into MEFs together with OSKM. Compound treatment (1μM) was started at day3 post transduction. GFP+ colonies were quantified at day12 post transduction. Error bar represents experiment with three independent wells. * $p<0.05$ (Student t-test).

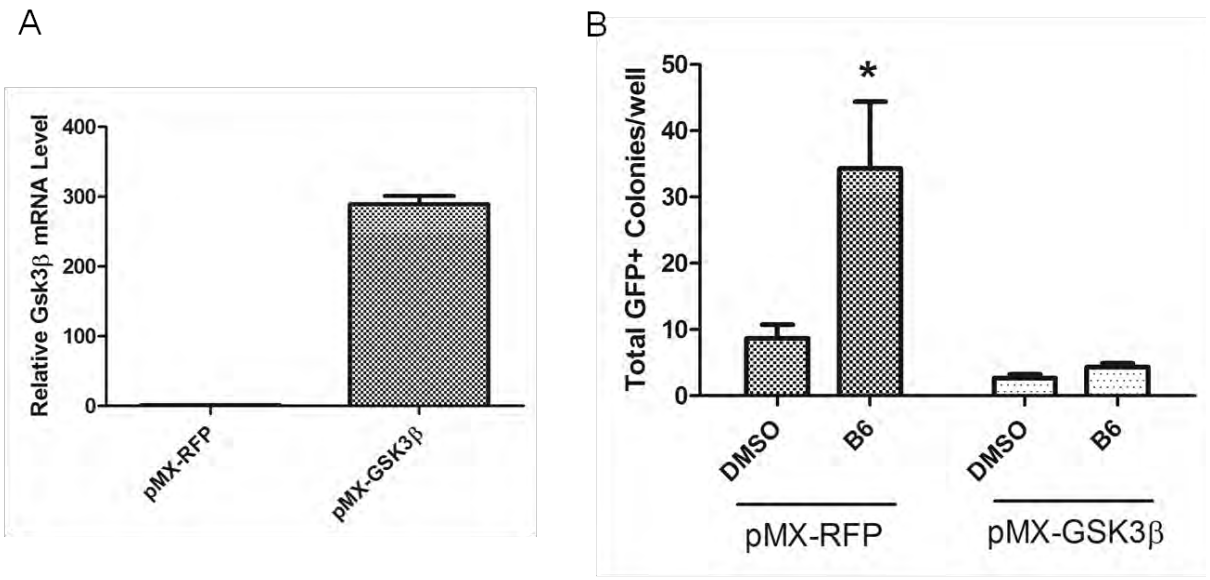


Figure 4.S13. B6 treatment does not alter expression of genes upstream of Akt

MEFs were first transduced with 4F for three days and then treated for another two days with DMSO or 1μM B6. Total RNAs were harvested and analyzed for indicated factors, which reportedly function upstream of Akt.

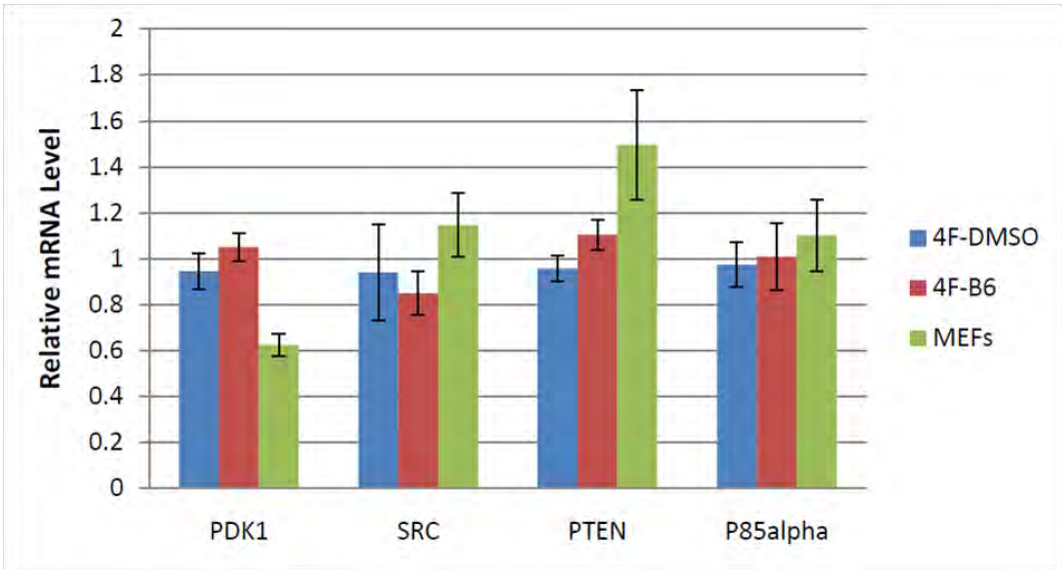


Figure 4.S14. MLN8237 dose-dependently enhances reprogramming

(A) The number of Oct4-GFP+ colonies was significantly increased upon MLN8237 treatment. Data is derived from analysis of three independent wells. (B) MLN8237 treatment induces mES-specific gene expression in 4F-transduced MEFs. Cells were harvested at day14 post 4F transduction. Nanog, Tet1 and Eras expression was analyzed in cells treated with DMSO or MLN8237. Shown are data derived from three independent wells.

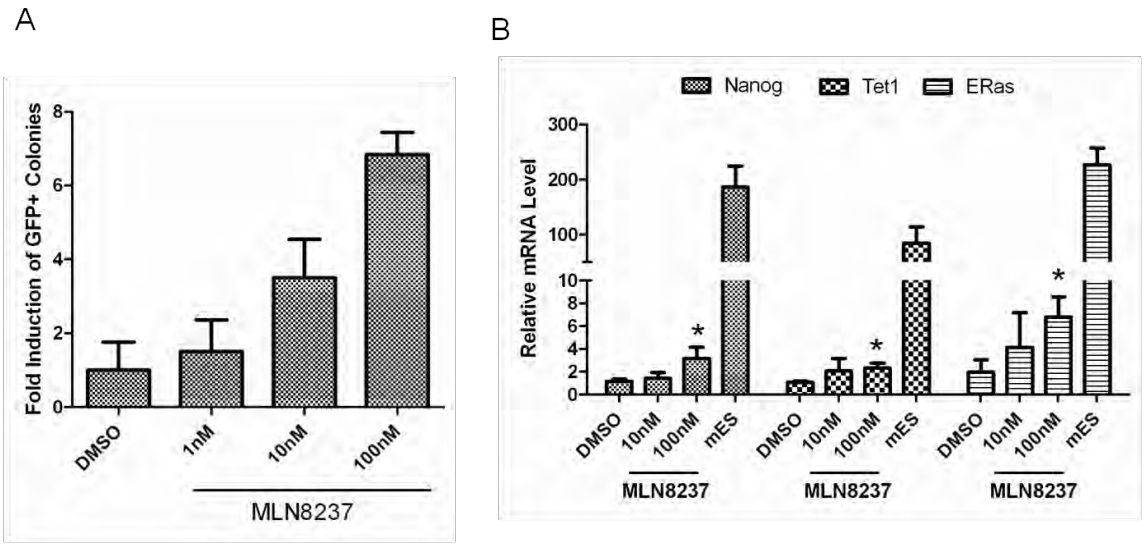


Figure 4.S15. Low concentrations of AurkA inhibitors do not alter the cell cycle

(A) 4F-infected MEFs were treated with different concentrations of B6 at day 3 post-infection. Cells were harvested after 48 hr of drug treatment, subjected to PI staining and analyzed for DNA content by flow cytometry. (B) MLN8237 treatment at 10 nM did not alter the cell cycle of 4F-infected MEFs. The experimental procedure was as in (A).

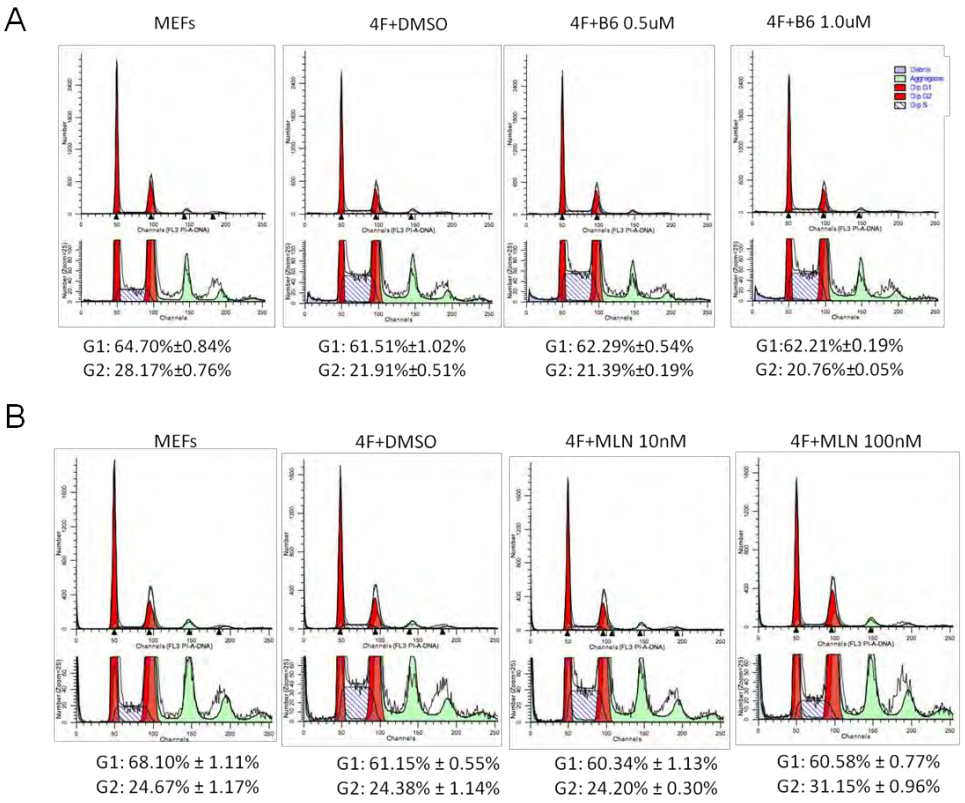


Figure 4.S16. AurkA inhibitor treatment does not inhibit mES cell differentiation

(A) Treatment with inhibitor B6 does not block changes in mES cell morphology following LIF withdrawal. mES cells were cultured both in LIF+ or LIF- medium for 4 days and harvested for AP staining. Differentiating cells become AP-negative and show a more scattered morphology. (B) Inhibitor B6 does not alter silencing of self-renewal genes following LIF withdrawal. The experimental procedure was as in (A). Cells were harvested for RNA extraction and RT-qPCR of indicated mES self-renewal genes. N.S.: Not significant.

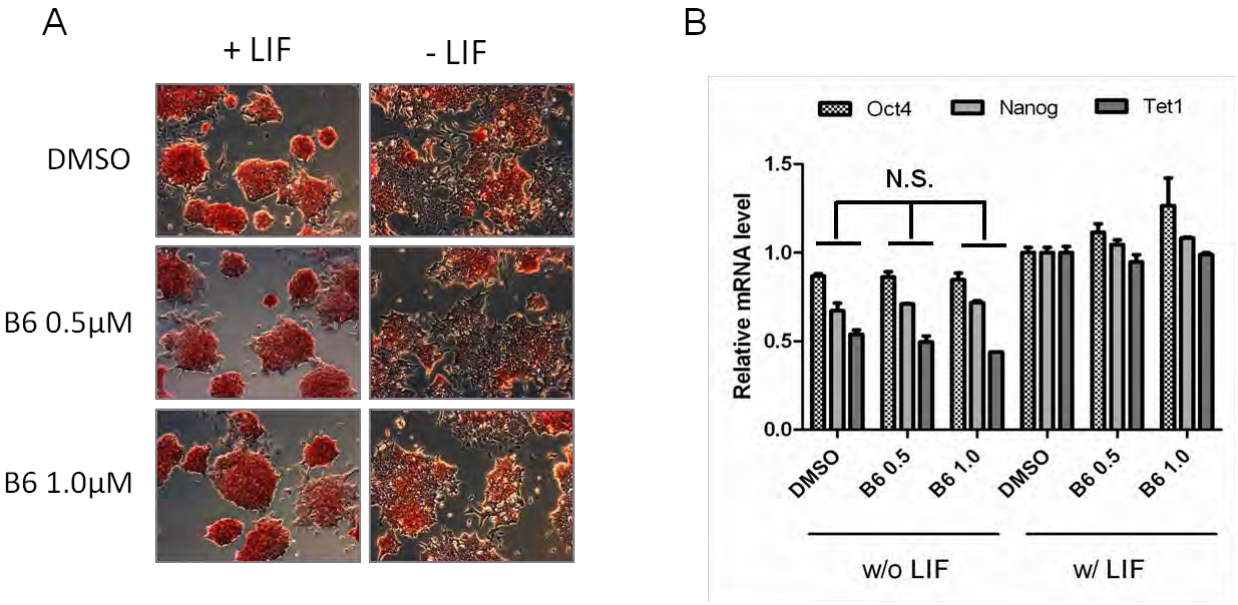


Table 4.1. List of barrier hits and potential targets

Compound ID	Compound Name	Potential Targets
BIM-0086769	KN-62	CaM kinase II (900nM)
BIM-0086727	Alsterpaullone	GSK-3β(4nM), Cdk1/cyclin B (35nM)
BIM-0207133	Arcyriaflavin A, Synthetic	cdk4/cyclin D1 (59nM), CaM kinase II (25nM),PKA (>2uM), PKC (>100uM)
BIM-0207164	IP3K Inhibitor	IP 3-K (10.2uM)
BIM-0050621	ML-7, Hydrochloride	myosin light chain kinase(300nM), PKA(21uM), PKC(42uM)
BIM-0086701	PP3	EGFR kinase (2.7uM)
BIM-0086714	Syk Inhibitor III	Syk kinase(2.5uM), Src(29.3uM)
BIM-0086660	Aurora Kinase Inhibitor III	Aurora A (42nM), Lck(131nM), Bmx(386nM), IGF1-R(591uM), SyK(887nM)
BIM-0086787	Sphingosine Kinase Inhibitor	sphingosine kinase(0.5uM)
BIM-0086716	TGF-b RI Inhibitor III	activin receptor-like kinase 4(129nM), 5 (47nM), p38 MAPKα(10.6uM)
BIM-0207174	p38 MAP Kinase Inhibitor IV	p38α MAPK (130nM), p38β MAPK (550nM)

Table 4.2. List of essential hits and potential targets

Compound ID	Compound Name	Potential Targets
BIM-0086655	AGL 2043	PDGFR(800nM), Flt3 and Kit (1-3uM)
BIM-0207149	Cdk2/9 Inhibitor	Cdk2/E(2nM), Cdk9/T1(4nM), GSK-3β, Cdk4/D1, Cdk7/H, Cdk1/B, and Abl (20, 53, 70, 80, and 160 nM)
BIM-0086744	Cdk/Crk Inhibitor	Cdks (IC50 = 48 nM, 15 nM, 9 nM, 10 nM, 71 nM, and 9 nM for Cdk1/B, Cdk2/E, Cdk3/E, Cdk5/p35, Cdk7/H/MAT1, and Cdk9, respectively), Cdk4/D1, Cdk6/D3, and GSK-3β(839 nM, 282 nM, and 754 nM)
BIM-0086751	Fascaplysin, synthetic	Cdk4/D1(0.35uM), Cdk6/D1(3.4uM)
BIM-0086708	Rho Kinase Inhibitor IV	11.8 nM, > 10 μM, > 10 μM, 3.26 μM, 2.35 μM, and 2.57 μM for ROCKII, PKA, PKC, PKG, Aurora A, and CaMKII, respectively
BIM-0086768	K-252a, Nocardiopsis sp.	CaM kinase II(1.8nM), myosin light chain kinase(20nM), PKA(18nM), PKC(25nM), PKG(20nM), gp140trk(3nM)
BIM-0207189	UCN-01	9 nM, 34 nM, 30 nM, 590 nM and 530 nM for PKCα, PKCβ, PKCγ, PKCδ and PKCε, 7 nM, 27 nM, 50 nM, 50 nM, 150 nM and 1.04 μM for Chk1, Cdc25C-associated protein kinase 1, Cdk1, PAK4, Cdk5/p25 and Chk2, 33 nM, 50 nM, 95 nM, 500 nM, 500 nM and 1.0 μM for PDK1, lck, MAPKAP kinase-2, Akt, GSK-3β and PKA
BIM-0207186	PI 3-Kα Inhibitor VIII	0.3, 40, 100, and 850 nM for p110α, p110γ, PI 3-K C2β, and p110β
BIM-0207192	Reversine	MEK1(8nM), Myosin II heavy chain(10nM), Aurora Kinases A, B and C (IC50 = 400, 500 and 400 nM)

CHAPTER V: FINAL SUMMARY AND PERSPECTIVES

In this dissertation, we have utilized both microRNAs and small molecules to understand a part of the molecular mechanism on how somatic cells could be reprogrammed to become pluripotent stem cells. Our findings revealed that microRNAs could play important functional roles in somatic cell reprogramming such as promoting mesenchymal-to-epithelial transition, modulating p53-p21 and TGFβ pathways and kinase functions. The finding that microRNAs could target the expression of signature ECM genes has also indicated that besides the common intracellular signaling regulation, extracellular networks are also regulated by microRNAs, which could result in profound effect on the fate and proper function of the target cells. Meanwhile, kinase inhibitor screen has discovered several new lines in understanding somatic cell reprogramming. The identification of p38, IP3K and Aurora A kinases as the new barrier genes as well as the discovery that essential kinases are enriched with cell cycle regulators has provided the proof of principle that studying kinases in genome wide scale could yield in depth mechanistic and functional insights of both reprogramming process and the target kinases themselves. In addition, we have identified that inhibition of Aurora A kinase, a gene that is involved in spindle formation and can act both as an oncogene and tumor suppressor could enhance the inactivation /phosphorylation of GSK3β in an Akt-dependent manner, a function that has not been reported before.

Meanwhile, our discoveries have also revealed new insights on the barrier pathways that inhibit reprogramming. Since the discovery that somatic cells could be reprogrammed to an ES-like state, it has long been proposed that certain barrier pathways may be the main inhibitor that prevents cells from becoming reprogrammed. With the mounting evidence of years‘ efforts, it has become much clearer that intracellular gene networks indeed have a primary role in inhibiting somatic cell reprogramming. In general, terminal differentiated cells are restricted or locked in their specific cell types due to the existence of two main regulatory networks: one is the cellular network that define or maintain their current identities. This network usually provides all the necessary regulations to maintain the cells in a healthy growing state and functioning properly. This network could be highly tissue specific due to the fact that different cell types from various tissues should have very different signature gene expressions and regulations to maintain their normal function. Once this network is compromised, the cells will lose their identities and could not function properly, which would trigger a series of events that may finally lead them to the senescence and apoptosis. The other network is the cell defense/safety mechanisms that prevent the cells from transformation or tumor initiation and usually consists of tumor suppressing and cell death related signaling pathways. These signaling pathways shall be the common mechanisms that exist in most of the cell types, as they all need to have such system to prevent the cells from becoming tumor cells. Once the cells become transformed, their normal function and growth

regulation will be lost, affecting the function of the whole tissue/organ. Activation of these defense/safety pathways could help the system to eliminate the cells that have accumulate enough genetic mutations and DNA damages to initiate tumors, without the risk of compromising the whole tissue/organ. Finally, the two regulatory networks will not act independently. They do not have clearer boundaries and function together to regulate the normal life of a particular cell.

Therefore, during the reprogramming process, once the four factors are introduced and overexpressed in the target cells, they would need to overcome both of the two regulatory networks in order to reprogramme the cells successfully. In our experiments, we have identified that one of the mechanisms that four factors utilized is through microRNA mediated regulatory machinery. We have found that miR-93 family microRNAs could target to p53-p21 and TGFβ signaling pathways, which are the common cell defense/safety mechanisms, and miR-135b could target to Wisp1-mediated extracellular matrix signaling pathway, which is a tissue-specific network required for maintaining MEFs‘ identity. These findings have provided the first evidence that both of these two regulatory networks could be regulated by microRNAs which are induced by 4F during the reprogramming process. Loss of microRNA biogenesis machinery thus could have huge impact on the repgoramming efficiency as the level of microRNAs available in the cells will be very limited and could not silence the targeted pathways to promote reprogramming.

Moreover, in our studies, the discovery that Wisp1-mediated ECM formation could affect the efficiency of somatic cell reprogramming has raised an interesting question: how is the reprogramming process regulated by extracellular signaling networks? Previous findings have all focused on the intracellular gene networks that could modulate reprogramming process. The effect of extracellular matrix and establishment of proper niche for iPSCs undergoing reprogramming has long been overlooked. In fact, certain evidences have clearly implied an important role of extracellular matrix networks in regulating reprogramming efficiency. For example, if cultured alone and long enough, most of the somatic cells (>80%) could become fully reprogrammed (Hanna et al, 2009). However, in reality, when a bulk of cells is being reprogrammed, only 0.01-0.2% cells could reach the fully reprogrammed state (Aoi et al, 2008; Meissner et al, 2007; Nakagawa et al, 2008; Takahashi & Yamanaka, 2006). Even with prolonged culture time, the total number of iPSCs colonies would maintain the same (Sridharan et al, 2009). These data clearly suggest that intracellular gene regulation is not the only force to affect the overall reprogramming efficiency and some other factors could contribute to the process as well. One such example is ECM mediated niche formation.

Mouse ES cells and iPSCs have distinct morphology from embryonic fibroblasts (MEFs). While MEFs could usually grow separately in a dish and are in general independent on cell-to-cell contacts, mESCs and iPSCs usually grow in colonies. These cells have tight junctions with nearby cells and form their own

niches when cultured in the feeder plates. It has been previously reported that establishing their own niches is required for ES cells to have robust and healthy growth and maintain self renewal state (Bendall et al, 2007). Therefore, ideally, in order to grow in such a state, reprogramming cells would also need to establish such niches to support their growth. Interrupting the process therefore will result in a decreased reprogramming efficiency. In addition to that, reprogramming cells also face some complicated situations. For example, since these cells are not growing in a stable state, the gene regulatory networks for mES/iPSCs self renewal are not established yet by four factors. At the same time, they still express a significant portion of MEFs specific genes and retain some of MEFs‘ original signaling networks. When these cells are in culture, they are surrounded by many feeder cells which constantly secret fibroblasts factors, such as WISP1, TGFBI, NOV and DKK2, as well as necessary nutrients that support iPSC/mES growth. For mES cells or iPSCs, since they do not have any MEF specific signaling network, they will not be affected by those fibroblasts factors. However, for reprogramming cells, they could be stimulated by them, which could make those target cells confused due to the existence of two networks, and thus resistant to reach fully reprogrammed state. Therefore, it is possible to modulate these processes and pathways in order to enhance the reprogramming efficiency. In this dissertation, we found that miR-135b mediated regulation of Wisp1-related ECM formation indeed could strongly affect the reprogramming efficiency and by

manipulating the expression of ECM components, we were able to push the balance toward full reprogrammed state.

Another aspect that modulates somatic reprogramming is cell cycle. In the kinase inhibitor screen, we have found that essential hits identified were enriched with cell cycle inhibitors. Recent reports from other colleagues also have indicated that cell cycle control could be one of the key events that affect the overall reprogramming efficiency (Ruiz et al, 2011). Why is cell cycle so important? Here comes another striking difference between mESC/iPSCs and MEFs. For fibroblast cells, their overall proliferation rate is quite low compared with mESCs/iPSCs. In addition, since the telomerases are active in mESCs/iPSCs, these cells could divide unlimitedly, while MEFs do not have such advantage due to lack of telomerase activity. Therefore, it is not surprising that MEFs will have a quite different cell cycle regulation from mESCs/iPSCs and when they are being reprogrammed, it is required that a suitable cell cycle control is established by overexpression of reprogramming factors. Furthermore, as mentioned previously, cell defense/safety mechanisms are one of the main barriers to inhibit reprogramming process. In fact, many of such mechanisms, for example, p53-p21 pathways and DNA damage responses, are closely associated with cell cycle control. Overexpression of some of the reprogramming factors could result in activation of such mechanisms that leads to DNA damages accumulation and cell cycle arrest. Therefore, it may be beneficial to eliminate

the cell cycle inhibitors and increase the overall proliferation for the reprogramming cells.

However, increasing the speed of proliferation could also have negative impact on the reprogramming efficiency. One such concern is the amount of reprogramming factors accumulated in the nucleus. Given the general speed of expression of reprogramming factors remains the same, increased the cell division will reduce the amount of proteins accumulated. It has been long known that reprogramming factors need to reach certain amounts or overall dose in order to initiate efficient reprogramming process (Sridharan et al, 2009). Reduced amount of transcription factors in the nucleus could lead to insufficient suppression of MEF specific gene networks and insufficient activation of mESC/iPSC signaling pathways and thus result in the overall reprogramming efficiency. In fact, according to our observations, 4F transduced population with large percentage of fast growing transformed cells usually generate much less full reprogrammed iPSC colonies, to which partial reason may be due to the low amount of accumulated reprogramming factors. Therefore, modest increase of cell proliferation may be more beneficial for the somatic cells to become efficiently reprogrammed.

In addition, when MEFs are reprogrammed to iPSCs, they need to change from slow proliferating cells to acquire a state with fast growing and short cell cycle. All these changes would require sufficient supply of nucleotides and other

nutrients so that DNA synthesis and replication could be coordinated. Therefore, it is expected that reprogramming cells would also undergo significant rewiring of their metabolic pathways. Recent report has also indicated that such process is required for efficient reprogramming (Folmes et al, 2011; Panopoulos et al, 2012). Interestingly, some of mostly highly regulated genes identified from miR-135b experiments are key components of important metabolic pathways. For example, 4EBP1 is one of the main downstream effector of mTOR pathway, an important metabolic regulatory pathway that is involved in many human diseases (Zoncu et al, 2011). 4EBP1 function through direct interaction with translation initiation factor 4E and inhibit the recruitment and formation of initiation complex in order to repress translation. It has been shown to have a central role in mTOR dependent oncogenic signaling pathway during tumorigenesis (Hsieh et al, 2010). In our experiments, we have found that 4EBP1 mRNA was strongly repressed upon miR-135b transfection, which could partially explain why miR-135b exhibits more specific enhancing effects on the reprogramming process than other miRs such as miR-93. Since miR-135b decreases the expression of 4EBP1, it may reduce the downstream effects of 4EBP1 and have inhibitory effects on other transformed tumor-like cells rather than real iPSCs due to the loss of oncogenic mTOR signaling. Therefore, we could have a specific increase of overall percentage of fully reprogrammed cells.

Finally, in summary, the discovery that somatic cells could be reprogrammed back to an embryonic stem cell-like pluripotent state has

revolutionized the current concept and understanding of cell fate determination. Terminally differentiated cells are used to be recognized as the final state of cell differentiation, which was thought not to be reversed. It is realized now that somatic cells indeed can be reprogrammed back to pluripotent stem cells and this process will only need a few transcriptional factors. This has led to a number of important questions: first, is there a signature set of transcriptional factors (TFs) for every kind of cells from various tissues? If so, by ectopic expression of these signature TFs, can we cross-differentiate cells of distinct origin? Indeed, recent progress has clearly proved that many kinds of cells could be trans-differentiated given the right set of transcriptional factors and culture conditions were provided, including cardiomyocytes, neurons and pancreatic cells (Ambasudhan et al, 2011; Ieda et al, 2010; Zhou et al, 2008). Therefore, it may be true that the cells from various tissues maintain their identities through the expression of tissue specific transcription factors, which are the main force to drive downstream tissue specific gene expressions and maintain the normal function of these cells. In order to change the fate or identity of terminal differentiated cells, one would just need to express a few new transcription factors as well as silencing the original gene networks in those target cells.

However, despite its great impact on understanding the fundamental biology of cell fate and pluripotency, somatic cell reprogramming still face a few challenges. In order to develop the therapeutic use of iPSCs, two main issues still need to be solved. One is the use of oncogenes as reprogramming factors

such as Klf4 and Myc. Overexpression of these genes could result in high tumorigenesis potential in derived iPSCs. In addition, overexpression of these could lead to accumulation of DNA damages which may significantly increase the risk of introducing additional mutations that could not be identified easily. These challenges will require further work in order to develop a more sophisticated method for deriving safe iPSC lines. Besides that, another challenge is the differentiation of iPSCs into pure downstream lineages. One of the main purposes of deriving iPSCs for personal medicine is to provide unlimited source for deriving important differentiated populations. However, there are only a few cell types for which a successfully method has been established to efficiently differentiate ES cells or iPSCs to the target population. Some of the key functional cells that iPSCs hold the most interest in, such as HSCs, are still difficult to get in an *in vitro* system. Therefore, it may be too early to see the therapeutic use of iPSCs. However, it could become a powerful tool to address some key issues of human diseases, such as the pathogenesis of ALS or other neural diseases. By using patient specific iPSCs, it is possible to model certain human diseases and understand their molecular mechanism, which could be extremely valuable for developing life-saving treatments and therapeutics. Alternatively, trans-differentiation could serve as a potential way for generating large number of target cell population both in vitro and in vivo, within a relatively shorter period of time. Since the signature transcriptional factors usually do not contain oncogenes, these cells could be safer to iPSCs and thus might be

suitable for developing novel treatments. Overall, it is remarkable to witness that the discovery of somatic cell reprogramming makes such an impact on both the basic knowledge of cell biology and development of new therapeutics. In years to come, more insightful studies are expected on the molecular mechanism of reprogramming process, which will benefits us all and let us understand how different cells are programmed by nature and how their functions are properly maintained.

BIBLIOGRAPHY

Abbas T, Dutta A (2009) p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer* **9:** 400-414

Adams JC, Watt FM (1993) Regulation of development and differentiation by the extracellular matrix. *Development* **117:** 1183-1198

Ahmed AA, Mills AD, Ibrahim AE, Temple J, Blenkiron C, Vias M, Massie CE, Iyer NG, McGeoch A, Crawford R, Nicke B, Downward J, Swanton C, Bell SD, Earl HM, Laskey RA, Caldas C, Brenton JD (2007) The extracellular matrix protein TGFBI induces microtubule stabilization and sensitizes ovarian cancers to paclitaxel. *Cancer Cell* **12:** 514-527

Amarzguioui M, Holen T, Babaie E, Prydz H (2003) Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Res* **31:** 589-595

Ambasudhan R, Talantova M, Coleman R, Yuan X, Zhu S, Lipton SA, Ding S (2011) Direct Reprogramming of Adult Human Fibroblasts to Functional Neurons under Defined Conditions. *Cell Stem Cell* **9:** 113-118

Ambros V (2004) The functions of animal microRNAs. *Nature* **431:** 350-355

Ambros V (2011) MicroRNAs and developmental timing. *Curr Opin Genet Dev* **21:** 511-517

Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, Morrisey EE (2011) Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. *Cell Stem Cell* **8:** 376-388

Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T, Yamanaka S (2008) Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321:** 699-702

Ashford J, Schoffstall C, Reddick WE, Leone C, Laningham FH, Glass JO, Pei D, Cheng C, Pui CH, Conklin HM (2010) Attention and working memory abilities in children treated for acute lymphoblastic leukemia. *Cancer* **116:** 4638-4645

Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, Eachus R, Pasquinelli AE (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122:** 553-563

Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V, Walsh M, Vallier L, Gil J (2009) Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev* **23:** 2134-2139

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281-297

Beattie J, Allan GJ, Lochrie JD, Flint DJ (2006) Insulin-like growth factor-binding protein-5 (IGFBP-5): a critical member of the IGF axis. *Biochem J* **395:** 1-19

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20:** 1885-1898

Bendall SC, Stewart MH, Menendez P, George D, Vijayaragavan K, Werbowetski-Ogilvie T, Ramos-Mejia V, Rouleau A, Yang J, Bosse M, Lajoie G, Bhatia M (2007) IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature* **448:** 1015-1021

Berschneider B, Konigshoff M (2011) WNT1 inducible signaling pathway protein 1 (WISP1): a novel mediator linking development and disease. *Int J Biochem Cell Biol* **43:** 306-309

Bissell MJ, Hines WC (2011) Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat Med* **17:** 320-329

Blelloch R, Wang Z, Meissner A, Pollard S, Smith A, Jaenisch R (2006) Reprogramming efficiency following somatic cell nuclear transfer is influenced by

the differentiation and methylation state of the donor nucleus. *Stem Cells* **24:** 2007-2013

Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R (2008) Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2:** 151-159

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128:** 1089-1103

Brown KM, Chu CY, Rana TM (2005) Target accessibility dictates the potency of human RISC. *Nat Struct Mol Biol* **12:** 469-470

Burnett JC, Rossi JJ, Tiemann K (2011) Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnol J* **6:** 1130-1146

Carey BW, Markoulaki S, Hanna JH, Faddah DA, Buganim Y, Kim J, Ganz K, Steine EJ, Cassady JP, Creyghton MP, Welstead GG, Gao Q, Jaenisch R (2011) Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell* **9:** 588-598

Chen CC, Lau LF (2009) Functions and mechanisms of action of CCN matricellular proteins. *Int J Biochem Cell Biol* **41:** 771-783

Chiu YL, Rana TM (2002) RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol Cell* **10:** 549-561

Chiu YL, Rana TM (2003) siRNA function in RNAi: a chemical modification analysis. *RNA* **9:** 1034-1048

Choi YJ, Lin CP, Ho JJ, He X, Okada N, Bu P, Zhong Y, Kim SY, Bennett MJ, Chen C, Ozturk A, Hicks GG, Hannon GJ, He L (2011) miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nat Cell Biol* **13:** 1353-1360

Chu CY, Rana TM (2006) Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol* **4:** e210

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature* **453:** 798-802

Desnoyers L, Arnott D, Pennica D (2001) WISP-1 binds to decorin and biglycan. *J Biol Chem* **276:** 47599-47607

Eminli S, Utikal J, Arnold K, Jaenisch R, Hochedlinger K (2008) Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cells* **26:** 2467-2474

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS (2003) MicroRNA targets in Drosophila. *Genome Biol* **5:** R1

Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, Li W, Weng Z, Chen J, Ni S, Chen K, Li Y, Liu X, Xu J, Zhang S, Li F, He W, Labuda K, Song Y, Peterbauer A, Wolbank S, Redl H, Zhong M, Cai D, Zeng L, Pei D (2010) Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6:** 71-79

Eulalio A, Huntzinger E, Izaurralde E (2008) Getting to the root of miRNA-mediated gene silencing. *Cell* **132:** 9-14

Fabian MR, Mathonnet G, Sundermeier T, Mathys H, Zipprich JT, Svitkin YV, Rivas F, Jinek M, Wohlschlegel J, Doudna JA, Chen CY, Shyu AB, Yates JR, 3rd, Hannon GJ, Filipowicz W, Duchaine TF, Sonenberg N (2009) Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* **35:** 868-880

Fabian MR, Sonenberg N, Filipowicz W (2010) Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79:** 351-379

Feng B, Ng JH, Heng JC, Ng HH (2009) Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* **4:** 301-312

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806-811

Folmes CD, Nelson TJ, Martinez-Fernandez A, Arrell DK, Lindor JZ, Dzeja PP, Ikeda Y, Perez-Terzic C, Terzic A (2011) Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab* **14:** 264-271

Garamszegi N, Garamszegi SP, Samavarchi-Tehrani P, Walford E, Schneiderbauer MM, Wrana JL, Scully SP (2010) Extracellular matrix-induced transforming growth factor-beta receptor signaling dynamics. *Oncogene* **29:** 2368-2380

Gartel AL, Ye X, Goufman E, Shianov P, Hay N, Najmabadi F, Tyner AL (2001) Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. *Proc Natl Acad Sci U S A* **98:** 4510-4515

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320:** 1077-1081

Giorgetti A, Montserrat N, Aasen T, Gonzalez F, Rodriguez-Piza I, Vassena R, Raya A, Boue S, Barrero MJ, Corbella BA, Torrabadella M, Veiga A, Izpisua Belmonte JC (2009) Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. *Cell Stem Cell* **5:** 353-357

Gorgun G, Calabrese E, Hideshima T, Ecsedy J, Perrone G, Mani M, Ikeda H, Bianchi G, Hu Y, Cirstea D, Santo L, Tai YT, Nahar S, Zheng M, Bandi M, Carrasco RD, Raje N, Munshi N, Richardson P, Anderson KC (2010) A novel Aurora-A kinase inhibitor MLN8237 induces cytotoxicity and cell-cycle arrest in multiple myeloma. *Blood* **115:** 5202-5213

Greber B, Lehrach H, Adjaye J (2007) Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal. *Stem Cells* **25:** 455-464

Gu S, Rossi JJ (2005) Uncoupling of RNAi from active translation in mammalian cells. *RNA* **11:** 38-44

Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:** 835-840

Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, Creyghton MP, Steine EJ, Cassady JP, Foreman R, Lengner CJ, Dausman JA, Jaenisch R (2008) Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133:** 250-264

Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R (2009) Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462:** 595-601

Hanna J, Wernig M, Markoulaki S, Sun CW, Meissner A, Cassady JP, Beard C, Brambrink T, Wu LC, Townes TM, Jaenisch R (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318:** 1920-1923

Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S (2009) Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460:** 1132-1135

Houbaviy HB, Murray MF, Sharp PA (2003) Embryonic stem cell-specific MicroRNAs. *Dev Cell* **5:** 351-358

Hsieh AC, Costa M, Zollo O, Davis C, Feldman ME, Testa JR, Meyuhas O, Shokat KM, Ruggero D (2010) Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. *Cancer Cell* **17:** 249-261

Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, Melton DA (2008) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* **26:** 795-797

Hutvagner G, Simard MJ, Mello CC, Zamore PD (2004) Sequence-specific inhibition of small RNA function. *PLoS Biol* **2:** E98

Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, Di Giorgio FP, Koszka K, Huangfu D, Akutsu H, Liu DR, Rubin LL, Eggan K (2009) A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell* **5:** 491-503

Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142:** 375-386

Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H, Boulos M, Gepstein L (2011) Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* **471:** 225-229

Ivanovska I, Ball AS, Diaz RL, Magnus JF, Kibukawa M, Schelter JM, Kobayashi SV, Lim L, Burchard J, Jackson AL, Linsley PS, Cleary MA (2008) MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol Cell Biol* **28:** 2167-2174

Judson RL, Babiarz JE, Venere M, Blelloch R (2009) Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol* **27:** 459-461

Jun JI, Lau LF (2011) Taking aim at the extracellular matrix: CCN proteins as emerging therapeutic targets. *Nat Rev Drug Discov* **10:** 945-963

Kahlem P, Newfeld SJ (2009) Informatics approaches to understanding TGFbeta pathway regulation. *Development* **136:** 3729-3740

Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Belmonte JC (2009) Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460:** 1140-1144

Kawano Y, Kypta R (2003) Secreted antagonists of the Wnt signalling pathway. *J Cell Sci* **116:** 2627-2634

Kessenbrock K, Plaks V, Werb Z (2010) Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141:** 52-67

Kim KS, Seu YB, Baek SH, Kim MJ, Kim KJ, Kim JH, Kim JR (2007) Induction of cellular senescence by insulin-like growth factor binding protein-5 through a p53-dependent mechanism. *Mol Biol Cell* **18:** 4543-4552

Kim VN, Han J, Siomi MC (2009a) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10:** 126-139

Kim YK, Yu J, Han TS, Park SY, Namkoong B, Kim DH, Hur K, Yoo MW, Lee HJ, Yang HK, Kim VN (2009b) Functional links between clustered microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. *Nucleic Acids Res* **37:** 1672-1681

Klattenhoff C, Theurkauf W (2008) Biogenesis and germline functions of piRNAs. *Development* **135:** 3-9

Konigshoff M, Kramer M, Balsara N, Wilhelm J, Amarie OV, Jahn A, Rose F, Fink L, Seeger W, Schaefer L, Gunther A, Eickelberg O (2009) WNT1-inducible signaling protein-1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis. *J Clin Invest* **119:** 772-787

Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* **11:** 597-610

Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, Chowdhury D, Dykxhoorn DM, Tsai P, Hofmann O, Becker KG, Gorospe M, Hide W, Lieberman J (2009) miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. *Mol Cell* **35:** 610-625

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P,

Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401-1414

Latronico MV, Condorelli G (2009) MicroRNAs and cardiac pathology. *Nat Rev Cardiol* **6:** 419-429

Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, Ganat YM, Menon J, Shimizu F, Viale A, Tabar V, Sadelain M, Studer L (2009) Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461:** 402-406

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75:** 843-854

Lens SM, Voest EE, Medema RH (2010) Shared and separate functions of polo-like kinases and aurora kinases in cancer. *Nat Rev Cancer* **10:** 825-841

Levine AJ, Puzio-Kuter AM (2010) The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science* **330:** 1340-1344

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15-20

Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, Blasco MA, Serrano M (2009a) The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460:** 1136-1139

Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q, Qin B, Xu J, Li W, Yang J, Gan Y, Qin D, Feng S, Song H, Yang D, Zhang B, Zeng L, Lai L, Esteban MA, Pei D (2010) A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7:** 51-63

Li W, Wei W, Zhu S, Zhu J, Shi Y, Lin T, Hao E, Hayek A, Deng H, Ding S (2009b) Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell Stem Cell* **4:** 16-19

Li Z, Yang CS, Nakashima K, Rana TM (2011) Small RNA-mediated regulation of iPS cell generation. *EMBO J* **30:** 823-834

Liao B, Bao X, Liu L, Feng S, Zovoilis A, Liu W, Xue Y, Cai J, Guo X, Qin B, Zhang R, Wu J, Lai L, Teng M, Niu L, Zhang B, Esteban MA, Pei D (2011) MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J Biol Chem* **286:** 17359-17364

Liu GH, Barkho BZ, Ruiz S, Diep D, Qu J, Yang SL, Panopoulos AD, Suzuki K, Kurian L, Walsh C, Thompson J, Boue S, Fung HL, Sancho-Martinez I, Zhang K, Yates J, 3rd, Izpisua Belmonte JC (2011) Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. *Nature* **472:** 221-225

Lowry WE, Richter L, Yachechko R, Pyle AD, Tchieu J, Sridharan R, Clark AT, Plath K (2008) Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci U S A* **105:** 2883-2888

Lyssiotis CA, Foreman RK, Staerk J, Garcia M, Mathur D, Markoulaki S, Hanna J, Lairson LL, Charette BD, Bouchez LC, Bollong M, Kunick C, Brinker A, Cho CY, Schultz PG, Jaenisch R (2009) Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. *Proc Natl Acad Sci U S A* **106:** 8912-8917

Lytle JR, Yario TA, Steitz JA (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104:** 9667-9672

Maherali N, Hochedlinger K (2009) Tgfbeta Signal Inhibition Cooperates in the Induction of iPSCs and Replaces Sox2 and cMyc. *Curr Biol*

Makeyev EV, Zhang J, Carrasco MA, Maniatis T (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell* **27:** 435-448

Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y, Chen G, Gage FH, Muotri AR (2010) A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143:** 527-539

Marson A, Foreman R, Chevalier B, Bilodeau S, Kahn M, Young RA, Jaenisch R (2008) Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell Stem Cell* **3:** 132-135

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123:** 607-620

Meissner A, Wernig M, Jaenisch R (2007) Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* **25:** 1177-1181

Meister G, Landthaler M, Dorsett Y, Tuschl T (2004) Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10:** 544-550

Mendell JT (2008) miRiad roles for the miR-17-92 cluster in development and disease. *Cell* **133:** 217-222

Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A (2008) Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454:** 49-55

Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, Nishikawa S, Tanemura M, Mimori K, Tanaka F, Saito T, Nishimura J, Takemasa I, Mizushima T, Ikeda M, Yamamoto H, Sekimoto M, Doki Y, Mori M (2011) Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell* **8:** 633-638

Moustakas A, Heldin CH (2009) The regulation of TGFbeta signal transduction. *Development* **136:** 3699-3714

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent

stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26:** 101-106

Neenhold HR, Rana TM (1995) Major groove opening at the HIV-1 Tat binding site of TAR RNA evidenced by a rhodium probe. *Biochemistry* **34:** 6303-6309

Ogawa K, Saito A, Matsui H, Suzuki H, Ohtsuka S, Shimosato D, Morishita Y, Watabe T, Niwa H, Miyazono K (2007) Activin-Nodal signaling is involved in propagation of mouse embryonic stem cells. *J Cell Sci* **120:** 55-65

Olsen PH, Ambros V (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216:** 671-680

Otto T, Horn S, Brockmann M, Eilers U, Schuttrumpf L, Popov N, Kenney AM, Schulte JH, Beijersbergen R, Christiansen H, Berwanger B, Eilers M (2009) Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. *Cancer Cell* **15:** 67-78

Overhoff M, Alken M, Far RK, Lemaitre M, Lebleu B, Sczakiel G, Robbins I (2005) Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J Mol Biol* **348:** 871-881

Panopoulos AD, Yanes O, Ruiz S, Kida YS, Diep D, Tautenhahn R, Herrerias A, Batchelder EM, Plongthongkum N, Lutz M, Berggren WT, Zhang K, Evans RM, Siuzdak G, Belmonte JC (2012) The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res* **22:** 168-177

Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, Hochedlinger K, Daley GQ (2008a) Disease-specific induced pluripotent stem cells. *Cell* **134:** 877-886

Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ (2008b) Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451:** 141-146

Peerani R, Rao BM, Bauwens C, Yin T, Wood GA, Nagy A, Kumacheva E, Zandstra PW (2007) Niche-mediated control of human embryonic stem cell self-renewal and differentiation. *EMBO J* **26:** 4744-4755

Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, Goddard AD, Hillan KJ, Gurney AL, Botstein D, Levine AJ (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc Natl Acad Sci U S A* **95:** 14717-14722

Petersen CP, Bordeleau ME, Pelletier J, Sharp PA (2006) Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* **21:** 533-542

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309:** 1573-1576

Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K (2010) Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28:** 848-855

Rana TM (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol* **8:** 23-36

Rozario T, DeSimone DW (2010) The extracellular matrix in development and morphogenesis: a dynamic view. *Dev Biol* **341:** 126-140

Ruiz S, Panopoulos AD, Herrerias A, Bissig KD, Lutz M, Berggren WT, Verma IM, Izpisua Belmonte JC (2011) A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr Biol* **21:** 45-52

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL (2010) Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7:** 64-77

Sanes JR (1989) Extracellular matrix molecules that influence neural development. *Annu Rev Neurosci* **12:** 491-516

Schubert S, Grunweller A, Erdmann VA, Kurreck J (2005) Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J Mol Biol* **348:** 883-893

Seoane J, Le HV, Massague J (2002) Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* **419:** 729-734

Shi Y, Desponts C, Do JT, Hahm HS, Scholer HR, Ding S (2008a) Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3:** 568-574

Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S (2008b) A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2:** 525-528

Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A (2008) Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6:** e253

Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K (2009) Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136:** 364-377

Stadtfeld M, Maherali N, Breault DT, Hochedlinger K (2008) Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2:** 230-240

Subramanyam D, Lamouille S, Judson RL, Liu JY, Bucay N, Derynck R, Blelloch R (2011) Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat Biotechnol* **29:** 443-448

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861-872

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663-676

Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455:** 1124-1128

Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K (2009) Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460:** 1145-1148

Venkatachalam K, Venkatesan B, Valente AJ, Melby PC, Nandish S, Reusch JE, Clark RA, Chandrasekar B (2009) WISP1, a pro-mitogenic, pro-survival factor, mediates tumor necrosis factor-alpha (TNF-alpha)-stimulated cardiac fibroblast proliferation but inhibits TNF-alpha-induced cardiomyocyte death. *J Biol Chem* **284:** 14414-14427

Venkatesan B, Prabhu SD, Venkatachalam K, Mummidi S, Valente AJ, Clark RA, Delafontaine P, Chandrasekar B (2010) WNT1-inducible signaling pathway protein-1 activates diverse cell survival pathways and blocks doxorubicin-induced cardiomyocyte death. *Cell Signal* **22:** 809-820

Vermeulen A, Robertson B, Dalby AB, Marshall WS, Karpilow J, Leake D, Khvorova A, Baskerville S (2007) Double-stranded regions are essential design components of potent inhibitors of RISC function. *RNA* **13:** 723-730

Wagner EF, Nebreda AR (2009) Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* **9:** 537-549

Wang WX, Wilfred BR, Xie K, Jennings MH, Hu YH, Stromberg AJ, Nelson PT (2010) Individual microRNAs (miRNAs) display distinct mRNA targeting "rules". *RNA Biol* **7:** 373-380

Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* **40:** 1478-1483

Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461:** 754-761

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A, Daley GQ, Brack AS, Collins JJ, Cowan C, Schlaeger TM, Rossi DJ (2010) Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7:** 618-630

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448:** 318-324

Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75:** 855-862

Xia HJ, Yang G (2005) Inositol 1,4,5-trisphosphate 3-kinases: functions and regulations. *Cell Res* **15:** 83-91

Xu L, Corcoran RB, Welsh JW, Pennica D, Levine AJ (2000) WISP-1 is a Wnt-1- and beta-catenin-responsive oncogene. *Genes Dev* **14:** 585-595

Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS (2009) MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* **137:** 647-658

Yang CS, Li Z, Rana TM (2011) microRNAs modulate iPS cell generation. *RNA* **17:** 1451-1460

Yoon IK, Kim HK, Kim YK, Song IH, Kim W, Kim S, Baek SH, Kim JH, Kim JR (2004) Exploration of replicative senescence-associated genes in human dermal fibroblasts by cDNA microarray technology. *Exp Gerontol* **39:** 1369-1378

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin, II, Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318:** 1917-1920

Yuan X, Wan H, Zhao X, Zhu S, Zhou Q, Ding S (2011) Brief report: combined chemical treatment enables oct4-induced reprogramming from mouse embryonic fibroblasts. *Stem Cells* **29:** 549-553

Yusa K, Rad R, Takeda J, Bradley A (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat Methods* **6:** 363-369

Zhang Y, Wen G, Shao G, Wang C, Lin C, Fang H, Balajee AS, Bhagat G, Hei TK, Zhao Y (2009) TGFBI deficiency predisposes mice to spontaneous tumor development. *Cancer Res* **69:** 37-44

Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA (2008) In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455:** 627-632

Zoncu R, Efeyan A, Sabatini DM (2011) mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cell Biol* **12:** 21-35

APPENDICES: PUBLISHED MANUSCRIPTS

MicroRNAs encoded by Kaposi's sarcoma-associated herpesvirus regulate viral life cycle.

Small RNA-mediated regulation of iPS cell generation.

microRNAs modulate iPS cell generation.

Molecular Mechanisms of RNA-Triggered Gene Silencing Machineries.