Title
MicroRNAs Promote Induced Pluripotency Through the Regulation of Cooperative Gene Networks

Author
Judson, Robert Laird

Publication Date
2012

Peer reviewed|Thesis/dissertation

MicroRNAs Promote Induced Pluripotency Through the Regulation of Cooperative Gene Networks

by

Robert L Judson

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

To my parents, Sarmite and William, for encouraging me to explore and learn;
To my sister, Lara, for tolerating me as I did;
To my niece, Charlie, for reminding me to continue to do so.

Acknowledgements

These last few years have been among the most enjoyable and successful of my existence, both in and outside of the laboratory. For that, I have everyone to blame but myself...

First and foremost, is Robert Blelloch. I consider myself the luckiest of graduate students to be able write in all honesty, after five years, that I could not have asked for a better mentor. It has been a true privilege to work with, to observe, and to learn from Robert (and to occasionally teach him a thing of two, as well). His creativity, intellect, support, ambition, friendship, and screw-it-we'll-do-it-ourselves mentality are the foundation and motivation for all of the experiments described in the following chapters.

One of Robert's most impressive skills (one I hope I have learned well enough to replicate one day) is his ability to build, maintain, and inspire a diverse, dynamic, creative, brilliant, as-hard-working-as-they-are-fun team of scientists. I have benefitted significantly from being around these wonderful people. Thank you to everyone past and present in the Blelloch lab for being amazing colleagues and friends. Special thanks to Cassie, the lab's pillar of stability, for running as tight a ship a herd of cats will allow. Thank you Archana for the intellectual and emotional support, the venting sessions, and the occasional friendly competition. Thank you Deepa, Yangming, and Collin for showing me the ropes and being great collaborators. Thank you Julia and Tobi for teaching me how much fun teaching can be, and for working hard on my crackpot ideas. Thank you Ron, Raga, Samy, Josh

and Matt for being epic teachers and role-models in both life and science.

My thesis committee, consisting of Andrei Goga, Alan Frankel, and Deepak Srivastava were pivotal for seeing this process through. Thank you for giving me focus and direction when I had spread myself too thin. Special thanks to Andrei for being a wonderful friend and mentor since my first day at UCSF.

Special thanks also to Lisa and Monique for administrating an exceptional graduate program. Your dedication and attention to details have meant that I got to focus all of my energy on research - the best gift someone can give a graduate student.

My education and the work presented here were funded by several different sources over the years, of which all I am very grateful. They include the National Sciences Foundation, the UC Cancer Research Coordinating Committee Fellowship, the Julius R. and Patricia A. Krevans Fellowship, the Northern California Association of Phi Beta Kappa, the National Institutes of Health, and the California Institute of Regenerative Medicine.

Thank you, very much, Alex Amiet of Dharmacon for your enthusiasm and support for my research. Our collaboration was instrumental for this project, and I hope you are as thrilled with the results as I am.

As amazing as my work environment has been, without my friends, I would have never made it through the past five years on this side of sanity (take that as you will). I am particularly grateful for those who have transcended the phases of my life, and remained sources of frequent mirth and support since college, high-

school or beyond - thank you Chris, JP, Jen, Sippy, Sarah, Mike, Roz, Anay, India, Rittik and Matt "of the woman" Donne. In addition, I have been extremely fortunate to have an exceptional number of wonderful friends introduced into my life during my time here in San Francisco. They include Laura, Rob, Dan, Sean, Jan, Mike Sachs, Yelena, Lyndsay, Karen, Marie, Sarah, Dragons, David, Julia, Brett, Timber, Copper, Andy, Aimee, Dale, Megan, Alex, the Sunset Gamers, the Shot Threw the Dart team, the Monday Morning Meat club and a whole host of others. Thank you all for keeping me sane when work became too much, and encouraging more insanity when I worked too much. Special thanks to David Jay for inspiring me to turn a scientific eye on myself and to realize that important complex problems can be solved, even if all the variables can't be isolated.

Finally, and most importantly, I want to acknowledge my foundation, my family. Thanks, Mom and Dad, for, quite literally, everything. Thank you, Lara, James, and Charlie, for the constant love and inspiration. Thank you, The Farm, for being the roots of my curiosity and work ethic. Thank you Frank and Susan for teaching me creativity, and thank you Billy for giving me confidence in my abilities to solve problems. Thank you, Andrea, for providing the love, the inspiration, and the small moments day after day that make them days well-lived.

Thanks again to everyone who has been a part of my research and my life the past five years. The next one-hundred pages wouldn’t have existed without you so, yes, if you read this far into the acknowledgements, you are now bound by honor to finish the entire thesis. Enjoy!

This manuscript contains data originally published in the following reports:

Judson RL, Babiarz JE, Venere M, Blelloch R. Embryonic stem cell-specific MicroRNAs promote induced pluripotency. Nature Biotechnology 2009; 27; 459-61.

Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature 2010; 463; 621-626.

Subramanyam, D, Judson RL*, Lamouille S*, Liu JY, Bucay N, Derynck R, Blelloch R. Multiple targets of miR-302b and miR-372 promote human induced pluripotency. Nature Biotechnology 2011; 29; 443-8.

Judson RL, Greve T, Parchem RJ, Blelloch R. "MicroRNA identification of gene networks that suppress induction of pluripotency." (Manuscript under review)

MicroRNAs Promote Induced Pluripotency Through the Regulation of Cooperative Gene Networks

Robert L Judson

MicroRNAs (miRNAs) are a class of endogenous non-coding RNAs that post-transcriptionally co-regulate networks of genes. The evolutionary history of miRNAs suggests they may play major roles in cell state transitions during the development of complex organisms. Here we characterized the functional role of miRNAs in the generation of induced pluripotent stem cells (iPSCs) from fibroblasts. We found that miRNAs specifically and endogenously expressed in embryonic stem cells (ESCs), including miR-294 and miR-302, greatly enhance the frequency, rate and specificity of fibroblast de-differentiation into iPSCs. Further, the Let-7 miRNAs, a family endogenously expressed in fibroblasts, are potent inhibitors of this same transition. Unexpectedly, a genome-wide screen revealed that the miR-181 family, normally activated in differentiating ESCs, also enhances de-differentiation. To determine the mechanisms by which these miRNA families enhance de-differentiation we developed unbiased high-throughput techniques for identifying and functionally characterizing candidate miRNA targets during this transition. We identified twenty-six miR-294 and miR-181 target genes that act as barriers to de-differentiation, many with cooperative relationships. We further found that both

miR-294 and miR-181 co-regulate Wnt and TGF-Beta signaling, with miR-294 additionally regulating Akt signaling. TGF-Beta inhibition cooperated with Akt or Wnt activation to enhance iPSC generation. We also identified miR-294 as a strong inhibitor of the epithelial-to-mesenchymal transition, a known barrier to de-differentiation. These data establish miRNAs as potent regulators of somatic cell reprogramming, demonstrate that single miRNAs act through co-inhibition of many genes, generate the most comprehensive functionally determined miRNA-mRNA networks to date, and elucidate interactive relationships among genes that normally suppress de-differentiation.

Table of Contents

List of Figures and Tables

Chapter 1: Introduction

Part I: Overview

Embryonic development is the story of a single cell with a single genome undergoing continuous division and differentiation to unfold into the hundreds of highly specific and functional cell types that comprise an adult organism. Understanding the mechanisms by which each cell identity, or cell state, is established, and further, how to manipulate these states, is the technological key for fulfilling the promise of regenerative medicine. Ultimately, it is the differential expression of large networks of co-regulated genes that underlies differentiation of cellular state and function. However, functional identification and verification of gene networks that regulate cell state transitions remains technically challenging. Increasing numbers of studies have demonstrated complex cross talk between major signaling pathways, blurring the boundaries between gene networks traditionally believed to be distinct[1–5]. Gene networks regulated by transcription factor binding vary greatly depending on the co-factors present and the local epigenetic landscape[6–8]. Diverse and significant forms of post-transcriptional regulation result in very little correlation between transcript levels and protein levels[9]. As a result, the interplay between of all of these networks can be highly cell state dependent, making established interactions difficult to generalize.

Recent advances in the techniques of systems biology have been instrumental in generating comprehensive snapshots of gene or protein expression

in specific cellular contexts. Less extensive are the methods to then concurrently manipulate networks of genes in appropriate cell types for functional studies. In this study, we establish a class of genetic elements, called microRNAs (miRNAs), as highly efficient tools for both identifying and functionally verifying novel gene networks **(Figure 1)**. We first demonstrate that miRNAs are potent regulators of a defined cell state transition called directed de-differentiation or reprogramming. We then present an experimental and bioinformatic workflow for using miRNAs to identify novel networks of genes and signaling pathways that cooperate to influence this cell state transition. These data establish a methodology to complement the current repertoire of systems approaches by which miRNAs can be used as molecular probes, highlighting which genes across known pathways, transcriptional networks and ontology groups should be experimentally tested for functional cooperation.

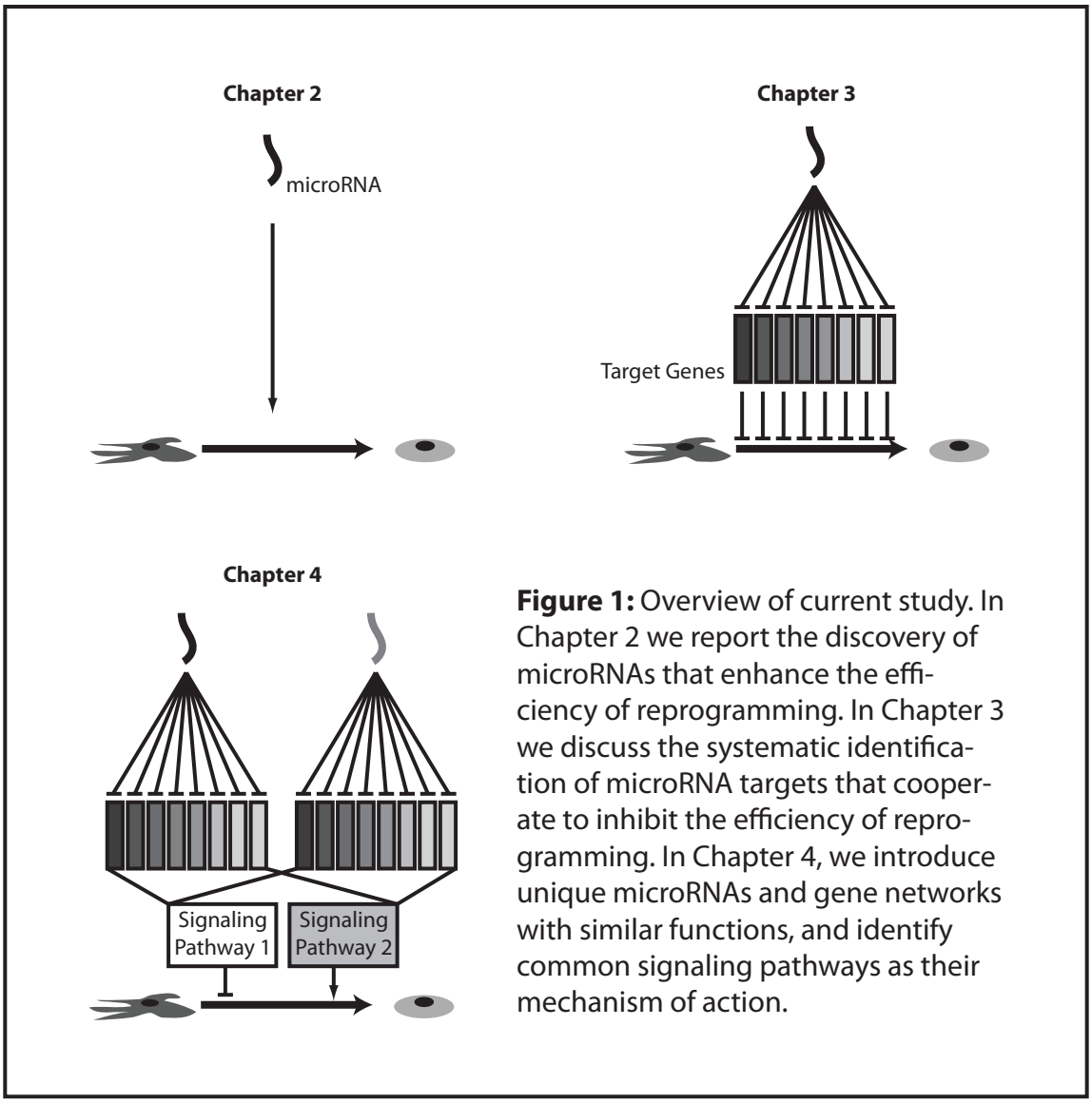


Figure 1: Overview of current study. In Chapter 2 we report the discovery of microRNAs that enhance the efficiency of reprogramming. In Chapter 3 we discuss the systematic identification of microRNA targets that cooperate to inhibit the efficiency of reprogramming. In Chapter 4, we introduce unique microRNAs and gene networks with similar functions, and identify common signaling pathways as their mechanism of action.

Part II: Cell State Transitions

Cell State Transitions

The terms "cell identity" or "cell state" refer to the collection of morphological, physiological, and functional characteristics, as well as the molecular and epigenetic profiles, of a single cell within a developing or mature multi-cellular

organism. Within an organism, cell states differ, not in their genome, which is identical in nearly all cells, but rather in the complex network of genes and genetic elements expressed from that genome. A "cell state transition" is the sequence of alterations to the characteristics and profiles of a specific cell identity causing it to adopt another functionally distinct cell identity. Within mammalian development exist thousands of cell states - some transient, some persisting to the adult organism - each connected through thousands of cell state transitions.

One property that defines cell state is the potential to undergo further cell state transitions. The first cell produced by a fertilized egg, called a zygote, is totipotent, and can give rise to any embryonic or extra-embryonic cell state in the developing organism. During development, cells that retain the ability to transition into any embryonic cell state are considered pluripotent stem cells. Further cell states that can transition into some, but not all, cell states exist in developing and adult tissue and are called multipotent progenitor stem cells. Finally, a cell that no longer undergoes cell state transitions is considered a somatic or differentiated cell.

Due to the potential clinical applications, significant resources have been devoted to developing a deeper understanding of the mechanisms that stimulate and supervise cell state transitions and stabilize cell identities. For example, such studies provide insight into a wide range of developmental diseases, such as congenital cardiac defects[10]. Further, tumorigenesis is an example of an aberrant cell state transition, often marked by the inappropriate re-acquisition of pluripotency[11]. Finally, directed cell state manipulation holds the potential for personalized regenerative medicine, whereby specific cells or even entire organs

could be synthesized using a small biopsy of a patient's own tissue.

The success of studies in these areas has greatly deepened our understanding of cell identity and cell state transitions. Two technological advancements in particular have greatly influenced the field: i) the advent of directed cell state manipulation and ii) the increased resolution and depth of molecular profiling.

Pluripotency is Plastic: Directed Cell State Manipulation

In the classic model of the cell state transitions that drive development, the differentiation of a lineage down more specific and more functionally distinct cell types is linked to a progressive loss of potentiality. A "terminally differentiated" cell, for example, was considered to be incapable of further cell state transitions, barring alterations and mutations to the sequence of the genome itself. This dogma, however, has recently been over-turned. Studies in somatic cell nuclear transfer and cell fusion demonstrated that a terminally differentiated nucleus could be fully reprogrammed into pluripotency when combined with the cytoplasm of a totipotent or pluripotent cell[12]. This was also true of cancer cells, indicating that the mutated and aberrant genome of a cell that had undergone a tumorigenic transition, could still, in a developmental sense, be "normal"[13]. The successful use of defined combinations of exogenously introduced transcription factors to directly de-differentiate somatic cells conclusively proved that, given the correct conditions, terminally differentiated cells could undergo a transition back into a pluripotent state and thereby, indirectly, into any cell in the embryo[14]. Most recently, variations

on this technology have been used to induce not only de-differentiation, but also trans-differentiation, both in vitro and in vivo[15–17]. These data blur definitions such as "pluripotent" or "terminally differentiated", and establish potentiality as a context-specific trait where differentiated cells are still capable of undergoing further cell state transitions.

Increased Resolution and Depth of Profiling

A series of technological advancements in the field of systems biology is also redefining conventional ideas of cell identity and cell state transitions. Systems biology uses a holistic perspective when studying the complex networks of interacting molecules within a biological system, as opposed to studying each interaction within that system individually and in isolation. Systems approaches generally take snapshots of the molecular profiles or molecular interactions within a specific cell type in a specific condition, allowing for global quantification of, for example, protein expression, RNA expression, epigenetic landscapes, protein modifications, protein-protein interactions or protein-nucleic acid interactions. Recent advances in systems technology have vastly increased the resolution of these assays, providing not only much more detailed profiles, but also the ability to profile individual cells, as opposed to populations of cells[18]. Studies using these technologies have demonstrated fluctuation of gene expression between individual cells within what was previously considered a homogenous population[19]. Of similar nature are the discoveries of small but consistent epigenetic differences between populations of cells of functional equivalence[20]. In other words, these technologies are able to identify molecular differences between cells that are functionally

identical. Such findings generate ambiguity around the definitions of cell identity and cell state transitions and demand a more thorough understanding of what measurable properties of a cell are functionally important for that identity.

Molecular Profiles Versus Functional Networks

Current technologies allow for direct manipulation of cell state transitions, and, further, allow for global molecular profiling of individual cells with unprecedented resolution. Despite these advances, fundamental questions surrounding the relationship between the molecular signatures and the functionality of cell states remain. Are two cells the same if they are functionally equivalent but have different molecular profiles or vice versa? To what degree must a molecular profile be identical to cause two cells to be functionally identical? To what degree of dissimilarity must a profile be altered to constitute a transition into another cell state?

These are controversial and unanswered questions that remain difficult to address. Systems biology is, by its nature, an observational and descriptive field, often unable to manipulate the networks defined by it to address the functional consequences. Experimental systems for defining which networks within a given profile are of functional importance are in need. The focus of the current study was to take advantage of several evolutionary traits of miRNAs to first manipulate cell states and then identify networks of cooperating genes responsible for that change in functionality.

Part III: MicroRNAs Regulate Cell State

MicroRNAs: Evolution and Function

MiRNAs are an example of an evolved mechanism of establishing novel networks of gene expression. These 20nt RNAs are encoded by the genome and function by guiding the RNA induced silencing complex (RISC) to specific gene transcripts, resulting in transcript degradation, translation inhibition and overall reduction in protein production[21,22]. MiRNAs recognize target transcripts through a 7-8nt partially complimentary seed sequence primarily located in the 3' untranslated region (UTR) of the mRNA. By targeting the UTRs, miRNAs provide strong evolutionary pressure for genes to form co-regulated networks of protein expression, without altering protein sequence[23].

Evidence suggests that miRNAs likely play pivotal roles in either defining or stabilizing cell states. With the growing number of fully sequenced eukaryote genomes, analysis of gene conservation and duplication has revealed an unexpected and fascinating conclusion - that the diversity and complexity of these species does not correlate with the diversity and complexity of the proteins encoded by their genomes[24]. The number of protein-coding genes within a genome appears unrelated to various measures of organism complexity such as morphological complexity, neuronal number, and diversity of cell types. Additionally, most classes of transcription factors and signaling molecules existed prior to the Cambrian Explosion - an era of unparalleled species diversification - with little gene duplication during this period[25]. In contrast, miRNA number and diversity correlate

extremely well with both the diversification of species during the Cambrian Explosion as well as measurements of organism complexity[26–28]. With each divergence of species, new miRNAs have been identified, whereas previously established miRNAs are rarely lost. The ancient and highly conserved miRNA families are also highly conserved in their tissue type of expression across complex species[29]. These data strongly support a potential role for miRNAs in the establishment, maintenance and/or diversification of specific cell identities.

MicroRNAs in Development and Cancer

As a class of molecules, miRNAs are required for post-implantation mouse development. Embryos genetically lacking *DGCR8*, which encodes an enzyme required for all canonical miRNA maturation, are embryonic lethal[30]. However, *DGCR8* knockout embryos do develop into E5.5 pre-implantation blastocysts[31]. These blastocysts are morphologically identical to their wildtype counterparts, with intact inner cell mass (ICM) and trophectoderm, showing no change in cell number or expression patterns of major transcription factors. Even more striking is the degree of similarity between the transcript profiles, which undergo virtually no change with the removal of all miRNAs. These data suggest that despite the high expression of specific miRNA families in the blastocyst, miRNAs play no role in gene regulation or early development, but become critical shortly after implantation.

Consistent with these *in vivo* studies, self-renewing and morphologically healthy *DGCR8* knockout embryonic stem cells (ESCs) can be maintained in culture, but possess a differentiation defect[30]. ESCs are stable cell lines derived from the cells

of the ICM and display similar characteristics[32]. For example, cells within the ICM are pluripotent and give rise to all tissue in the embryo proper. ESCs are also pluripotent and when injected back into a blastocyst, they integrate and contribute to all tissues in the developing organism. Further, ESCs can be induced to differentiate into a multitude of cell types *in vitro*. However, when miRNA-null ESCs are induced to differentiate, they temporarily turn on markers of differentiation, but are unstable, and ultimately re-activate markers of undifferentiated ESCs[30].

Interestingly, differentiation can be induced in *DGCR8* knockout ESCs through exogenous introduction of miRNAs endogenously expressed in somatic cells[33]. Conversely, the re-introduction of miRNAs endogenously expressed in wildtype ICM and ESCs prevents this induced differentiation. These studies support a role for miRNAs as stabilizers of cell identities, buffering against transitions into related but distinct cell states. Consistent with this model is the bioinformatic observation that the predicted mRNA targets of a miRNA are most frequently expressed not in the miRNA-expressing cell, but rather in cell types temporally or spatially similar to the miRNA-expressing cell during development[23]. Indeed, many miRNAs have been shown to be pivotal regulators of stratifying similar cell states in the development of many tissues, such as the nervous and cardiovascular systems[34,35]. Interestingly, tumorigenesis in several tissue types has been found to be associated with globally reduced miRNAs[36]. Together, these data indicate that miRNAs play pivotal roles in stabilizing or de-stabilizing specific cell states.

MicroRNAs as tools to identify functional gene networks

In this study, we hypothesized that through unbiased identification of the mRNA targets of a miRNA that stabilizes a cell state, we could identify both novel genes and novel gene networks that regulate that state. Given that miRNAs have evolved to co-regulate hundreds of genes, we aimed to use them as biological highlighters, marking networks of cooperating genes expressed during cell state transitions that could then be experimentally tested for functional relevance. To test this hypothesis, we chose to use the directed de-differentiation of mouse embryonic fibroblasts (MEF) as a system for addressing our hypothesis. We chose this system for several reasons. First, directed de-differentiation is a defined transition between cell states with distinct morphology, function and molecular markers[14,37]. Second, the miRNA profiles of both the initial and terminal cell populations were previously well-defined[38]. Third, this transition holds significant clinical potential in the field of regenerative medicine, by providing an unlimited source of patient-specific stem cells. At the onset of this study, barriers to the realization of this potential included the low efficiency of the assay and the use of tumorigenic integrating retroviruses to complete the assay. Finally, although the start and end points of the assay were well-defined, that transition itself was completely undefined, leaving ample room to investigate mechanism.

Part IV: Transcription Factor Mediated Reprogramming

The Reprogramming Assay

During directed de-differentiation, a set of defined transcription factors is introduced into somatic cells. Over the course of days to weeks, various cell state transitions occur, eventually resulting in a small fraction of the original population of cells adopting the morphology and molecular profile of ESCs **(Figure 2)**. Importantly, these cells functionally re-establish both self-renewal and pluripotency, and are thus called induced pluripotent stem cells (iPSCs). As the epigenetic landscape of the original cells is fully reprogrammed during this transition, the term "reprogramming" has become synonymous with "directed de-differentiation", and the terms will be used interchangeably in this report.

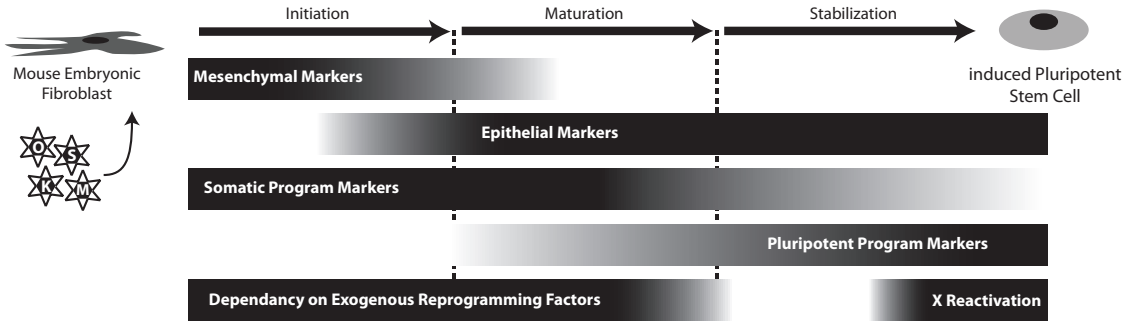


Figure 2: Schematic depicting the phases and markers in Oct4, Sox2, Klf4 & cMyc (OSKM)-directed reprogramming. Infected Mouse Embryonic Fibroblasts (MEFS) first undergo an Initiation Phase, completion of which is marked by down-regulation of mesenchymal markers and up-regulation of both epithelial and pluripotency markers. During the Maturation Phase, the transcription continues to shift away from the somatic program towards the pluripotent program. The Stabilization Phase is defined by the pluripotency program achieving independence from the exogenous reprogramming factors. Reprogramming continues during Stabilization with late markers including silencing of reprogramming factors and X reactivation.

The starting population of cells, the exogenous transcription factors, the

method of factor expression and the media conditions of the assay vary from study to study. The most commonly used starting cell populations in mouse have been MEFs derived from E13.5 embryos containing an *Oct4-GFP* transgene, which is activated in iPSCs[39]. However, the assay has been successfully conducted using a myriad of starting cell populations including T-cells, hair-follicle keratinocytes, liver cells, and stomach cells[40–42]. The most commonly used set of transcription factors are the "Yamanaka Factors" used by Shinya Yamanaka and colleagues in their original description of direct de-differentiation, and include Oct4 (Pou5f1), Sox2, Klf4, and cMyc[14]. In most studies, these factors are introduced either through infection with retrovirus or lentivirus or using a system of "secondary MEFs". In this latter system, dox-inducible lentiviruses are used to generate iPSCs which are, in turn, used to generate chimeric mice[43]. Secondary MEFs containing the integrated lentivirus are derived from the chimeric embryos, resulting in somatic cells that will express the exogenous transcription factors in the presence of doxycycline. Generally, MEFs expressing the transcription factors are then cultured in standard ESC-supporting medium - DMEM with 10-15% FBS supplemented with Lif. Alternatively, the use of defined serum-replacement supplement (KnockOut) instead of FBS, also supports directed de-differentiation, and enhances the efficiency of colony formation[39]. Using these conditions, some of the mechanisms that govern reprogramming have been elucidated since the onset of the present study. These mechanisms are discussed below with a focus on the reprogramming of mouse cells.

The Reprogramming Factors

Of the original four reprogramming factors, Oct4 (Pou5f1) has proven to be

the most irreplaceable for direct de-differentiation. A member of the POU transcription factor family, Oct4 is a well-established master regulator of the pluripotency network. It is expressed in and required for proper development of the ICM, down-regulated upon differentiation, and required for pluripotency in ESCs[44,45]. However, Oct4 is not sufficient to establish or maintain the pluripotency network, but rather co-occupies promoters with other transcription factors, such as Sox2 and Nanog[46–49]. Thus far, in standard MEF reprogramming conditions, Oct4 has only been replaced by two orphan nuclear receptors, Nr5a1 or Nr5a2, which are thought to act through directly binding and activating the Oct4 locus[50].

Sox2 is a transcription factor that directly binds to DNA through an HMG domain, and is stabilized when bound to other co-activators, including Oct4[48]. Sox2 is expressed in the ICM and other developing tissue and is required for development[51]. In standard conditions, Sox2 has been replaced through supplementation of an inhibitor of Tgfbr1[52,53].

Klf4 (Kruppel like factor 4) is expressed in both ESCs and, at lower levels, MEFs. Klf4 belongs to a larger family of transcription factors including Klf2 and Klf5, which, though redundant with each other, have been well established as regulators of proliferation and stemness[54]. Klf4 alone can reprogram primed epiblast stem cells into the naive ESC state[55]. Klf4 is also known to co-occupy many promoters with Oct4 and Sox2[56]. Interestingly, Klf4 has been easier to replace than Oct4 or Sox2, either with other transcription factors (Esrrb), signaling molecules (BMPs) or compounds (Kenpaullone and Valproic Acid)[57–60].

The fourth Yamanaka factor, cMyc, is a well-characterized oncogene which can greatly promote cell proliferation[61]. Not surprisingly, animals grown from iPSCs generated using cMyc retrovirus are significantly more likely to generate spontaneous tumors due to reactivation of the cMyc transgenes[62]. cMyc has other deleterious effects on reprogramming as well, including the generation of highly proliferative non-iPSC transformed cells and, at higher levels, induction of cell death. Fortunately, cMyc was the first Yamanaka factor found to be completely dispensable for reprogramming, acting only as a potent enhancer of the process[63].

The Reprogramming Process

Many other factors have since been discovered that can replace or supplement these four transcription factors. These include other key transcription factors in the pluripotency network such as Nanog and Sall4, inhibitors of epigenomic modification such as BIX, 5'-azaC and VPA, inhibition of regulators of either apoptosis/senescence, such as p53 and Cdkn1a, or regulators of metabolism, such as Ampk[64–6869]. Together with systems approaches to profile global transcript changes in reprogramming cells, the following model describing the mechanisms of directed de-differentiation is emerging **(Figure 2&3)**.

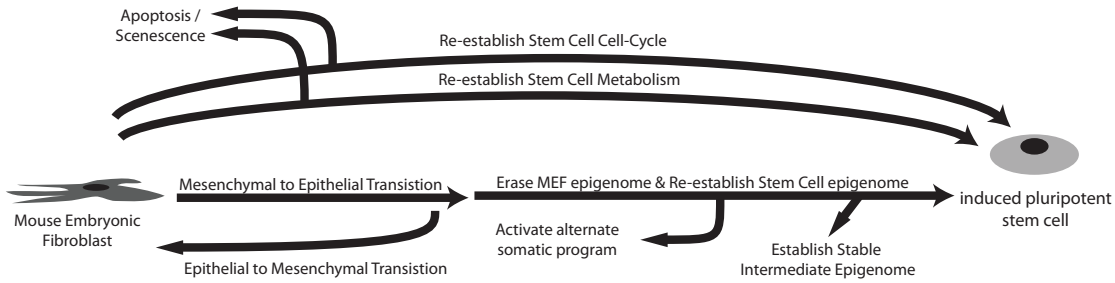


Figure 3: Schematic depicting the required cellular changes that must occur during the reprogramming of MEFs into iPSCs (arrows facing right) and the known barriers to these changes (arrows facing left).

A somatic cell, such as a MEF, upon expression of the reprogramming factors must first undergo an initiation step. Although unclear, molecularly, all that underlies initiation, this stage is marked by the completion of a mesenchymal to epithelial transition (MET) [70,71]. Barriers to successful initiation include induction of cellular senescence and apoptosis although prevention of these events is not sufficient to induce initiation in all cells[72]. Upon successful initiation, a cell must then enter into maturation, which is defined by the activation of several known markers of pluripotency including Nanog and Dnmt3l[70]. Barriers to maturation include reversion to a mesenchymal state (EMT), activation of other somatic cell programs, and, again senescence and apoptosis[37,70,73]. In successfully reprogramming cells, markers of pluripotency progressively activate, eventually resulting in a self-renewing iPSC, which is no longer dependant on exogenous expression of the reprogramming factors, and is thus considered to have entered the stabilization phase, marked both X re-activation and silencing of the retroviral transgenes[37,70]. Throughout this process, on an epigenetic level, MEF-specific DNA methylation and histone modifications must be erased and replaced with the ESC program. Active inhibitors of chromatin modifying enzymes and enhancement of

proliferation both aid this process and thus aid reprogramming[66,74]. Profiling of the epigenetic landscape has revealed that iPSCs continue to reprogram during the stabilization phase, often through many passages, even though, functionally, they are indistinguishable from ESC lines[75]. Physiologically, a metabolic switch must occur from mitochondrial oxidative phosphorylation in MEFs to anaerobic glycolysis in iPSCs, and the cell cycle must shorten and lose the G1-S checkpoint[68,72,76]. Indeed, factors that promote these two switches appear to function independently to enhance de-differentiation. Likewise, all of the other morphological, functional, and molecular differences between a MEF and a ESC must occur, though other switches that act as significant barriers have yet to be defined.

In summary, to reprogram a MEF into an iPSC requires more than just expression of the master transcription factor regulators of the ESC identity, but also requires that the MEF identity must be erased, and that other somatic programs remain silent. All this must happen without activating the myriad of apoptosis or senescence-inducing checkpoints in place to prevent such transitions from occurring. Although the networks of transcription factors that drive these processes are relatively well understood, the networks of upstream and downstream effectors and signaling pathways involved in driving this transition are not.

Signaling Pathways in Reprogramming

One pathway implicated in direct de-differentiation is the Wnt/Beta-catenin pathway. Activated Beta-catenin-expressing retrovirus was one of the original

twenty-four factors screened by Yamanaka and colleagues for its ability to induce pluripotency[14]. Interestingly, removal of this virus from either their twenty-four- or ten- factor cocktails did cause a reduction in reprogramming efficiency, albeit minor compared to Oct4, Sox2, Klf4 or cMyc removal, which warranted its exclusion from their finalized and published four-factor cocktail. In other studies, Oct4/Sox2/Klf4 (OSK)-reprogramming conducted in Wnt3a conditioned media demonstrated a two-fold enhancement in iPSC-colony forming efficiency[77]. However, the concentration of Wnt3a, the required duration of exposure, and the phase of reprogramming affected were undefined. Further, the control-conditioned media enhanced iPSC-colony formation over non-conditioned media to nearly the same degree, leaving open the possibility of secondary secreted factors playing a role. There is significant evidence for a critical role of Wnt/Beta-catenin signaling in maintaining pluripotency, suggesting that this pathway may enhance reprogramming through stabilizing cells that have already reached pluripotency[78–81]. In cell fusion reprogramming experiments, Wnt3a enhances reprogramming, but in a very concentration dependant manner – with both high and low doses having no or negative effects[82]. In short, preliminary studies suggest that Wnt/Beta-catenin signaling likely enhances direct reprogramming, but the mechanism and timing of action remain unaddressed.

TGF-Beta/Activin/Nodal signaling has a more established, though still conflicting, role in direct de-differentiation. Inhibition of TGF-Beta signaling using an Alk5 inhibitor was found to be a potent enhancer of OSKM reprogramming[52,53]. In addition, it could replace Sox2 entirely, earning the molecule the nickname, RepSox.

The mechanism for RepSox-mediated reprogramming enhancement was attributed to transcriptional activation of the pluripotency gene, Nanog. Interesting, reports from other pluripotent systems found that TGF-Beta signaling activated, not silenced, Nanog transcription[83,84]. The mechanism is further confounded by conflicting reports as to whether RepSox acted in the initiation or the stabilization phases of de-differentiation[52,53]. Other ligands similar to TGF-b act in two different stages of induced pluripotency. They include BMP4, which aids in maintaining ESCs in culture, and BMP7, which potently enhances OSKM-reprogramming specifically in the initiation phase of reprogramming[70,85]. Clearly, further characterizations of the timing and the mechanism of TGF-Beta signaling are required.

Akt signaling is suspected to play a role in direct de-differentiation, but has yet to be characterized in context. Activation of Akt signaling allows for ESC culturing the absence of Lif and is required to maintain pluripotency[86–88]. However, in the transition from a somatic cell to a pluripotent stem cell, the role of this pathway is controversial. Akt activation helps fusion-based de-differentiation, inhibits SCNT-based de-differentiation, and remains unstudied in direct de-differentiation[89].

The roles of other pathways, such as MAPK and Lif signaling during de-differentiation are less ambiguous. Lif signaling is required for maintenance of ESCs in culture and is thus included in virtually all medias during reprogramming[90,91]. Its mechanism of action is thought to be through activation of Stat3 signaling, which is required for ESC maintenance, and has been further verified as limiting for de-differentiation into fully pluripotent stem cells[92,9394]. Inhibition of MAPK / MEK

signaling also promotes the maintenance of pluripotent ESCs and, in some conditions, has been shown to enhance de-differentiation[81,95].

Despite pioneering studies implicating many of these pathways in direct de-differentiation, they remain poorly characterized in both their timing and mechanism of action. Further, whether there is cooperating or inhibiting cross-talk between these pathways is completely unknown.

MicroRNAs in Reprogramming

Given their evolutionary history and role in the cell state transitions of early development, we hypothesized that miRNAs would be pivotal players in direct de-differentiation. Indeed, one of the most characterized functions of the RNA-binding protein Lin28, among the first discovered reprogramming factors, is to inhibit a family of miRNAs expressed in somatic cells[96]. This manuscript describes our observations and conclusions on this topic, including the discovery of three families of miRNAs that are potent regulators of reprogramming. Our first report, described again here in Chapter 2, identified members of the miR-290 cluster and miR-302 cluster as enhancers of de-differentiation. Since that time many groups in addition to our own have contributed to our understanding of miRNA regulation of de-differentiation. Significant contributions include i) the identification of novel miRNA enhancers of reprogramming, both related to miR-290/302 (miR-106b, miR-93, miR-130, miR-372) and independent (miR-200, miR-181) ii) the identification of miRNA inhibitors of de-differentiation (Let-7, miR-21, miR-29a, miR-34), iii) the finding of miRNA cocktails that can induce de-differentiation in the absence

exogenous transcription factors (miR-302/miR-367, miR-302/miR-200/miR-369), and iv) and the discovery of miRNA cocktails that can induce trans-differentiation of fibroblasts into neurons or cardiomyocytes[15,70,97–106]. Together these data conclusively demonstrate miRNAs as potent regulators of cell identity, capable of inducing transitions through a variety of mechanisms.

In the chapters that follow, we will describe our contributions to the field, including the identification of the miR-290/302 and miR-181 families as an enhancers of direct de-differentiation, the Let7 family as stabilizers of the MEF identity, the discovery of novel networks of miRNA regulated genes that cooperate to inhibit direct de-differentiation from a MEF to an iPSCs, and the implication of novel cellular processes as barriers to differentiation. We further demonstrate co-regulation of the TGF-B, Wnt and Akt signaling pathways by these miRNA, pinpoint one mechanism of TGF-B inhibition during reprogramming as EMT inhibition, and show cooperation between these pathways during the initiation stage of reprogramming. Together, these data provide a much more in depth understanding of the genes and pathways that regulation reprogramming initiation, functionally define the two most comprehensive experimentally validated miRNA-mRNA networks to date, and present a workflow that can be used in other systems to efficiently identify networks of genes of functional importance in cell state transitions.

Chapter 2: MicroRNAs are Regulators of Pluripotency

Part I: ESC MicroRNAs Promote the Pluripotent Cell State

We first hypothesized that miRNAs endogenously expressed in mouse ESCs were enriched for species that stabilized the ESC state, and would thus enhance direct de-differentiation of MEFs into iPSCs. The miR-290 cluster constitutes over 70% of the entire miRNA population in mouse ESCs[38] **(Figure 4)**. Its expression is rapidly down-regulated upon ESC differentiation, and is silenced in MEFs[38,97,107]. A subset of the miR-290 cluster, called the embryonic stem cell cycle (ESCC) regulating miRNAs, enhances the unique ESC cell cycle[108]. This subset includes miR-291-3p, miR-294, and miR-295. To test whether ESCC miRNAs could promote the induction of pluripotency, we introduced these miRNAs along with retroviruses expressing Oct4, Sox2, and Klf4 (OSK) into MEFs[14]**.** The MEFs carried two reporters: an Oct4-GFP reporter that activates GFP with the induction of pluripotency and ubiquitous expression of a β-galactosidase/neo fusion from the Rosa26 locus[39].

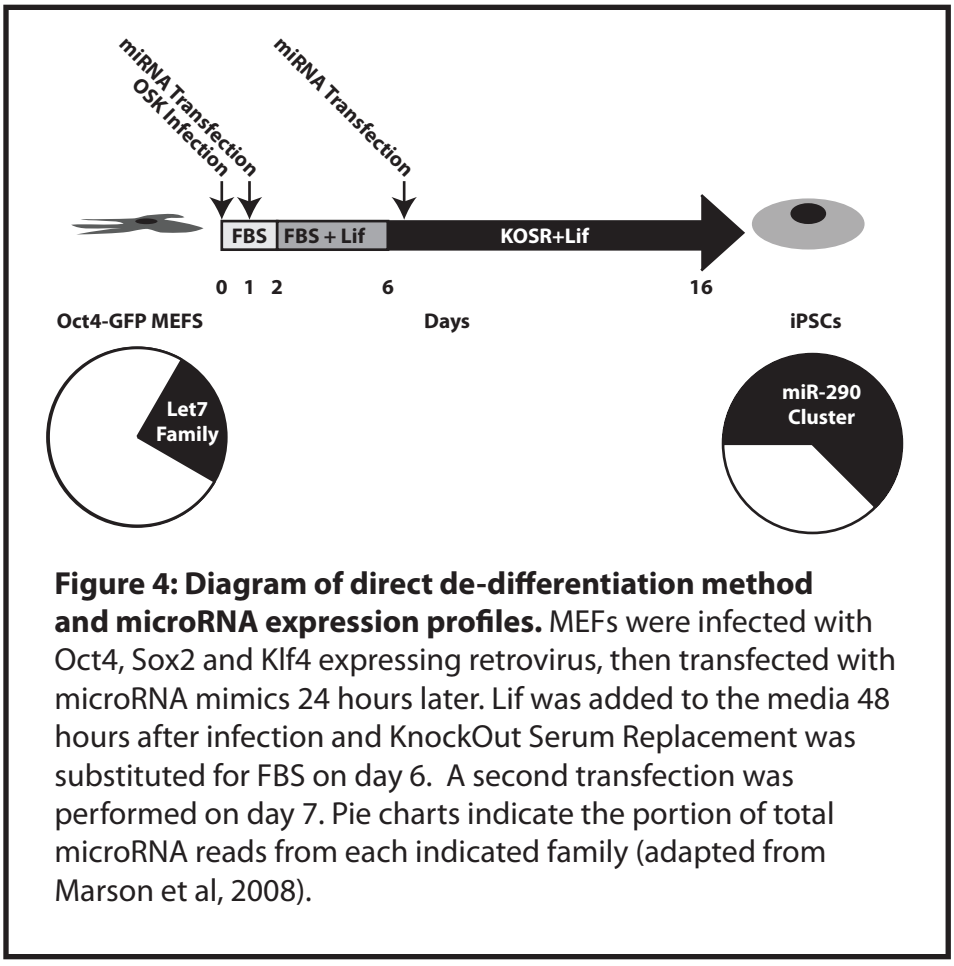


Figure 4: Diagram of direct de-differentiation method and microRNA expression profiles. MEFs were infected with Oct4, Sox2 and Klf4 expressing retrovirus, then transfected with microRNA mimics 24 hours later. Lif was added to the media 48 hours after infection and KnockOut Serum Replacement was substituted for FBS on day 6. A second transfection was performed on day 7. Pie charts indicate the portion of total microRNA reads from each indicated family (adapted from Marson et al, 2008).

MiRNAs were introduced on days 0 and 6 post-infection by transfection of synthesized double-stranded RNAs that mimic their mature endogenous counterparts **(Figure 4)**. This method transiently recapitulates ES-like levels of the miR-290 cluster miRNAs **(Figure 5)**. We used our previously reported media conditions containing defined KnockOut serum replacement instead of FBS[39] **(Figure 4)**.

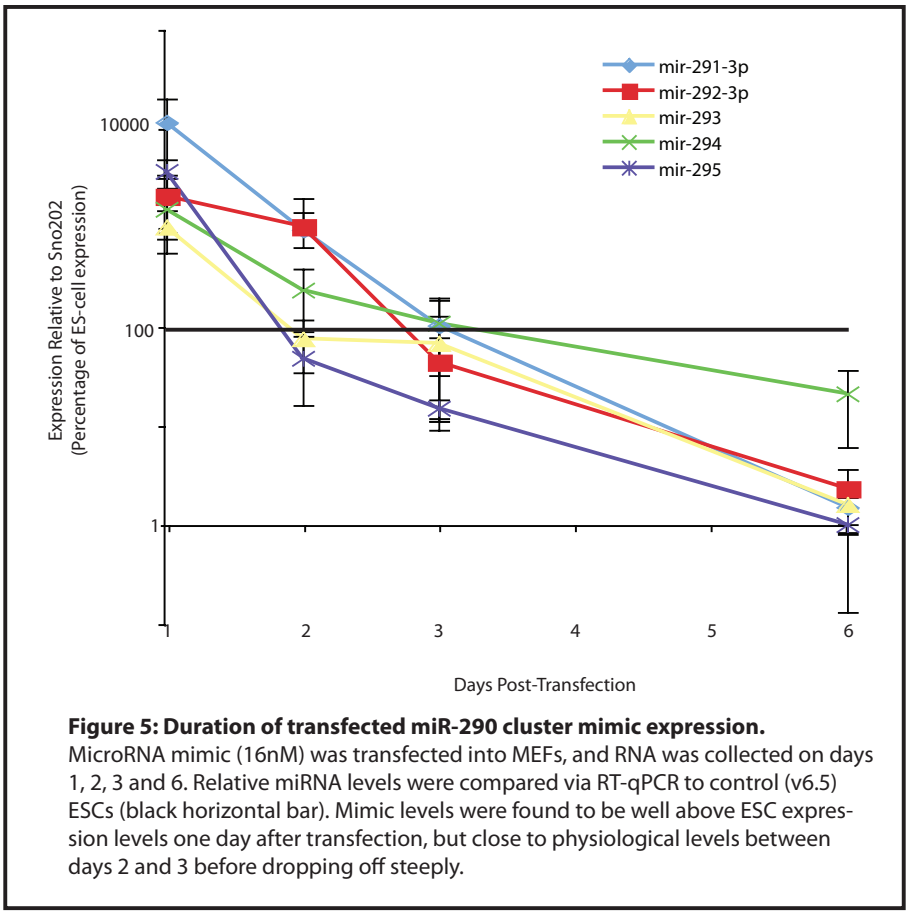


Figure 5: Duration of transfected miR-290 cluster mimic expression. MicroRNA mimic (16nM) was transfected into MEFs, and RNA was collected on days 1, 2, 3 and 6. Relative miRNA levels were compared via RT-qPCR to control (v6.5) ESCs (black horizontal bar). Mimic levels were found to be well above ESC expression levels one day after transfection, but close to physiological levels between days 2 and 3 before dropping off steeply.

OSK plus miR-291-3p, miR-294, or miR-295 consistently increased the number of Oct4-GFP+ colonies as compared to controls transduced with OSK plus transfection reagent **(Figure 6a)**. The miR-294 mimic showed the greatest effects, increasing efficiency from 0.01-0.05% to 0.1-0.3% of transduced MEFs. Introduction of a chemically synthesized miR-294 pre-miRNA similarly enhanced reprogramming (**Figure 6b**). Two other members of the miR-290 cluster that are not ESCC miRNAs, miR-292-3p and miR-293, did not increase colony number **(Figure 6a)**. The ESCC miRNAs share a conserved seed sequence, which largely

specifies target mRNAs (**Figure 6c)**. MiR-302d, a member of another miRNA cluster that has the same seed sequence also enhanced reprogramming (**Figure 6d**). Mutation of the seed sequence in miR-294 blocked the increase in colony number (**Figure 6c&d**). In summary, together with Oct4, Sox2, and Klf4, the ESCC miRNAs and related miRNAs with a common seed sequence promote the de-differentiation of fibroblasts into Oct4-GFP+ ES cell-like colonies.

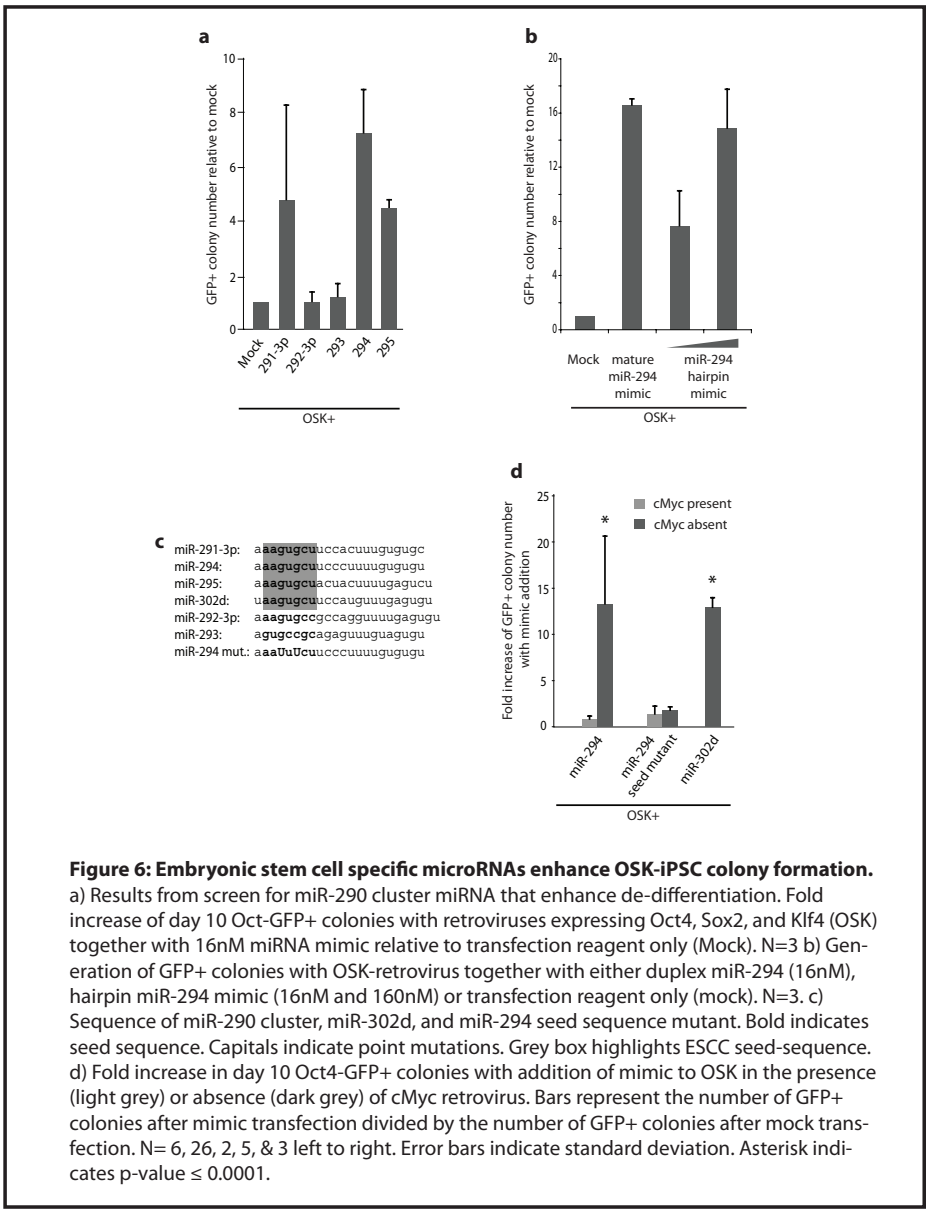


Figure 6: Embryonic stem cell specific microRNAs enhance OSK-iPSC colony formation. a) Results from screen for miR-290 cluster miRNA that enhance de-differentiation. Fold increase of day 10 Oct-GFP+ colonies with retroviruses expressing Oct4, Sox2, and Klf4 (OSK) together with 16nM miRNA mimic relative to transfection reagent only (Mock). N=3 b) Generation of GFP+ colonies with OSK-retrovirus together with either duplex miR-294 (16nM), hairpin miR-294 mimic (16nM and 160nM) or transfection reagent only (mock). N=3. c) Sequence of miR-290 cluster, miR-302d, and miR-294 seed sequence mutant. Bold indicates seed sequence. Capitals indicate point mutations. Grey box highlights ESCC seed-sequence. d) Fold increase in day 10 Oct4-GFP+ colonies with addition of mimic to OSK in the presence (light grey) or absence (dark grey) of cMyc retrovirus. Bars represent the number of GFP+ colonies after mimic transfection divided by the number of GFP+ colonies after mock transfection. N= 6, 26, 2, 5, & 3 left to right. Error bars indicate standard deviation. Asterisk indicates p-value ≤ 0.0001.

Consistent with previous observations that ESCC miRNAs act redundantly**,** mixes of the different ESCC miRNAs did not further enhance reprogramming efficiency[108] (**Figure 7a)**. Therefore, further studies focused on miR-294. Increasing doses of miR-294 further enhanced Oct4-GFP+ colony formation and the Oct4-GFP+ cellular fraction (**Figure 7b&c**). At the highest doses, miR-294 increased the number of colonies to approximately 75 percent of that achieved with OSK and cMyc (OSKM) (0.4-0.7% of starting MEFs) (**Figure 7b&c**). Addition of miR-294 mimic increased the kinetics of OSK reprogramming to rates comparable to OSKM reprogramming **(Figure 8a**). Transfection of miR-294 did not further enhance the reprogramming efficiency of any other three-factor combination or OSKM (**Figure 8b**). Therefore, miR-294 substituted for, but did not further enhance, cMyc's contribution to reprogramming efficiency.

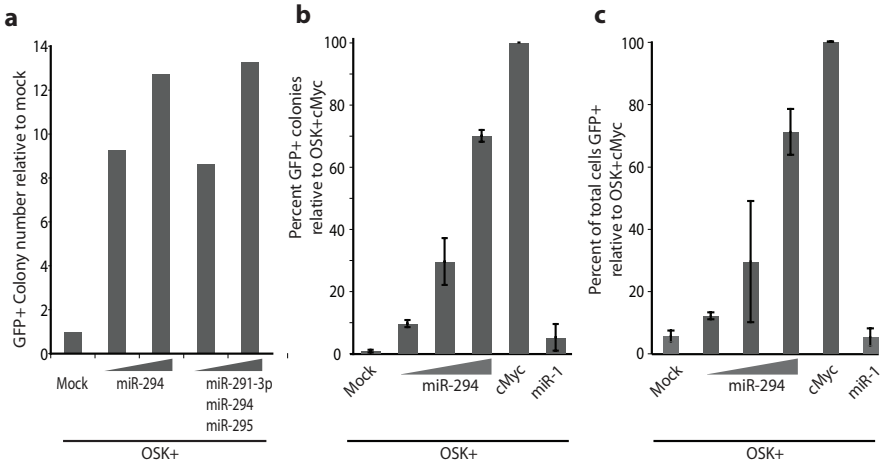


Figure 7: ESCCs enhancement of de-differentiation is redundant and dose-dependant a) Effect of combining ESCC miRNAs on reprogramming. Generation of GFP+ colonies with OSK-retrovirus together with either miR-294 (16m or 48nM) or a mixture of miR-291-3p, miR-294 and miR-295 (5.4nM each or 16nM each) or transfection reagent only (mock). b) Percent day 10 Oct4+ colonies for OSK plus 1.6, 16 and 160nM transfected miR-294 mimic or 160nM miR-1 relative to OSKM. N=3 c) FACS analysis of day 12 GFP+ cells from MEFs infected with OSK-retrovirus and transfected with miRNA mimics. Wedge indicates increasing concentrations (1.6, 6, and 160nM) of mimic. N=3. Error bars indicate standard deviation.

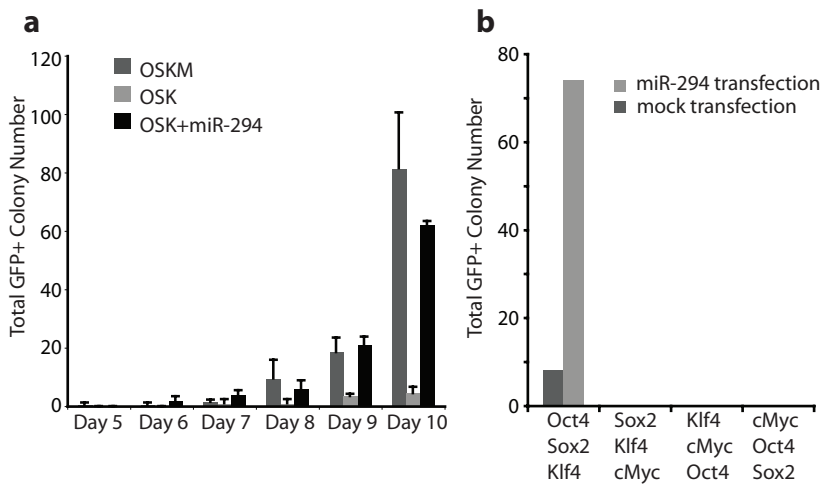


Figure 8: Kinetics of reprogramming and effects of miR-294 on other combinations of reprogramming factors

a) Generation of GFP+ colonies with OSK-retrovirus alone (OSK), with cMyc (OSKM), or with transfection of 16nM miR-294 mimic (OSK+miR-294) over time. GFP+ colonies were counted on days 5-10. First GFP+ microcolonies were visible in OSKM and OSK+miR-294 by day 7 and in OSK by day 8. Error bars represent standard deviation for N=3. b) Generation of GFP+ colonies with combinations of retroviruses expressing Oct4, Sox2, Klf4 or cMyc with and without transfection of miR-294 (16nM).

ESC-like Oct4-GFP+ colonies induced by OSK and miR-294 (miR-294-iPS) were expanded and verified as iPSCs. MiR-294-iPS lines expressed endogenous Oct4, Sox2, and Klf4, while retrovirus expression was silenced **(Figure 9a&b)**. Colonies showed an ESC-like morphology and stained positively for the ES cell markers, Nanog and SSEA-1 **(Figure 9c**). The cell lines had normal karyotypes and efficiently induced teratoma formation with differentiation down all three germ layers (**Figure 9d&e**). Injection of miR-294-iPS cells into blastocysts resulted in high-grade chimeras, with contribution of donor iPS cells to all three germ layers, and to germ line, confirming the occurrence of a complete cell state transition

(Figure 9 f-h).

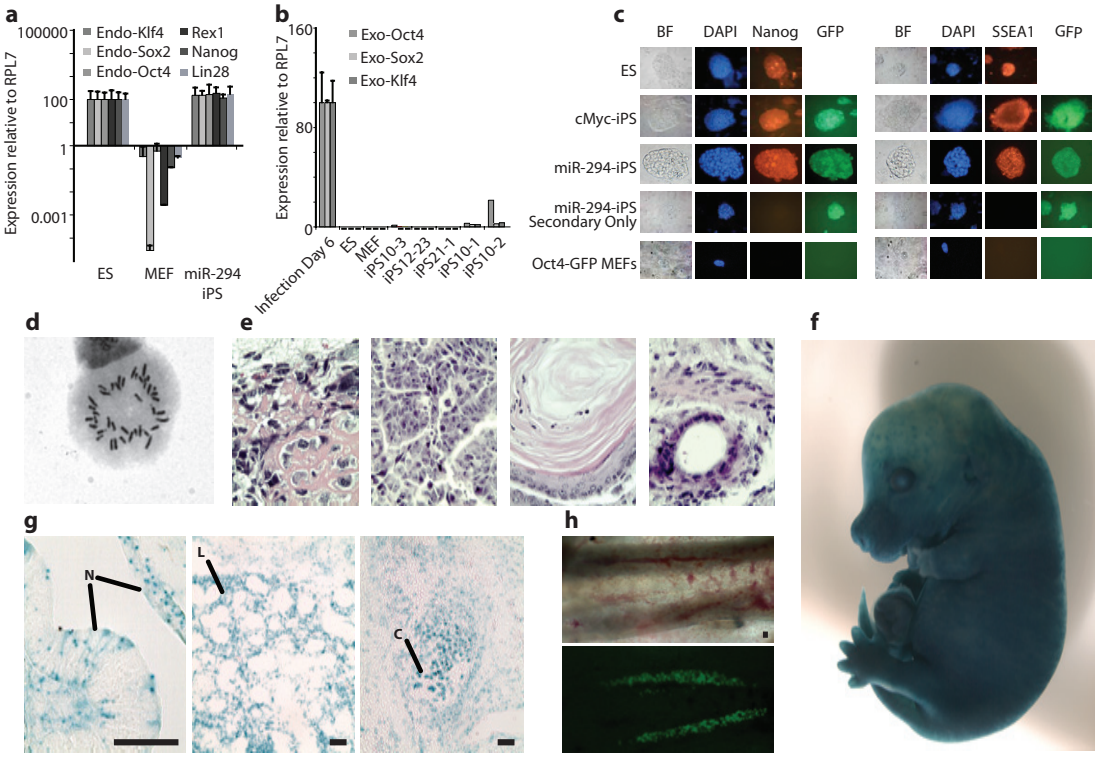


Figure 9: Verification of miR-294 derived iPSC pluripotency
a) RT-qPCR for endogenous pluripotency markers in control (V6.5) ESCs, MEFs, and miR-294-iPSC lines. N=3,3,&5. RPL7 was used as input control. Data was normalized to ESC expression. b) RT-qPCR for exogenous Oct4, Sox2, and Klf4 in MEFs 6 days after viral infection, control ESCs, and MEFs (each N=3) and 5 individual miR-294-iPSC lines. Horizontal black bars indicate Ct>40. RPL7 was used as input control. Data was normalized to MEF expression 6 days after viral infection. c) Brightfield and immunofluorscent images of miR-294-iPSC colonies. Images are representative of six iindependent iPSC lines. Staining controls include ESCs, cMyc-iPSC colonies, miR-294-iPSC colonies with secondary antibody only, and MEFs. d) Representative karyotype of iPSCs induced with OSK and miR-294 mimic. e) H&E staining of miR-294-iPSC derived teratomas. Images depict, left to right, bone, neural tissue, keratinizing squamous epithelial tissue, and glandular tissue. f) An E15 chimera derived from blastocyst injection of miR-294-iPSCs carrying a ubiquitously expressed B-galactosidase reporter. g) X-gal staining demonstrating miR-294-iPSC chimeric contribution to ectoderm (neural tissue, N), endoderm (lung, L), and mesoderm (cartilage, C) h) GFP expression in the genital ridges of E12.5 chimera demonstrates Oct4-GFP miR-294-iPSC contribution to germline. All error bars indicate standard deviation.

The mechanism for how ESCC miRNAs substitute for cMyc in reprogramming is not entirely clear. However, bioinformatic analysis of ES ChIP-seq data[109] showed that both c-Myc and n-Myc bind to the promoter region of the miR-290 cluster (**Figure 10a**). Oct4, Sox2 and Nanog have also been reported to bind the promoter of the miR-290 cluster[38].

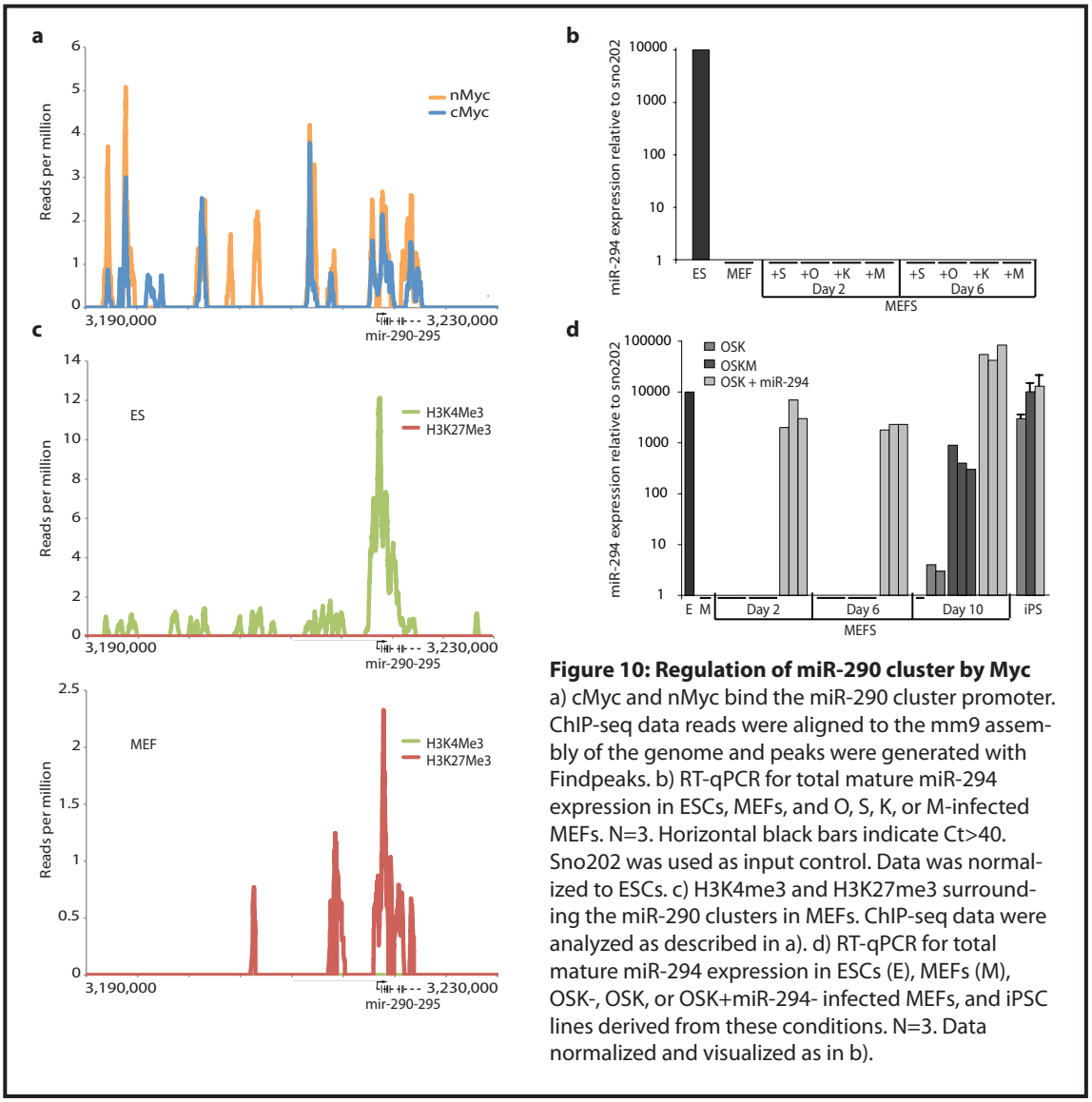
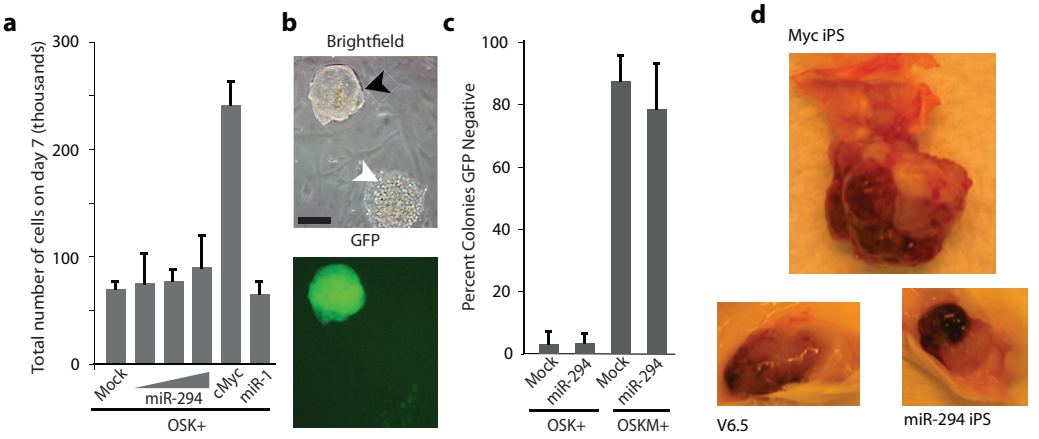


Figure 10: Regulation of miR-290 cluster by Myc
a) cMyc and nMyc bind the miR-290 cluster promoter. ChIP-seq data reads were aligned to the mm9 assembly of the genome and peaks were generated with Findpeaks. b) RT-qPCR for total mature miR-294 expression in ESCs, MEFs, and O, S, K, or M-infected MEFs. N=3. Horizontal black bars indicate Ct>40. Sno202 was used as input control. Data was normalized to ESCs. c) H3K4me3 and H3K27me3 surrounding the miR-290 clusters in MEFs. ChIP-seq data were analyzed as described in a). d) RT-qPCR for total mature miR-294 expression in ESCs (E), MEFs (M), OSK-, OSK, or OSK+miR-294- infected MEFs, and iPSC lines derived from these conditions. N=3. Data normalized and visualized as in b).

Transduction of cMyc, Oct4, Sox2, or Klf4 expressing retrovirus individually failed to induce expression of the miR-290 cluster in fibroblasts (**Figure 10b**). Analysis of ChIP-seq data for different histone modifications showed that the miR-290 promoter is H3K27 methylated in MEFs, a modification associated with transcriptional silencing[110] (**Figure 10c)**. In contrast, the promoter is H3K4 methylated in ES cells, a modification associated with transcriptional activity. Therefore, these transcription factors likely can only induce the expression of the

miR-290 cluster as cells replace promoter-associated H3K27 with H3K4 methylation during the reprogramming process. Indeed, with OSKM transduction, miR-294 was robustly activated late in the reprogramming process, similar to the reported timing for expression of endogenous Oct4, and other critical members of the core ES machinery (**Figure 10d**)[37,111]. These data suggest that miR-294 is downstream of cMyc, but requires epigenetic remodeling for expression.

Interestingly, the downstream effects of the ESCC miRNAs versus cMyc on the reprogramming process were not identical. Unlike cMyc, miR-294 did not promote proliferation of MEFs early in the reprogramming process (**Figure 11a)**. Furthermore, as previously reported, approximately 80% of the OSKM colonies failed to express GFP and lacked ESC-like morphology[63] **(Figure 11b&c)**. In contrast, OSK+miR-294 produced a predominantly uniform population of ESC-like GFP+ colonies (**Figure 11c**). The Oct4-GFP– colonies were induced by cMyc, not inhibited by miR-294, as the introduction of both produced a similar number of GFP–, non-ESC-like colonies as cMyc alone (**Figure 11c**). Finally, when cells were injected into immunodeficient mice to produce teratomas, more than a third of the teratomas resulting from cMyc-iPS cells invaded into the underlying body wall, while none of teratomas resulting from miR-294-iPS cells did so **(Figure 11d&e)**. These findings show that while miR-294 can substitute for cMyc to enhance reprogramming, its effects on the cell population are not identical, as it induces a more uniform population of ESC-like self-renewing cells.



e

Cell Lines	Mice Injected	Teratomas Formed	Teratomas Invasive	Day Post-injected Collected	Percent Teratoma	Percent Invasive
V6.5	6	10	0	1x15, 4x31, 1x55	**100**	**0**
OSK+Myc						
10-4	1	2	0	24		
10-5	1	2	1	15		
10-6	1	2	1	24		
12-2	2	3	0	31	**78**	**43**
12-3	3	4	1	31		
12-4	1	0	0	31		
12-5	1	0	0	31		
14-1	1	1	0	31		
14-2	1	1	0	31		
OSK+294						
10-1	1	2	0	31		
10-2	1	2	0	31		
10-3	1	2	0	31		
12-19	2	3	0	31		
12-20	2	1	0	31	**100**	**0**
12-21	2	2	0	31		
12-22	2	3	0	15		
12-23	1	1	0	15		

Figure 11: Effect of miR-294 on proliferation and transformation
a) Total cell number during reprogramming. Cells were counted on day 7 after infection with OSKM or OSK +/- miRNA mimic. Concentrations of miR-294 mimic: 1.6, 16, and 160nM. Concentration of miR-1 mimic: 160nM. b) GFP negative colonies in presence of cMyc. Oct4-GFP+, ESC-like colonies (black arrow) and GFP-negative, non-ESC-like colonies (white arrow). c) Quantification of number of day 10 GFP-negative colonies after infection with OSKM or OSK +/- miR-294 mimic. All error bars indicate standard deviation of N=3. d) Representative images of non-invasive (v6.5 and miR-294-iPSC) and invasive (Myc iPSC) teratomas. All images to scale. e) Number of invasive and non-invasive tumors with different cell lines injected. Columns display from left to right, independent cell lines, number of mice injected, number of total teratomas isolated, number of total teratomas found to be invasive, and the number of days after injection teratomas were harvested. Percent teratoma refers to the percentage of cell lines that formed teratomas. Percent invasive refers to the percentage of teratomas found to migrate through the underlying body wall.

Part II: Fibroblast MicroRNAs Inhibit the Pluripotent Cell State

Having demonstrated that physiologically relevant miRNAs, normally expressed in pluripotent stem cells, can enhance the transition of somatic cells into their cell type of origin, we next hypothesized that if the general function of miRNAs were to stabilize a cell state, than the inverse should also be true. MEF specific miRNAs should stabilize the fibroblast cell state, and thereby inhibit the transition into iPSCs. Profiling studies identified the let-7 family as highly expressed in MEFs, but silenced in ESCs[38] **(Figure 4)**. Indeed, upon induction of differentiation in culture, ESCs up-regulate let-7, while simultaneously silencing the miR-290 cluster[112]. Further, exogenous introduction of let-7 into *Dgcr8* knockout ESCs is sufficient to induce differentiation[112]. Co-introduction with miR-294, however, block this effect, demonstrating antagonistic functions of these miRNAs.

We hypothesized that inhibition of the let-7 family would enhance reprogramming of somatic cells to iPSCs. To test the effect of the let-7 family on reprogramming, we used a let-7 antisense inhibitor. This inhibitor was able to suppress several let-7 family members simultaneously **(Figure 12a)**. Again, we used Oct4-GFP transgenic MEFs to quantify changes in reprogramming efficiencies. MEFs were transduced with retroviral vectors expressing Oct4, Sox2, Klf4, with or without c-Myc on day 0, as well as being transfected with let-7 or a control inhibitor on days 0 and 6. When OSK-retrovirus was used, let-7 inhibition increased the number of GFP-positive colonies on day 10 by 4.3-fold compared to mock whereas a

control inhibitor had no significant effect **(Figure 12b)**. In the presence of all four transcription factors, let-7 inhibition resulted in a 1.75-fold increase **(Figure 12c)**.

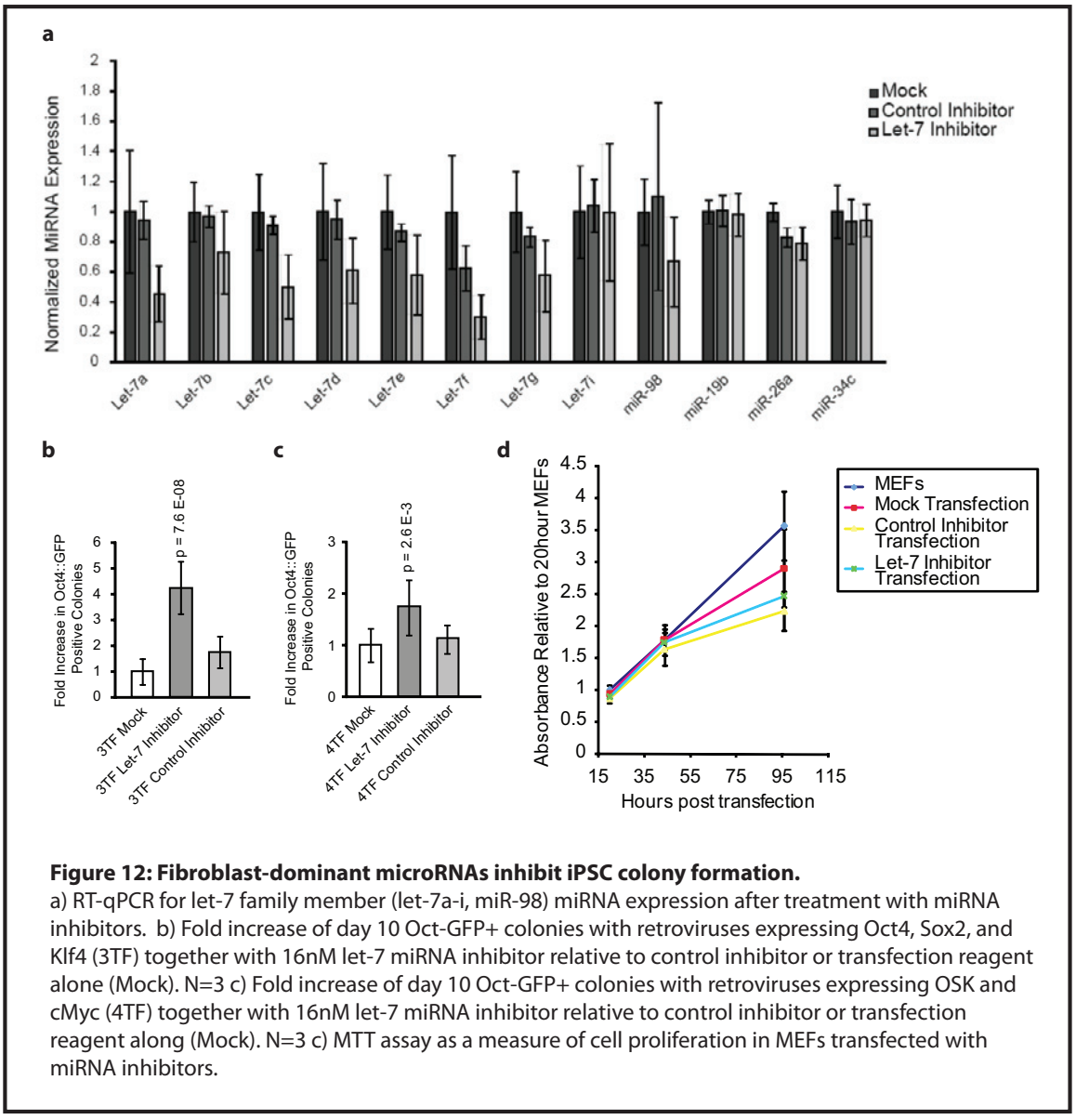


Figure 12: Fibroblast-dominant microRNAs inhibit iPSC colony formation.
a) RT-qPCR for let-7 family member (let-7a-i, miR-98) miRNA expression after treatment with miRNA inhibitors. b) Fold increase of day 10 Oct-GFP+ colonies with retroviruses expressing Oct4, Sox2, and Klf4 (3TF) together with 16nM let-7 miRNA inhibitor relative to control inhibitor or transfection reagent alone (Mock). N=3 c) Fold increase of day 10 Oct-GFP+ colonies with retroviruses expressing OSK and cMyc (4TF) together with 16nM let-7 miRNA inhibitor relative to control inhibitor or transfection reagent along (Mock). N=3 c) MTT assay as a measure of cell proliferation in MEFs transfected with miRNA inhibitors.

The effect of the let-7 inhibitor is not due to enhanced proliferation of the MEFs as there was actually a subtle decrease in proliferation after transfection of either the let-7 or control inhibitor **(Figure 12d)**. Immunofluorescence confirmed expression of Nanog in reprogrammed cells **(Figure 13a)**. Furthermore, the resulting iPS cells expressed endogenous pluripotency markers at levels similar to wild-type ESCs and

did not express the exogenously introduced factors **(Figure 13b&c)**, as expected for fully reprogrammed cells. These findings show that inhibition of the let-7 family of miRNAs enhances the reprogramming of somatic cells.

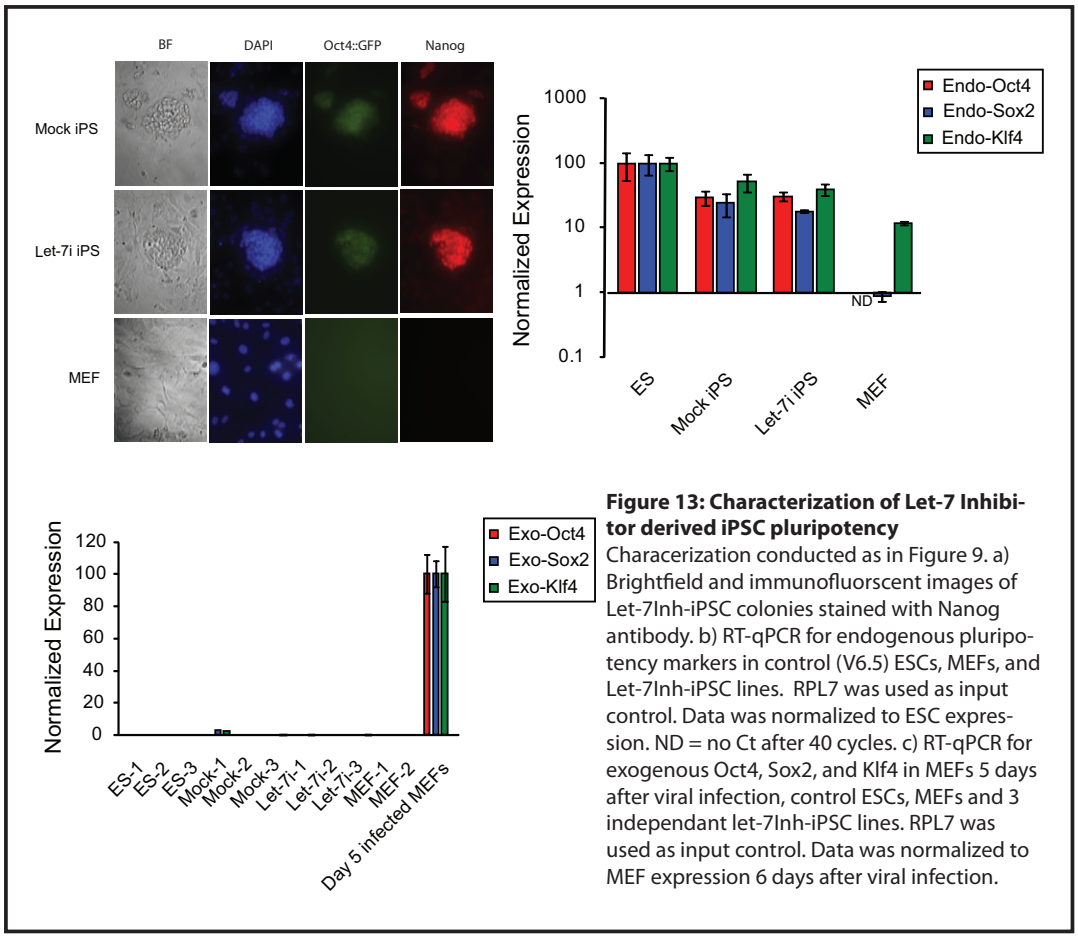


Figure 13: Characterization of Let-7 Inhibitor derived iPSC pluripotency
Characerization conducted as in Figure 9. a) Brightfield and immunofluorscent images of Let-7Inh-iPSC colonies stained with Nanog antibody. b) RT-qPCR for endogenous pluripotency markers in control (V6.5) ESCs, MEFs, and Let-7Inh-iPSC lines. RPL7 was used as input control. Data was normalized to ESC expression. ND = no Ct after 40 cycles. c) RT-qPCR for exogenous Oct4, Sox2, and Klf4 in MEFs 5 days after viral infection, control ESCs, MEFs and 3 independant let-7Inh-iPSC lines. RPL7 was used as input control. Data was normalized to MEF expression 6 days after viral infection.

Part III: Conclusions

These data characterize a potent role for miRNAs in regulating somatic cell reprogramming. Over-expression of miRNAs normally expressed in ESCs increases the efficiency of MEF de-differentiation, whereas miRNAs endogenous to MEFs inhibit it. Further, the use of miRNA mimics appears to have advantages over virus-based transcription factor delivery, generating cleaner reprogramming assays, with less non-iPSC colonies and less tumorigenic potential.

Chapter 3: MiR-294 Enhances Reprogramming through Networks of Cooperating Genes

Part I: Identification of High-Confidence miR-294 Targets

In the next set of experiments, we sought to identify which miR-294 targets were responsible for its role in enhancing de-differentiation. We hypothesized that through unbiased identification and functional characterization of miR-294 targets, we could identify a network of genes that cooperate in the inhibition of this cell state transition. We first sought to generate a database of high confidence gene targets of miR-294. Traditionally, miRNA targets are identified by first predicting likely targets using computational algorithms, followed by experimental verification of individual candidates. This technique is notoriously inefficient, as the most accurate algorithms retain false positive and false negative rates of approximately 50%, and most cannot take into account cell type specific effects or variations in splicing and UTR length[113,114]. Recently, these approaches have been complimented by the development of several systems approaches for globally identifying candidate miRNA targets in the appropriate cellular context, including mimic-induced mRNA destabilization (MIMD), Ago2-CLIP, SILAC, ribosomal pull-down and biotin-tagged mimic pull-down[97,113–117]. Each of these approaches quantifies different molecular consequences of a miRNA-mRNA interaction and, interestingly, when directly compared, candidate targets sets derived from these various techniques overlap

poorly[115]. Thus, it remains unclear which technique most accurately identifies bona fide miRNA targets, defined as direct miRNA-mediated reduction in protein levels.

To generate a high confidence target list for miR-294, we first consolidated the available experimental and computational data on candidate miR-294 targets. MiR-294 is expressed in ESCs as part of a mouse-specific locus known as the miR-290 cluster[107]. This cluster expresses eleven distinct miRNAs, three of which (miR-291a-3p, miR-294 and miR-295) are highly conserved in total sequence, each containing the identical AAGUGCU seed sequence **(Figure 6c)**. As many of the above candidate target identification techniques cannot differentiate between targets of miRNAs with the same seed sequence expressed in the same cell type, we conducted our analysis for grouped miR-291a-3p/294/295, using three independent means of target identification. 1) Previously, MIMD-microarray was used to identify targets of miR-294 in the context of ESCs[97]. Synthesized mature miR-294 mimetics were transfected into ESCs genetically lacking *DGCR8*, a gene essential for miRNA biogenesis, and changes in mRNA levels were detected via microarray **(Figure 14a)**. As *DGCR8-/-* ESCs lack all canonical miRNAs, the effects of individually introduced miRNA mimics are likely exaggerated and more easily detectable. From this dataset, 1079 genes were identified as being down-regulated with the addition of miR-294 and containing at least one miR-294 binding site in the 3' UTR or open reading frame **(Figure 14a&b)**. 2) Another dataset of potential miR-294 targets has been generated through Ago2-CLIP-seq in mESCs[118]. Anti-Ago2 antibody was used to precipitate RISC-associated RNA and associated genes were identified via deep sequencing. After bioinformatic filters were applied, 200 candidate targets were

identified. 3) Finally, we chose the TargetScan algorithm to computationally predict miR-294 targets, as this technique has been experimentally verified as the most accurate bioinformatic approach for predicting targets, and generated a list of 417 candidates[113,114].

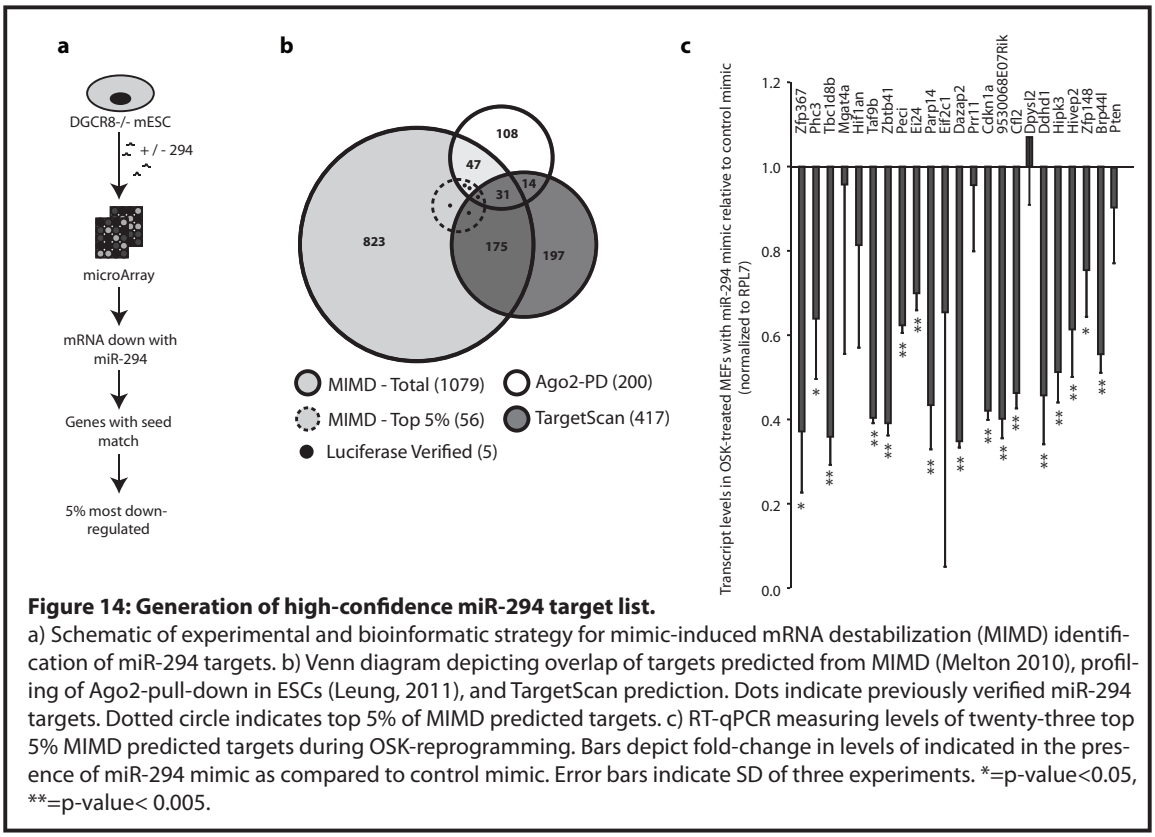


Figure 14: Generation of high-confidence miR-294 target list.
a) Schematic of experimental and bioinformatic strategy for mimic-induced mRNA destabilization (MIMD) identification of miR-294 targets. b) Venn diagram depicting overlap of targets predicted from MIMD (Melton 2010), profiling of Ago2-pull-down in ESCs (Leung, 2011), and TargetScan prediction. Dots indicate previously verified miR-294 targets. Dotted circle indicates top 5% of MIMD predicted targets. c) RT-qPCR measuring levels of twenty-three top 5% MIMD predicted targets during OSK-reprogramming. Bars depict fold-change in levels of indicated in the presence of miR-294 mimic as compared to control mimic. Error bars indicate SD of three experiments. *=p-value<0.05, **=p-value< 0.005.

Consistent with previous observations, we observed only partial overlap of the candidate lists generated by these three techniques **(Figure 14b)**. To determine which subset of these gene sets were most enriched for bona fide targets, we compiled a list of genes previously verified to be translationally inhibited by miR-miR-291a-3p/294/295 via luciferase reporter assays. In this assay, the effect of exogenous miRNA on translation can be directly quantified by fusing the UTR of the candidate gene to a luciferase ORF[108,119]. Only the MIMD-derived gene set contained

all five verified miR-291a-3p/294/295 targets (Cdkn1a, Lats2, Rbl2, Ei24, Casp2) **(Figure 14b)**. We next sought methods by which to refine this list, while retaining the known true positives. As previous systems approaches to identifying miRNA targets have consistently concluded that the degree of miRNA-induced translational inhibition is well correlated with the degree of miRNA-induced mRNA destabilization, we filtered the MIMD candidate set based upon degree of miR-294 mediated mRNA knockdown[113,114,116]. By choosing the top 5% of genes most reduced by miR-294, we significantly reduced the size of the candidate list (55 genes), without excluding any of the five known targets **(Figure 14b)**.

To verify that our refined gene set was enriched for miR-294 targets in the context of somatic cell de-differentiation, we transfected MEFs with either miR-294 or control mimic 24 hours after infection with Oct4, Sox2 and Klf4-expression retrovirus (OSK). Total RNA was collected 72 hours after transfection and RT-qPCR was conducted using primers against 22 of the genes. Of these, 17 genes were significantly down-regulated on a RNA level in the presence of miR-294 **(Figure 14c)**. These data indicate strong enrichment in our gene list for endogenous targets of miR-294 during de-differentiation. We also added to our list Pten, a gene previously shown to be a target of the miR-294 orthologue, miR-302, in human cells[120], resulting in a list of 56 high confidence miR-294 targets **(Table 1)**.

Table 1: High Confidence miR-294Target List

Gene ID	Gene Name	Gene ID	Gene Name
213673	9530068E07Rik	15944	Irgm1
237615	Ankrd52	16784	Lamp2
55951	Brp44l	50523	Lats2
12366	Casp2	13590	Lefty1
216001	Cbara1	269181	Mgat4a
235505	Cd109	69188	Mll5
78334	Cdk19	547253	Parp14
12575	Cdkn1a	23986	Peci
83815	Cenpq	18645	Pfn2
74107	Cep55	241915	Phc3
12632	Cfl2	11757	Prdx3
12753	Clock	105787	Prkaa1
225995	D030056L22Rik	270906	Prr11
23994	Dazap2	19211	Pten
67665	Dctn4	19334	Rab22a
114874	Ddhd1	19651	Rbl2
12934	Dpysl2	20706	Serpinb9b
13663	Ei24	67043	Syap1
236511	Eif2c1	407786	Taf9b
80898	Erap1	245638	Tbc1d8b
59079	Erbb2ip	21822	Tgtp1
209416	Gpkow	71929	Tmem123
231086	Hadhb	72477	Tmem87b
319594	Hif1an	22223	Uchl1
15259	Hipk3	30940	Usp25
15273	Hivep2	226470	Zbtb41
15441	Hp1bp3	22661	Zfp148
		238673	Zfp367
		98999	Znfx1

Part II: miR-294 Targets are Enriched for Inhibitors of Reprogramming

We next characterized which high confidence miR-294 targets functionally inhibited somatic cell de-differentiation using a siRNA approach to knock-down each target individually. Pools of siRNA with four different sequences against single genes were synthesized to minimize off target effects (On-Target Plus - Dharmacon). MEFs containing an Oct4-GFP transgene were infected with retrovirus expressing OSK then transfected with siRNAs after 24 hours **(Figure 15a)**. The number of Oct4-

GFP positive colonies was tallied on day 16 post-transfection to measure de-differentiation efficiency. Transfection of miR-294 mimic and four different non-targeting siRNA sequences were used as positive and negative controls, respectively, and the experiment was performed in triplicate using independently generated MEFs and OSK virus. The number of day 16 Oct4-GFP+ colonies in each mimic-containing well was compared to negative controls using strictly standardized mean difference (SSMD)[121]. siRNA against 10 of the 56 miR-294 targets demonstrated fold increases in iPSC formation that achieved our cut-offs for statistical significance (p-value <0.01 and SSMD >2) **(Figure 15b&c)**. One of these genes, Cdkn1a, was previously identified as a potent inhibitor of de-differentiation[72]. To determine if the effects on de-differentiation were due to targeting the expected gene, and not an off-target effect, pools of four siRNA with independent sequence (siGenome - Dharmacon) were synthesized. Both sets of siRNA were introduced into the de-differentiation assay. Of the 10 miR-294 target siRNA screen hits, 8 siRNAs (against Cdkn1a, Zfp148, Hivep2, Ddhd1, Dpysl2, Pten, Cfl2 and 9530068E07Rik) enhanced de-differentiation with both pools of siRNA, indicating that the enhanced efficiency is due to inhibition of the target gene **(Figure 15d)**. As both pools of siRNA against Brp44l and Hipk3 were not effective, these genes were removed from further analysis.

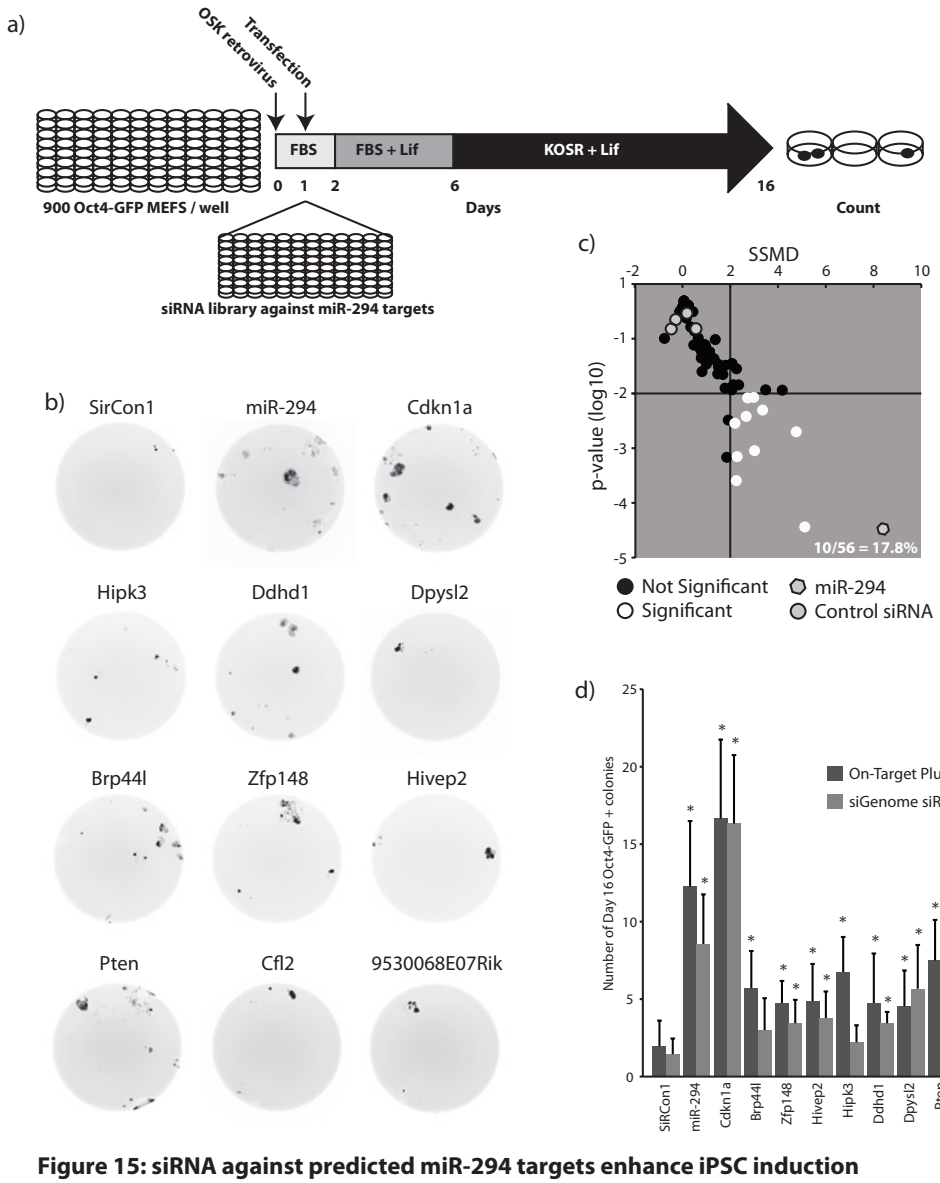


Figure 15: siRNA against predicted miR-294 targets enhance iPSC induction
a) Schematic of reprogramming screen. b) Full well represented images of Oct4-GFP expression day 16 after OSK-infection. Wells were transfected with either non-targeting control siRNA (siRCon), miR-294 or siRNA pools against indicated genes. c) Quantification of de-differentiation efficiency in siRNA screens. Axis are average SSMD of wells transfected with experimental siRNA and 4 different negative control siRNA, and p-value of all experimental wells against all negative control wells (grey circles). White dots indicate strongly significant hits (SSMD>2, p-value<0.01). n=3. d) Follow-up of hits from c) using two independent pools of siRNA transfected on day 1 of OSK-reprogramming and measuring number of day 16 Oct4-GFP+ colonies. Dark grey bars = OnTarget-Plus, Light grey bars = siGenome. *=p-value < 0.05 compared to non-targeting siRNA control (siRCon1) as determined by Student T-test, n=4.

To verify siRNA targeting of these genes, the assay was repeated and total RNA was collected 72 hours post-transfection for RT-qPCR analysis. Each siRNA

pool significantly reduced expression of the expected target genes **(Figure 16a)**. Our previous MIMD-RTqPCR indicated that each of the 8 gene hits, other then Dpysl2 and Pten, demonstrated significant mRNA reduction in the presence of miR-294 **(Figure 14c)**. To determine whether miR-294 inhibited Pten and Dpysl2 protein levels, total protein was collected from OSK-infected MEFs with and without miR-294 transfection and analyzed via Western. Total Pten and Dpysl2 protein were down-regulated by miR-294 **(Figure 16b)**. To determine whether translation of these genes was directly suppressed by miR-294, regions of the Pten and Dpysl2 3'UTRs containing predicted miR-294 binding sites were cloned and fused to a luciferase open reading frame **(Figure 16c)**. Luciferase reporters were co-transfected with miR-294 into Dgcr8 KO ESCs, and luciferase activity was measured after 24 hours. Luciferase activity was significantly reduced in the Dpysl2 construct in the presence of the microRNA and this repression required the miR-294 binding site, verifying Dpysl2 as a direct miR-294 target **(Figure 16d)**. Interestingly, miR-294 had no effect on the Pten construct, suggesting that either Pten is indirectly down-regulated by miR-294, or that the other binding sites in the Pten 3'UTR are required for miR-294 repression **(Figure 16c)**. These data verify that miR-294 enhances iPSC colony formation through inhibition of at least eight genes, seven of which are novel inhibitors of de-differentiation.

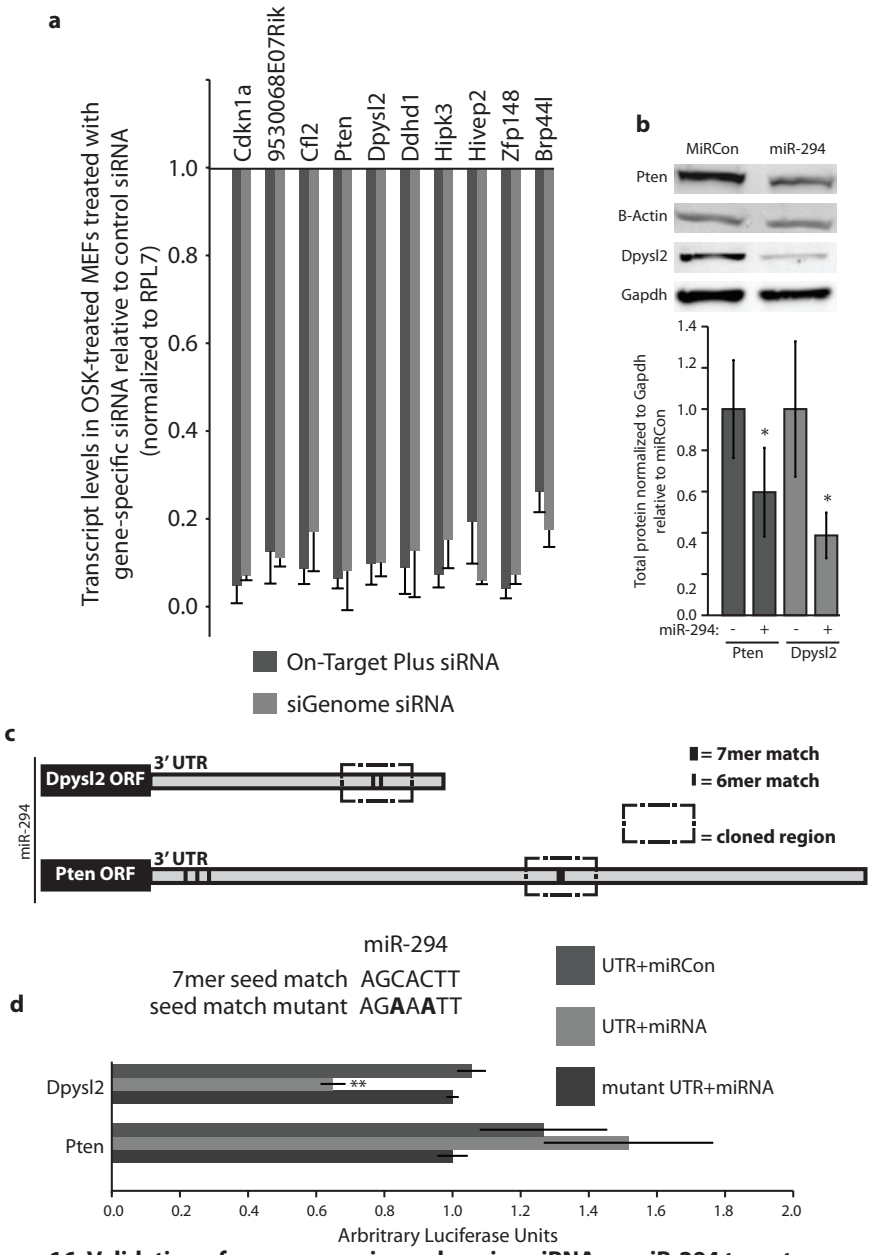


Figure 16: Validation of reprogramming enhancing siRNA as miR-294 targets

a) RT-aPCR measuring levels of each indicated gene 48 hours after addition of corresponding siRNA for two independent pools of siRNA. b) Western Blot representative image and quantification demonstrating miR-294-mediated Pten and Dpysl2 protein knockdown. c) Schematic of Dpysl2 and Pten 3′UTRs with miR-294 binding sites indicated. Boxed regions were cloned adjacent to a luciferase reported. d) Luciferase expression of contructs containing cloned regions from c) with either miR-294, mutant miR-294 or control miRNA transfection. Error bars = SD of n=4. **=p-value<0.005.

Part III: High-content Imaging of Reprogramming Reveals Distinct Mechanisms of miR-294 Function

Reprogramming efficiency can be enhanced either through increasing the percentage of founder cells susceptible to de-differentiation (frequency), or through increasing the kinetics of reprogramming (rate). MiR-294, in addition to increasing Oct4-GFP+ colony number, also induced Oct4-GFP activity earlier during de-differentiation **(Fig 17a)**. Colony area is an independent measure of accelerated rate, as earlier forming colonies are larger at any fixed time point chosen for measurement. Accordingly, miR-294 significantly increased average colony area **(Fig 17b)**. These data demonstrate that miR-294 enhances both the frequency and the rate of reprogramming.

To determine whether siRNAs against targets of miR-294 also enhance both the frequency and rate of reprogramming events, full-well images were captured for each condition, and both colony number and average colony size were measured. Based on these parameters, siRNAs could be separated into three groups: Group 1 (increasing number, but having insignificant effects on area), Group II (increasing area with little effect on number), or Group III (increasing both)**(Fig 17c)**. Similar to the miRNA itself, most miR-294 targets altered colony area, indicating miR-294 accelerates the rate of reprogramming through many targets. Interestingly, some siRNAs (Cfl2) only increased colony frequency, whereas six others (Pfn2, Erap1, Ankrd52, Prkaa1, Lats2, Zbtb41) only increased colony size, and were accordingly

missed by the initial screen. These data indicate that these two parameters of reprogramming efficiency, rate and frequency, can be independently manipulated, and likely represent different physiological processes. Further, they support a model whereby miR-294 enhances both processes at least partially through distinct targets.

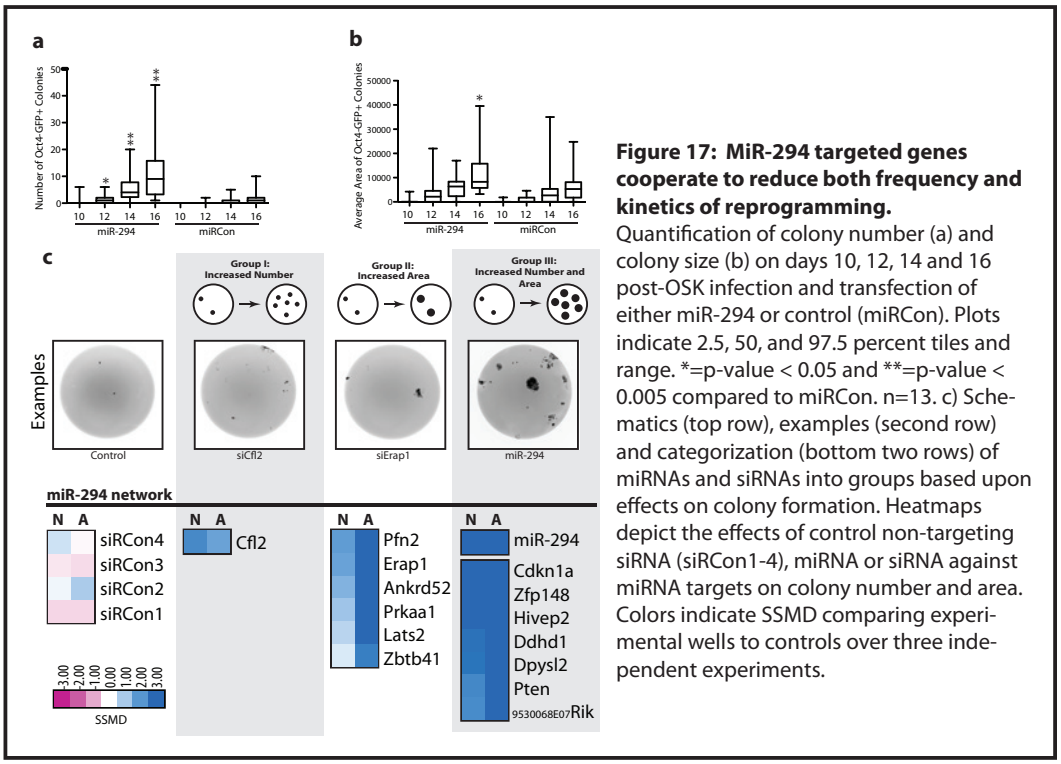


Figure 17: MiR-294 targeted genes cooperate to reduce both frequency and kinetics of reprogramming. Quantification of colony number (a) and colony size (b) on days 10, 12, 14 and 16 post-OSK infection and transfection of either miR-294 or control (miRCon). Plots indicate 2.5, 50, and 97.5 percent tiles and range. *=p-value < 0.05 and **=p-value < 0.005 compared to miRCon. n=13. c) Schematics (top row), examples (second row) and categorization (bottom two rows) of miRNAs and siRNAs into groups based upon effects on colony formation. Heatmaps depict the effects of control non-targeting siRNA (siRCon1-4), miRNA or siRNA against miRNA targets on colony number and area. Colors indicate SSMD comparing experimental wells to controls over three independent experiments.

Part IV: miR-294 Targeted Genes Define Functional Networks of Cooperating Genes

As miR-294 inhibits the expression of multiple genes, which, in turn, regulate different measurable parameters of de-differentiation efficiency, we hypothesized

that the different targets of miR-294 may be regulating distinct cellular processes that converge to enhance de-differentiation. To address this question, we performed a pair-wise screen for cooperative functionality to unbiasedly determine which siRNA against miR-294 targets can work-together to enhance reprogramming. We chose siRNA against the 14 targets of miR-294 that enhanced either colony number or area, and introduced them into the de-differentiation assay in every combination of two. We then compared the number and area of day 16 Oct4-GFP colonies in assays transfected with each set of two siRNAs to assays transfected with each individual siRNA combined with a control siRNA **(Fig 18a)**. Among the 14 miR-294 targets, 20 combinations of siRNAs showed an increase in colony number and/or area with an SSMD score of 2 or higher **(Fig 18b)**. Interestingly, most of these cooperative relationships occurred between genes that influenced different parameters of reprogramming, consistent with distinct cellular pathways between groups (**Fig 18c**). These data suggest that these three sets represent parallel pathways by which miR-294 enhances de-differentiation.

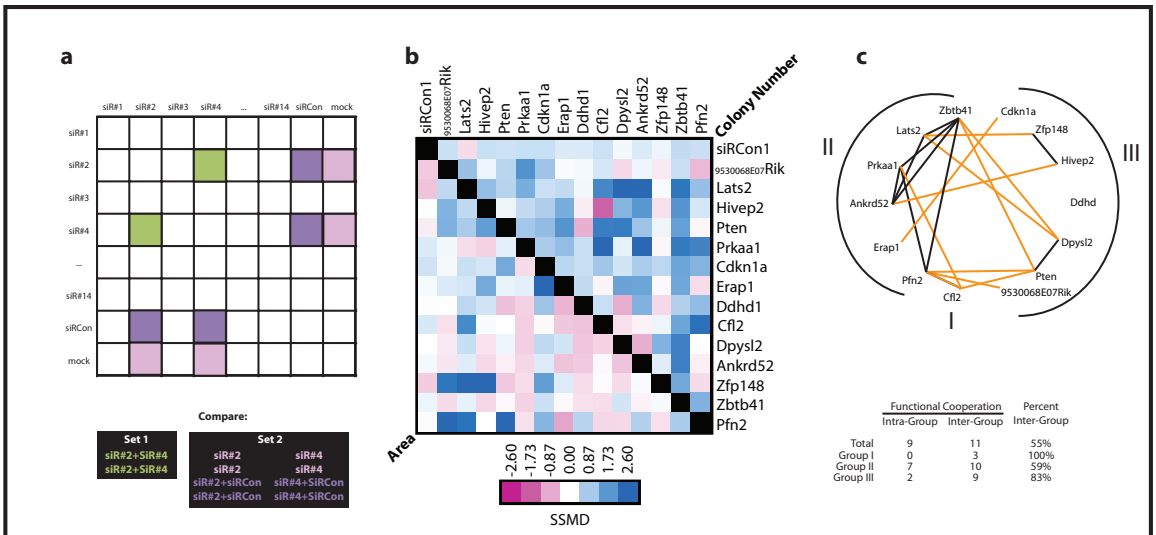


Figure 18: Screen for functional cooperation among siRNA against miR-294 targets
a) Schematic of measurement for functional cooperation of siRNA against miR-294 targets. Matrices of all siRNA were performed in OSK-reprogramming. For any combination of two siRNA (green), day 16 Oct4-GFP colony number or area in wells containing both siRNA, were compared to wells containing only the single siRNAs (pink), or each individual siRNA in combination with control siRNA (siRCon) (purple). b) Heatmaps depicting effects of combining siRNA against targets of miR-294. Colors indicate SSMD comparing combinations of siRNA to each siRNA individually or in combination with siRCon. Top right depicts changes in colony number. Bottom left indicates changes in colony area. c) Depiction of cooperative relationships between miR-294 targets (top). Lines indicate cooperation (SSMD>2) of either area or colony number. Orange lines indicate cooperation occurring across groups. Table depicting percentages of intra-group synergy (bottom).

We next looked for overlap between the miR-294 target sets and genes involved in previously established barriers to de-differentiation, such as induction of apoptosis or senescence, activation of somatic programs, inhibition of proliferation, or remodeling of cell cycle and metabolic pathways. Although the functional targets of miR-294 were involved in every one of these processes, the three sets of genes defined with our epistatic analysis were not over-represented in any single association set, suggesting that these networks are comprised of previously un-associated genes. Together, our results demonstrate that miR-294 enhances de-differentiation through at least three novel functions gene networks, comprised of at least 14 target genes, defining the most comprehensive known miRNA regulated network of functional consequence.

Part V: Summary

Through unbiased and systematic identification of miR-294 targets, followed by functional characterization of knockdown of each target, we have identified fourteen inhibitors of somatic cell de-differentiation. Through multiplex manipulation of these genes and high content analysis of full-well images, we have found that these targets regulate distinct properties of the de-differentiation process, with several subsets working together cooperatively. The genes represent a myriad of functions, establishing novel networks from previously unassociated genes. These data establish a methodology to complement the current repertoire of systems approaches by which miRNAs can be used as a molecular probe, highlighting which genes across known pathways, transcriptional networks and ontology groups should be experimentally tested for functional cooperation.

Chapter 4: The Neural Lineage microRNA-181 Enhances De-differentiation Through Distinct Networks of Genes but Common Pathways

Part I: miR-181 Enhances Reprogramming Through Novel Gene Networks

Given the success of using the de-differentiation enhancing miRNA, miR-294, to discover novel networks of cooperating genes that regulate reprogramming, we wanted to ask if this workflow could be used as a stream-lined methodology for identifying functionally-relevant gene networks in cells state transitions. To address this question, we repeated our methods using different miRNAs that enhance de-differentiation.

To identify individual miRNAs that enhance the production of iPSCs, 570 chemically synthesized mature mouse miRNAs (mimics) were screened for their ability to enhance OSK induced de-differentiation of mouse embryonic fibroblasts (MEFs) to iPSCs. Individual wells of OSK-infected MEFs possessing an *Oct4-GFP* transgene were transfected with mimic on days 1 and 7 post-infection **(Fig 19a)**. The number of day 16 Oct4-GFP+ colonies in each mimic-containing well was compared to sixteen mock-transfected wells per plate using SSMD. When performed in duplicate, 16 miRNA mimics enhanced the frequency of Oct4-GFP+ colony

formation in both screens **(Fig 19b)**. OSK-mimic induced colonies were morphologically similar to mESCs and expressed comparable levels of endogenous Oct4, Sox2, Klf4, Rex1, SSEA1 and NANOG (**Fig 20a-b**). Oct4-GFP+ colonies also efficiently silenced the exogenous retroviruses, indicating an advanced stage of reprogramming into iPSCs **(Fig 20c)**.

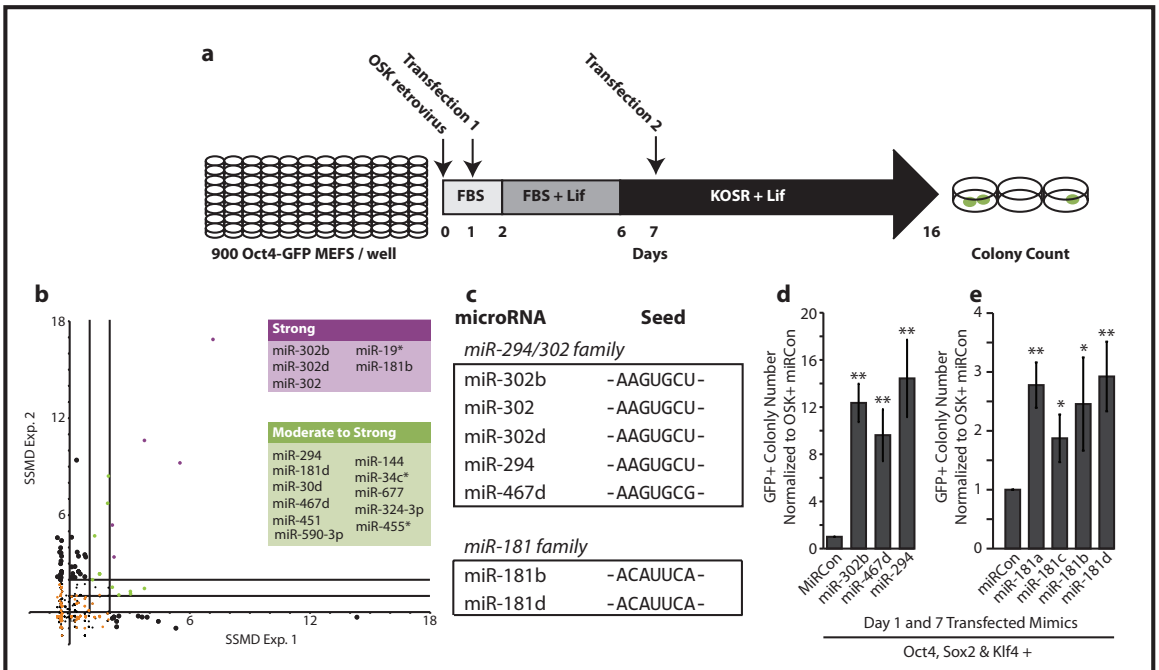


Figure 19: A genome wide screen identifies the miR-294/302 and miR-181 families as enhancers of reprogramming.
a) Schematic of 96 well-based mimic screen for miRNA enhancers of de-differentiation. b) SSMD of two genome-wide screens for miRNA that enhance OSK-driven de-differentiation. Data points represent SSMD between the number of Oct4-GFP+ colonies on day 16 in the presence of an exogenous miRNA mimic compared to 16 mock transfections per plate (shown as orange dots). Significance defined as strong (SSMD>2), moderate (SSMD>1), or weak (SSMD <1). Large dots represent SSMD >2 in at least one experiment with purple being strong in both and green being strong in one and moderate in second experiment (miRNAs corresponding to purple and green dots are shown in inlay). c) miRNA families represented by multiple hits in the screen. d-e) Verification of two families. MicroRNA mimics transfected at days 1 and 7. Data represents number of Oct4-GFP+ colonies from OSK-reprogramming supplemented with indicated microRNA, normalized to OSK + non-targeting miRNA mimic (MirCon). n=3.

Several of the miRNAs that enhanced reprogramming share a common seed sequence **(Fig 19c)**. The most represented seed sequence was that of the ESCC miRNAs, including miR-294 and miR-302 **(Fig 19c-d)**. The second most enriched

seed sequence was that of the miR-181 family, previously unassociated with reprogramming **(Fig 19c)**. Validation experiments to all members of the family confirmed their ability to enhance the frequency of iPSC colony formation (**Fig 19e**). For further experiments, miR-181d was chosen as a representative family member. When injected into E3.5 blastocysts, OSK-miR-181 generated iPSCs contributed significantly to all germ layers, including the germ line, signifying complete reprogramming to an iPSC **(Fig 20d-e)**.

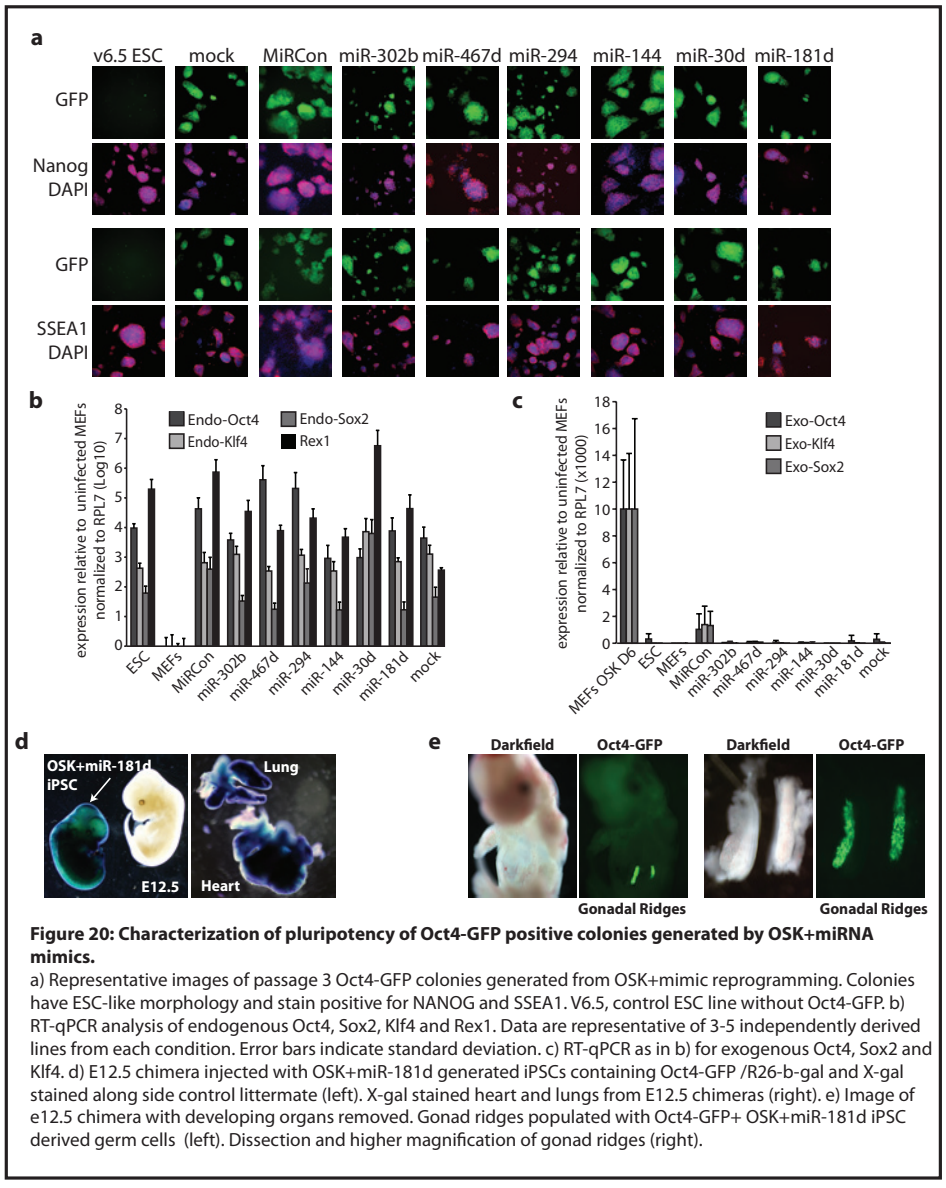


Figure 20: Characterization of pluripotency of Oct4-GFP positive colonies generated by OSK+miRNA mimics.

a) Representative images of passage 3 Oct4-GFP colonies generated from OSK+mimic reprogramming. Colonies have ESC-like morphology and stain positive for NANOG and SSEA1. V6.5, control ESC line without Oct4-GFP. b) RT-qPCR analysis of endogenous Oct4, Sox2, Klf4 and Rex1. Data are representative of 3-5 independently derived lines from each condition. Error bars indicate standard deviation. c) RT-qPCR as in b) for exogenous Oct4, Sox2 and Klf4. d) E12.5 chimera injected with OSK+miR-181d generated iPSCs containing Oct4-GFP /R26-b-gal and X-gal stained along side control littermate (left). X-gal stained heart and lungs from E12.5 chimeras (right). e) Image of e12.5 chimera with developing organs removed. Gonad ridges populated with Oct4-GFP+ OSK+miR-181d iPSC derived germ cells (left). Dissection and higher magnification of gonad ridges (right).

Unlike the ESCC miRNAs, miR-181 is not expressed significantly in ESCs or iPSCs[38]. Furthermore, miR-181 has been associated with both early neural differentiation and destabilization of mESCs[122]. Therefore, miR-181 is unlikely to function through establishing or stabilizing an ESC program. We reasoned that it may enhance de-differentiation through suppression of MEF-stabilizing programs. To test this assumption, we asked whether miR-181 could enhance reprogramming when exclusively introduced early during the transition from a somatic cell to a pluripotent state, prior to down-regulation of MEF genes or activation of the pluripotency network. During iPSC reprogramming, cells undergo an initiation phase followed by maturation/stabilization. This transition is marked by the down-regulation of mesenchymal markers, such as Slug, and the activation of epithelial markers, such as Cdh1 (E-cadherin) and a subset of ESC expressed genes including Dnmt3l[70]. During OSK reprogramming, Cdh1 and Dnmt3l activation and Slug repression were first detectable at day 8 **(Fig 21a)**. To determine the duration of transfected mimic function, a reporter for miR-302 activity was generated **(Fig 21b)**. An immortalized MEF line stably expressing the reporter was generated and transfected with miR-302 mimic. Reporter activity was monitored everyday via flow cytometry **(Fig 21c)**. Introduced mimics were active from one to six days following transfection **(Fig 21d)**. Therefore, a microRNA mimic added at day 1 of reprogramming would function in the initiation phase, whereas a mimic added late would function on a mixed population (**Fig 21e**). Introduction of miR-181 and miR-294 family members only at day 1 largely recapitulated the effect of adding the miRNA mimic at days 1 and 7 **(Fig 21f)**, showing that both miRNA families

efficiently enhance colony formation during the initiation phase. These data show that the miR-181 and miR-294 miRNA families function in significant part to suppress the fibroblast program.

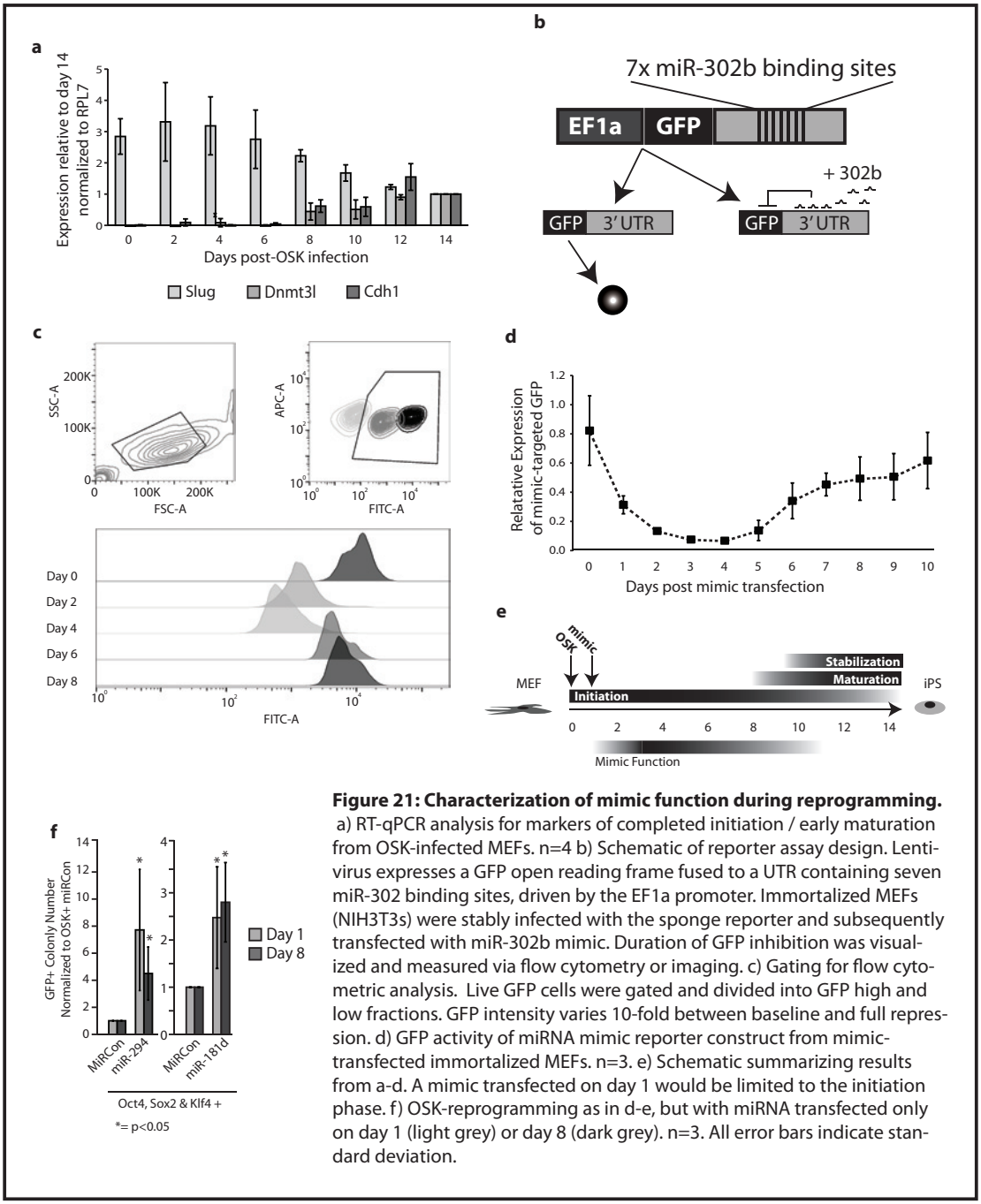


Figure 21: Characterization of mimic function during reprogramming. a) RT-qPCR analysis for markers of completed initiation / early maturation from OSK-infected MEFs. n=4 b) Schematic of reporter assay design. Lentivirus expresses a GFP open reading frame fused to a UTR containing seven miR-302 binding sites, driven by the EF1a promoter. Immortalized MEFs (NIH3T3s) were stably infected with the sponge reporter and subsequently transfected with miR-302b mimic. Duration of GFP inhibition was visualized and measured via flow cytometry or imaging. c) Gating for flow cytometric analysis. Live GFP cells were gated and divided into GFP high and low fractions. GFP intensity varies 10-fold between baseline and full repression. d) GFP activity of miRNA mimic reporter construct from mimic-transfected immortalized MEFs. n=3. e) Schematic summarizing results from a-d. A mimic transfected on day 1 would be limited to the initiation phase. f) OSK-reprogramming as in d-e, but with miRNA transfected only on day 1 (light grey) or day 8 (dark grey). n=3. All error bars indicate standard deviation.

To identify networks of genes that prevent de-differentiation we next

generated a database of high confidence targets miR-181, similar to our analysis of miR-294 targets described in chapter 3. We first consolidated available experimental datasets including miR-181 induced translational inhibition measured via mass spectrometry in the context of HeLa cells, as well as individually verified targets of miR-181 various cellular contexts[114,123–126] **(Fig 22a)**. We next required that each gene contain an appropriate miR-181 binding site that was conserved between mouse and human. Finally, each gene was required to be expressed at some point during MEF de-differentiation, as measured by microarray, resulting in a set of 58 genes, with no genes overlapping with the miR-294 target set[37,111] **(Table 2)**. Similar to our previous verification of miR-294 targets, we transfected MEFs with miR-181 or control mimic 24 hours after OSK-infection, and collected total RNA after 48 hours. We assayed for mRNA destabilization of 13 randomly sampled genes via RT-qPCR. The majority of the genes demonstrated microRNA induced mRNA destabilization, although the down-regulation was not as enriched or pronounced as the miR-294 target list **(Fig 22b)**. This is most likely due to how the targets were identified. By using a MIMD dataset for our miR-294 targets, we likely selected for genes that were highly destabilized by the miRNA on a mRNA level. For miR-181 we began with a dataset measuring translational inhibition instead of transcript degradation.

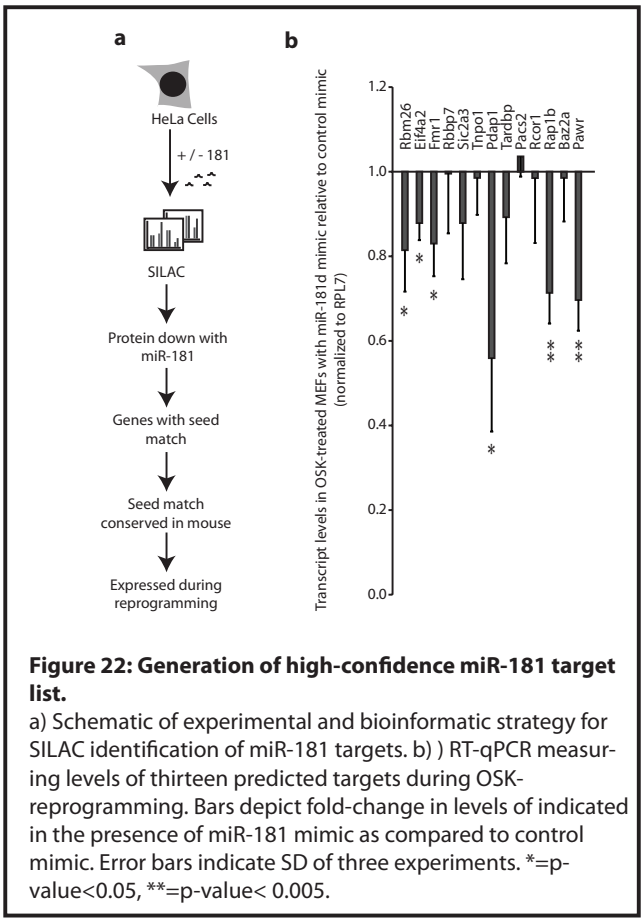


Figure 22: Generation of high-confidence miR-181 target list.
a) Schematic of experimental and bioinformatic strategy for SILAC identification of miR-181 targets. b)) RT-qPCR measuring levels of thirteen predicted targets during OSK-reprogramming. Bars depict fold-change in levels of indicated in the presence of miR-181 mimic as compared to control mimic. Error bars indicate SD of three experiments. *=p-value<0.05, **=p-value< 0.005.

Table 2: High Confidence miR-181Target List

11845	Arf6
11920	Atm
116848	Baz2a
72567	Bclaf1
207165	Bptf
12454	Ccnk
12591	Cdx2
12593	Cdyl
12831	Col5a1
432508	Cpsf6
12912	Creb1
13205	Ddx3x
110052	Dek
235567	Dnajc13
13626	Eed
13682	Eif4a2
14265	Fmr1
17425	Foxk1
72313	Fryl
23881	G3bp2
14465	GATA6
14828	Hspa5
319765	Igf2bp2
140488	Igf2bp3
16201	Ilf3
22343	Lin7c
17118	Marcks
17257	Mecp2
75624	Metap1
105689	Mycbp2

105689	Mycbp2
77579	Myh10
18099	Nlk
70930	Nol8
22026	Nr2c2
217893	Pacs2
114774	Pawr
231887	Pdap1
19046	Ppp1cb
19130	Prox1
80912	Pum1
19317	Qk
215449	Rap1b
245688	Rbbp7
74213	Rbm26
19652	Rbm3
217864	Rcor1
20527	Slc2a3
230908	Tardbp
81004	Tbl1xr1
21676	Tead1
21859	Timp3
69981	Tmem30a
238799	Tnpo1
252838	Tox
22221	Ubp1
56531	Ylpm1
213541	Ythdf2
22628	Ywhag

Similar to the technique described in chapter 3, we next characterized which predicted targets functionally inhibited somatic cell de-differentiation using a siRNA approach. In addition to using siRNA against the 56 miR-181 targets we also generated siRNA against 54 genes randomly selected from the genome. We transfected MEFs 24 hours after OSK-infection, and tallied the number of Oct4-GFP positive colonies on day 16 post-transfection to measure de-differentiation efficiency. The experiment was performed in triplicate using independently generated MEFs and OSK virus, and four different non-targeting siRNA sequences as negative controls. When compared to the negative controls, siRNA against 12 of the 58 miR-181 targets demonstrated fold increases in iPSC formation that achieved our

cut-offs for statistical significance (p-value <0.01 and SSMD >2) **(Fig 23a-b)**. In contrast, only 3 of the random pools of siRNA demonstrated similar effects. To verify siRNA targeting of these genes, the assay was repeated and total RNA was collected 72 hours post-transfection for RT-qPCR analysis. Each siRNA pool significantly reduced expression of the expected target genes **(Fig 23c)**. Genes not expressed in this context were removed from further analysis. To determine if the effects on de-differentiation were due to targeting the expected gene, and not off-target, siRNA with independent sequence were synthesized and verified **(Fig 23c)**. Both sets of siRNA were introduced into the de-differentiation assay. Eight miR-181 targets (Bptf, Lin7c, Cpsf6, Nr2c2, Bclaf1, Nol8, Igfbp2, and Marcks) were verified, whereas only 1 of the random genes remained consistent, demonstrating significant enrichment for siRNA that enhance de-differentiation in sets against experimentally predicted miR-294 and miR-181 targets **(Fig 23d-e)**. Nearly all of the functional miR-181 targeted genes were destabilized in the presence of the miRNA **(Fig 23f)**. We asked whether the remaining three genes were translationally inhibited by these miRNA using luciferase reporter assays and found all of them to exhibit seed-sequence dependent translational repression **(Fig 24)**, confirming our identification of eight novel and functional targets of miR-181. These data demonstrate that using miRNA targeting to refine siRNA screens is an efficient method for identifying genes of functional interest.

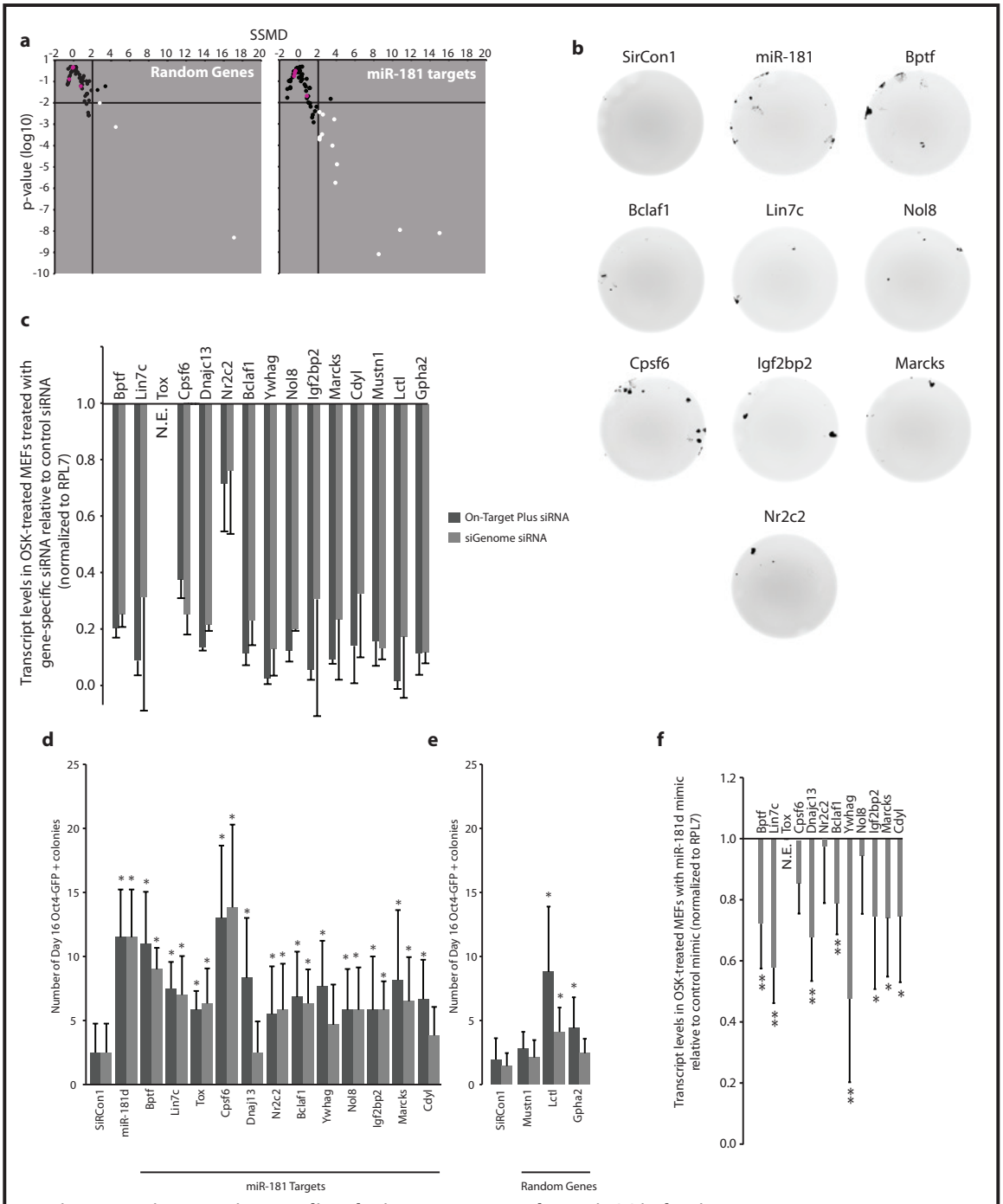


Figure 23: siRNA against predicted miR-181 targets enhance iPSC induction.
Quantificaiton (a), representative images (b), verification (c-e), and MIMD (f) of siRNA against miR-181 targets and random genes during OSK-reprogramming as in Figures 15 & 16.

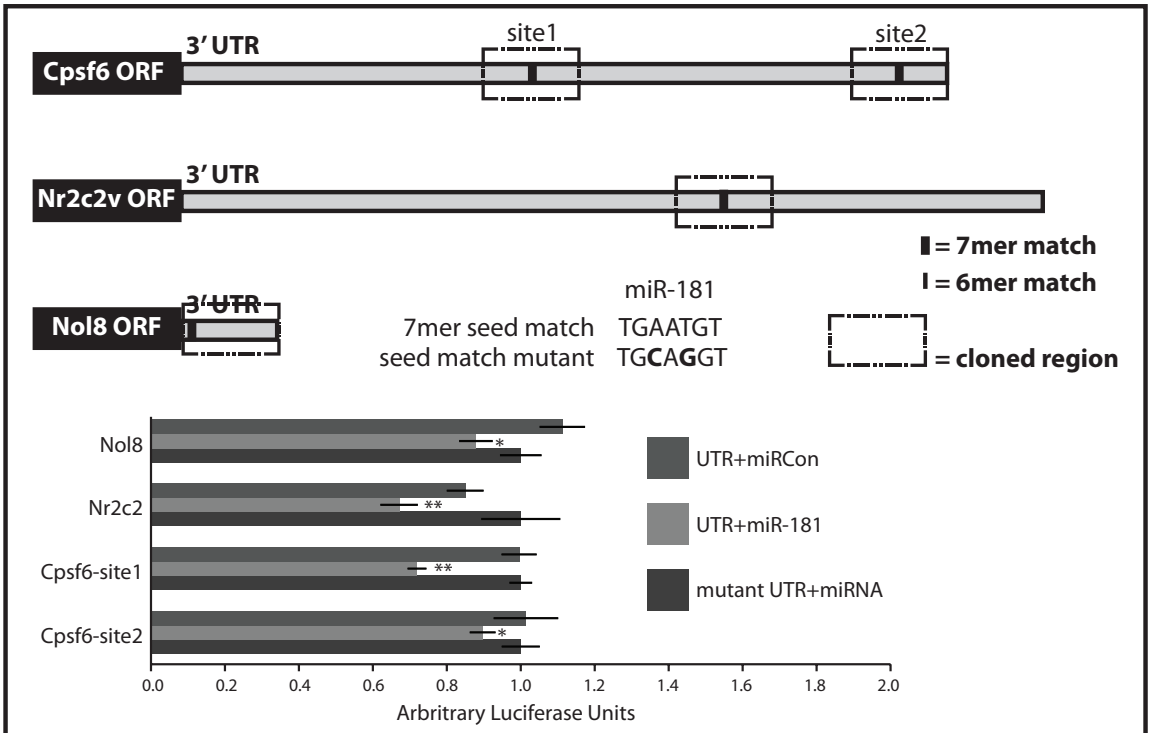


Figure 24: MiR-181 directly targets Cpsf6, Nr2c2, and Nol8.
Top Panel: Schematic of 3′UTRs of candidate miR-181 target genes.
Bottom Panel: Expression of luciferase fused to indicated cloned regions in the presence or miRNA control (miRCon) or miR-181.

Unlike miR-294, miR-181 did not accelerate the rate of colony formation or day 16 colony size, suggesting that this miRNA only increased the frequency of reprogramming events **(Fig 25a)**. Consistently, when colony area was measured, the majority of siRNA against miR-181 targets that enhanced reprogramming only increased colony number (Group 1)**(Fig 25b)**. Synergism through combinatorial knockdown of miR-181 targets was less prevalent, suggesting a higher degree of epistatic relationships **(Fig 25c-d)**. These data support the conclusion that miR-294 and miR-181 enhance de-differentiation through multiple and partially overlapping mechanisms, each regulated by different sets of cooperating target genes.

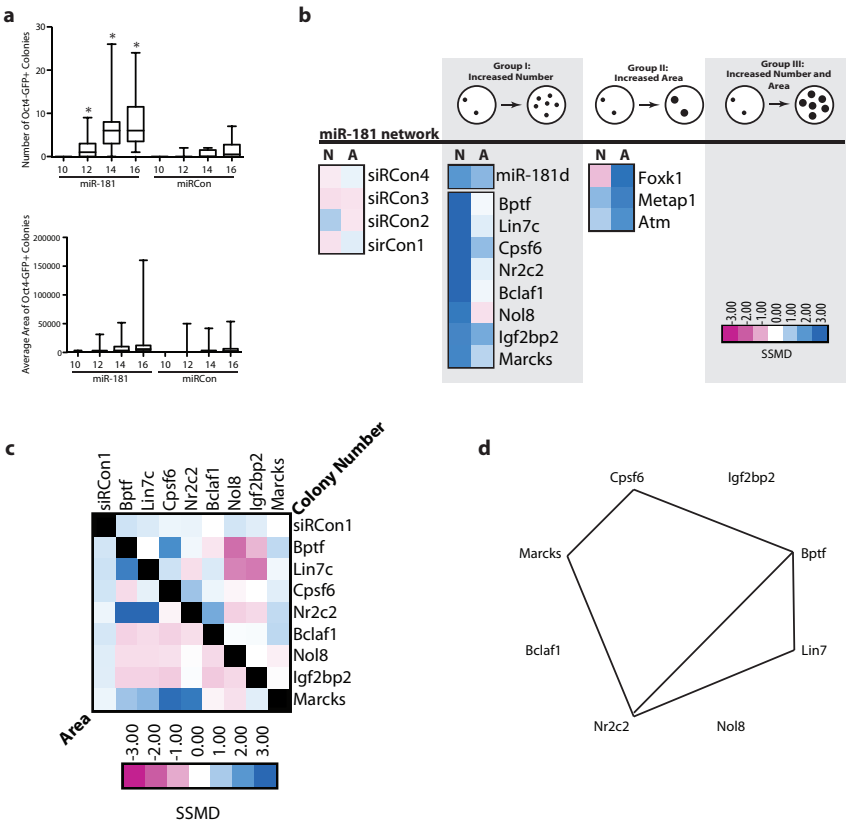


Figure 25: MiR-181 functions through multiple targets to increase reprogramming frequency, not rate. a) Quantification of colony number (top) and colony size (bottom) on days 10, 12, 14 and 16 post-OSK infection and transfection of either miR-181 or control (miRCon). Plots indicate 2.5, 50, and 97.5 percent tiles and range. *=p-value < 0.05 and **=p-value < 0.005 compared to miRCon. n=13. b) Schematics (top row), examples (second row) and categorization (bottom two rows) of miRNAs and siRNAs into groups based upon effects on colony formation. Heatmaps depict the effects of control non-targeting siRNA (siRCon1-4), miRNA or siRNA against miRNA targets on colony number and area. Colors indicate SSMD comparing experimental wells to controls over three independent experiments. c) Heatmaps depicting effects of combining siRNA against targets of miR-181 as in Figure 18. Colors indicate SSMD comparing combinations of siRNA to each siRNA individually or in combination with siRCon. Top right depicts changes in colony number. Bottom left indicates changes in colony area. d) Depiction of cooperative relationships between miR-181 targets. Lines indicate cooperation (SSMD>2) of either area or colony number.

Part II: miR-294 and miR-181 Regulation Converges on Wnt and TGF-Beta Signaling

Given that miR-294 and miR-181 demonstrated partially overlapping functions, but no overlapping targets, we hypothesized that their individual mechanisms might converge on known cellular processes or signaling mechanisms.

As our lists of functional target genes were too small to conduct meaningful enrichment analyses, we turned to high scoring computationally predicted targets (Targetscan, context score <-0.25). Among the top signaling pathways were Cadherin, Wnt, p53 and TGF-beta, each of which have been demonstrated to regulate de-differentiation[52,53,77,127] (**Fig 26a**). In addition, several pathways previously unassociated with de-differentiation were identified including several containing Akt signaling at their core (Insulin, Jak-STAT, Neurotrophin, Apoptosis). To test whether miR-294 or miR-181 influence Akt signaling during early reprogramming, Westerns for phosphorylated T308 and S473 AKT (phospho-AKT) were measured. Forty-eight hours after transfection of miR-294, but not miR-181, the ratio of activated AKT to total AKT was increased **(Fig 26b-c)**. This activity is likely due to miR-294 mediated suppression of PTEN, a known inhibitor of AKT activity. Further, ectopic expression of a tamoxifen-inducible active AKT (*M+Akt:ER*) enhanced colony formation only during the initiation phase, not the maturation phase, of OSK reprogramming compared to controls (*M-Akt:ER*)[128,129] **(Fig 26d**. These data establish Akt signaling as a novel positive regulator of early de-differentiation that is influenced by miR-294.

Next we addressed whether miR-294 and miR-181 functionally regulated Wnt and TGF-beta signaling, both pathways known to influence reprogramming. Both miRNAs activated Wnt signaling during reprogramming as measured by TopFlash activity and localization of B-catenin (**Fig 26e-f**). Similarly, TGF-Beta signaling was regulated by both miRNAs as demonstrated by decreased endogenous phosphorylated-SMAD2 during OSK reprogramming **(Fig 26g)**.

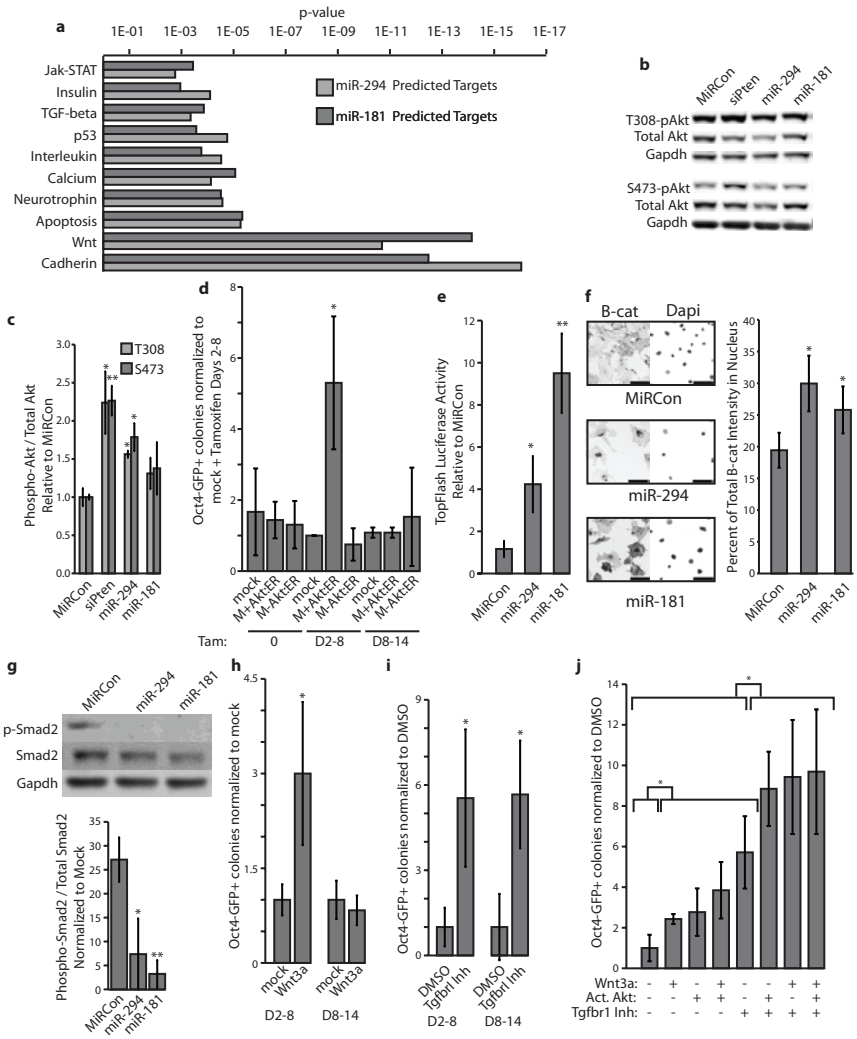


Figure 26: MiR-294 and miR-181 converge on multiple signaling pathways. a) Ten most enriched signaling pathways common to predicted miR-294 and miR-181 targets. b) Representative Western blot detecting total and phoshpo-Akt levels in reprogramming MEFs treated with indicated siRNA and miRNA, serum starved for 24 hours, and treated with IGF. c) Quantification of Westerns represented in b. n=3. d) Day 16 colony count of MEFs infected with OSK and either inducible activated (M+) or inactivated (M-) Akt, treated with tamoxifen during indicated time periods. n=5. e) Relative luciferase units from TopFlash reporter co-transfected into serum starved and Wnt3a-treated reprogramming MEFS with indicated miRNAs. n=3. f) Representative images (left) and quantification (right) of immuno-fluorescent B-cat staining in reprogramming MEFs treated with indicated miRNA as in e. n=3. g) Representative images (top) and quantification (bottom) of Westerns detecting Smad2 and phospho-Smad2 in reprogramming MEFs treated with miRNA after 24 hours serum starvation. n=4. h) Day 16 colony count of MEFs infected with OSK and treated with recombinant Wnt3a (left) or Tgfbrl Inh (right) during indicated time periods. n=4. j) Day 16 colony count of MEFs infected with OSK and treated with indicated combinations of recombinant Wnt3a, Tgfbr1 Inh and M+Akt:ER+Tamoxifen (Act. Akt) on days 2-8. n=4. All error bars indicate standard deviation. *=p<0.05, **=p<0.005.

Previous studies did not address when Wnt signaling had its greatest effects, while conflicting reports exist for TGF-Beta signaling inhibition[52,53]. Therefore, we

tested the effect of recombinant WNT3A and a small molecule TGFBRI inhibitor (Tgfbr Inh) on colony formation when added at either days 2-8 or 8-14. WNT3A enhanced colony formation exclusively during early reprogramming, whereas Tgfbr Inh functioned equally in both stages (**Fig 26h-i**). As each of these pathways functioned early in de-differentiation and are regulated by miR-294 or miR-181, we predict that they may cooperate in a similar fashion to miRNA targets. Combinations of activated *M+Akt:ER*, WNT3A, and Tgfbr Inh were added on days 2-8 of OSK-reprogramming. Activated Wnt and Akt signaling together did not further enhance colony formation, indicating potentially redundant or converging roles of these pathways **(Fig 26j)**. Conversely, TGF-Beta signaling inhibition demonstrated significant cooperation with either activated Wnt or Akt signaling **(Fig 26j)**. These data show that miR-294 and miR-181 converge to enhance reprogramming through both the inhibition of TGF-Beta and activation of Wnt signaling, and miR-294 further enhances reprogramming through activation of Akt signaling.

Part III: miR-294 Inhibits TGF-Beta Induced EMT

One of the initial steps in successful reprogramming is a MET. Interestingly, a subsequent reversal EMT is a barrier to full reprogramming. TGF-Beta signaling has been previously shown to induce EMT in several model epithelial cell lines, including HaCat cells[130]. We therefore asked whether miR-294 could inhibit TGF-Beta induced EMT. HaCat cells were treated with TGF-Beta 24 hours after transfection with miRNA mimics. After 3 days of TGF-Beta treatment, HaCat cells

underwent EMT with dramatic morphological changes accompanied by cell individualization, actin reorganization into stress fibers, and loss of E-cadherin and ZO-1 expression **(Fig 27a-b)**. In contrast, expression of miR-294, but not miR-294 seed mutant, substantially inhibited TGF-Beta-induced EMT. To determine whether this inhibition was due to direct alteration of TGF-Beta signaling, levels of TGFBRI, TGFBRII and phospho-SMAD2/3 were determined via Western analysis. Strikingly, miR-294 greatly reduced TGFBRII protein levels **(Fig27c)**. Consistently, TGF-Beta induced SMAD2 and SMAD3 phosphorylation was significantly abrogated as well. These data suggested that TgfbrII may be an additional direct target of miR-294. Indeed, transcript analysis revealed three ESCC binding sites in the 3'UTR of TgfbrII **(Figure 27d)**. The ESCC miRNAs miR-302b and the human miR-294 ortholog, miR-372, inhibited translation of luciferase fused to this UTR, which was rescued by mutation of the binding sites. Together, these data suggest that one mechanism by which miR-294 inhibited TGF-B signaling enhances reprogramming is through inhibition of Tgf-B-induced EMT, and that this inhibition is, in part, through direct targeting of TgfbrII.

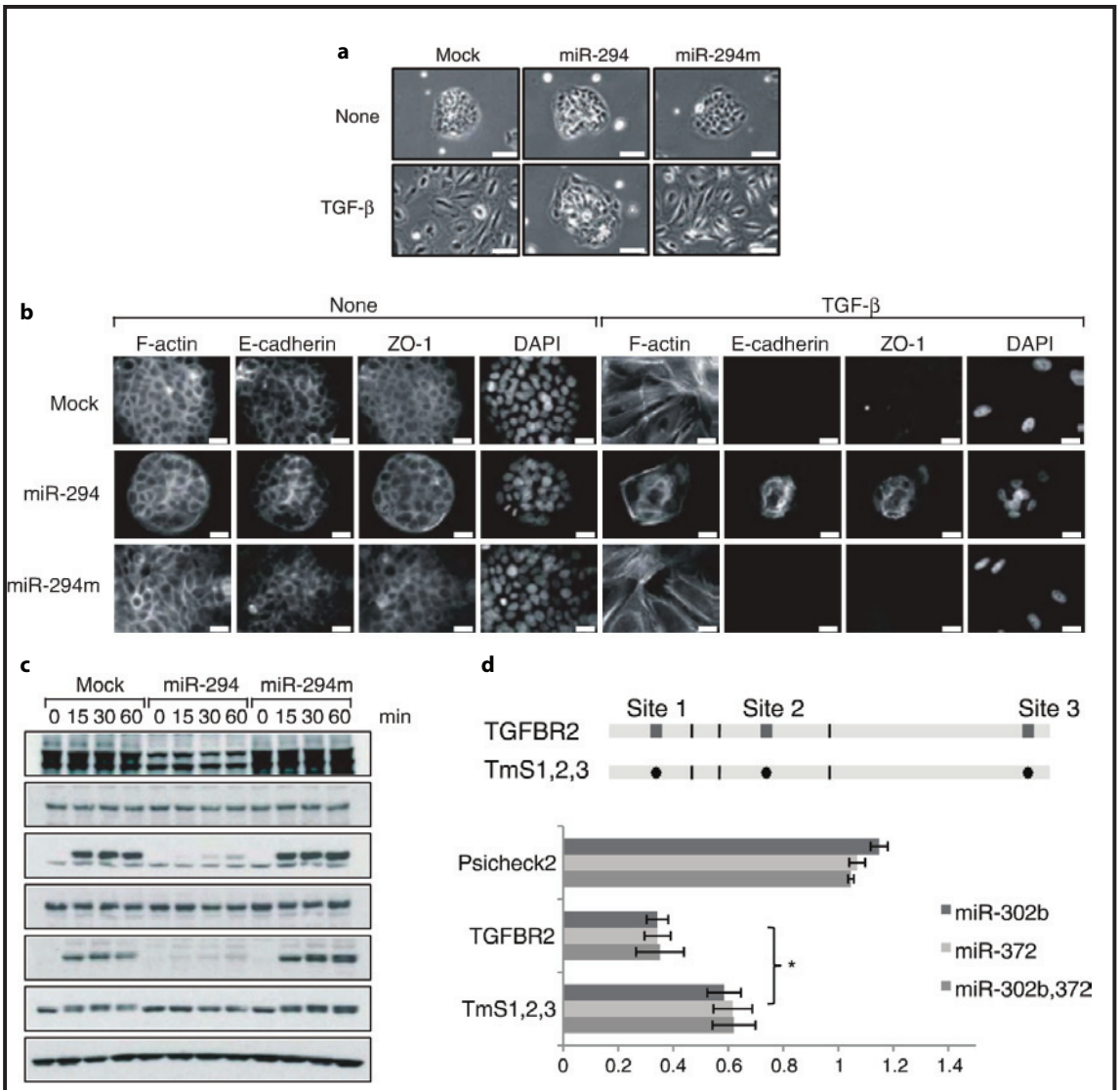


Figure 27: miR-294 inhibits TGF-Beta-induced epithelial to mesenchymal transition through direct targeting of TGFBRII.

HaCaT cells were transfected with indicated miRNAs, then treated or not with TGF-Beta for 73 h and observed by phase contrast microscopy (a), or fixed and subjected to immunostaining for F-actin, E-cadherin and ZO-1 (b). c) Western blot showing levels of TFG-Beta receptors, phospho-Smad2 and phospho-Smad3 in HaCaT cells 0-60min after TGF-Beta exposure in the presence of miRNA mimics. Cells were transfected with the indicated miRNAs 48 h before TGF-Beta treatment. d) Luciferase analysis of TGFBRII 3'UTR. Seed matches for ESCC miRNAs in the 3'UTRs along with mutant construct shown in top panel. Luciferase results after co-transfection with ESCC miRNAs relative to mock transfection are shown in the lower panel after normalization to firefly luciferase values. Error bars = SD. *=p-value <0.05.

Part IV: Summary

Our data produce the two most extensive experimentally-verified miRNA-mRNA networks yet reported, solidifying the postulate that miRNAs function though co-regulation of many genes and pathways. In addition, this work supports the use of miRNAs and their mRNA target interactions to investigate the mechanisms behind developmental and cellular phenomena. Through unbiased screening of miRNAs as well as identification and screening of individual mRNA targets, we identified two miRNA families that target a combined twenty-five genes and three signaling pathways that cooperatively function as barriers to early de-differentiation (**Fig 28**). These data uncover diverse cellular pathways that can cooperate to regulate cell state transitions, and provide significant insight into how miRNAs coordinately regulate the gene networks that make up these pathways.

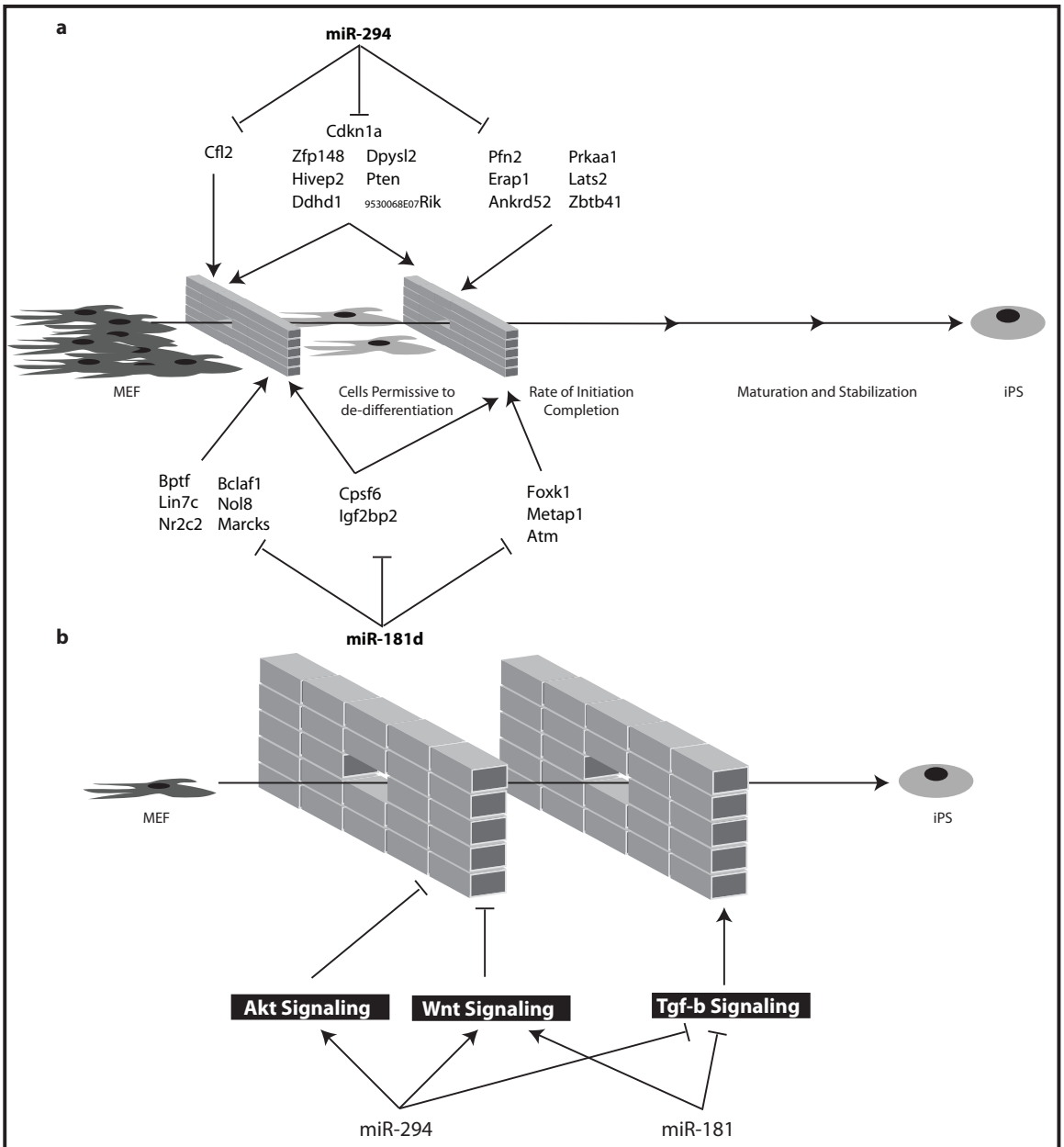


Figure 28: Schematic summary of mechanistic insight into reprogramming provided by these studies. a) miR-294 and miR-181 target two distinct networks of genes acting as barriers to reprogramming. Genes represent two physiological barriers - 1) inhibition of the number of cells permissive to reprogramming and 2) inhibition of the rate of reprogramming. b) miR-294 and miR-181 converge on two pathways regulating distinct barriers to reprogramming.

Chapter 5: Discussion and Future Directions

MicroRNAs Function Through Many Targets

Recently developed profiling techniques have confirmed that individual miRNAs can have hundreds of target transcripts[113–116]. Accordingly, it has been commonly postulated that miRNAs likely work through regulation of large networks of genes. However, experimental data for functional relevance of this magnitude of co-regulation have not been reported. Indeed, hundreds of previous studies have attributed miRNAs' mechanisms of action to one to three targets[131]. Frequently, if siRNA knock-down of a single miRNA target recapitulates the biological effect of over-expression of that miRNA, that gene is called the "dominant target". Here, we have demonstrated that siRNA knockdown of many miRNA targets can recapitulate the effect of the miRNA itself, suggesting that this approach cannot be used to accurately identify the dominant target. Instead, these results solidify the postulate that miRNAs function though co-regulation of many genes. Our workflow also serves as a proof of principle for the utility of using miRNA target analysis to investigate the mechanisms behind cell state transitions.

MicroRNAs Inhibit Distinct Barriers of Reprogramming

Through unbiased screening of miRNA, identification of targets and screening of individual targets, we identified two miRNA families that target a combined twenty-five genes expressed in early reprogramming that prevent de-differentiation. Combining these networks with previously reported functions of the target genes, we find that miR-294 and miR-181 target genes involved in

promoting several previously reported barriers to differentiation including activation of apoptosis / senescence / cell cycle arrest (Cdkn1a, Pten, Atm, Zfp148, Hivep2, Bclaf1, Lats2), EMT (TgfbRII), and the AMPK regulated metabolic switch (Prkaa1)[68,70,72]. In addition, the functional targets were enriched for regulators of cytoskeleton dynamics / endocytosis / vesicular transport (Cfl2, Dpysl2, Pfn2, Pten, Prkaa1, Lin7c, Atm). Although these processes have no known role in somatic cell de-differentiation, these data suggest that a distinct barrier to de-differentiation may be the restructuring of the cytoskeleton. Further, at least three of the target genes, Bclaf1, Dpysl2 and Erap1 promote early differentiation of somatic tissue[132–134]. In fact, ESCs that express dominant negative Erap1 or are deficient in Bclaf1 fail to differentiate at all. This suggests that part of successful iPSC generation requires the silencing of somatic programs and that miRNAs contribute to this process. Finally, many of the functional miR-294 and miR-181 targets are poorly characterized genes with little known function. It will be interesting to see how these particular genes are inhibiting de-differentiation, especially considering that their own cooperative relationships imply they function through independent mechanisms.

MicroRNAs are De-stabilizers of Cell State

These studies demonstrate that the primary mechanism of miR-294 and miR-181 enhancement is the down-regulation of MEF stabilizing genes. Indeed, given the antagonistic relationship of miR-181 with mESCs, it is very likely that miR-181, in particular, is functioning through inhibition of fibroblast programs. We also identified the let-7 family as MEF-expressed miRNAs that inhibit the transition away

from the somatic cell state. Consistent with miR-294 and miR-181 destabilizing the fibroblast program, Let-7 targets many of the down-stream effectors of the pluripotency network, as well as a few reprogramming factors themselves, such as Myc and Lin28[97]. Indeed, Let-7 functionally destabilizes the pluripotency network and mESCs. These data are consistent with the model that miRNAs function through destabilization of cells types that are temporally or spatially similar in development **(Fig 29)**. Were this true, we would hypothesize that miR-294 and miR-181 could prime MEFs for any induced cell state transition, and, conversely, that miR-181 could not prime cell states in the neuronal lineage for transitions. Experimental approaches to answering these questions would make very interesting follow up studies.

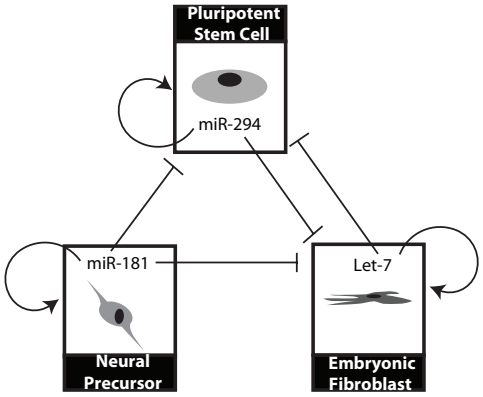


Figure 29: Model depicting miRNAs destabilizing similar cell state. Representative dominant miRNAs in each cell type are shown. Arrows and inhibition bars summarize data from this study and previous publications. The emerging model suggests that dominant miRNAs buffer against transitions to other cell states by destabilizing the programs of related cell types, thereby indirectly stabilizing the cell type endogenously expressing the miRNA.

Of particular note, consistent with our Let-7 data, this model predicts that inhibition of endogenous miRNAs would enhance cell state transitions. Were this true, it follows that over-expression of any non-toxic small RNA would enhance de-

differentiation to some degree, due to over-saturation of RISC, effectively shutting down endogenous miRNA activity. Indeed, during our screens, we noticed that each siRNA set, including the randomly selected control set, contained 20-30% siRNA which caused subtle increases in Oct4-GFP colony number. We predict this low level of enhancement is likely due to indirect Let-7 inhibition. Researchers should be aware of this potential side-effect of small RNA over-expression when studying cells state transitions, and plan appropriate controls.

Conclusion

Together, our data support the evolutionary postulate that miRNAs play a critical role in establishing and maintaining the myriad of highly specific and functional cell states found in complex organisms. MiRNA are an efficient, specific, and non-integrating tool for inducing cell state transitions, and their targeting mechanism makes down-stream analysis of functional target genes an efficient process. We look forward to the results of similar workflows applied to other cell state transitions, and their potential impact on both clinically translatable and basic developmental research.

Materials & Methods

Cell Culture

MEF Generation

MEF generation was conducted as previously described[39]. In brief, either *rosa26-Bgal;Oct4-GFP* or *Oct4-GFP* embryos were harvested on E13.5. Heads and visceral tissue were removed. Remaining tissue was disassociated with trypsin and physical disruption and plated (P0) in MEF media (high glucose (H-21) DMEM, 10%FBS**,** non-essential amino acids, L-glutamine, Penn/Strep, 55uM beta-mercaptoethanol). MEFs were expanded to P3 and frozen.

Virus Production

HEK293T cells grown to approximately 70% confluence were transfected with pCL-Eco and pMXs- or pWZL-expression plasmids at a ratio of 1:2 following the Fugene 6 manufacture's protocol. At 24 hours, media was replaced with fresh MEF media. At 48 hours, supernatant was harvested, filtered (0.45uM) and frozen at -80 degrees. Virus preparations were only thawed once before use.

Lentivirus: HEK293T cells grown to approximately 70% confluence were transfected with pMDL, pRSV, pVSVG and pSIN-expression plasmids at a ratio of 1:1:1:2 following the Fugene 6 manufacture's protocol. Cells were left for 48 hours, then harvested as above.

De-differentiation

Oct4-GFP MEFs (P5) were plated onto gelatin coated Whatman Clear View or Greiner uClear black-walled 96-well imaging plates at 900 cells / well. The next day, 50ul of each retrovirus-containing supernatant with 4ug/mL polybrene was added. Day 1 post infection, virus was replaced with fresh MEF media. Thereafter, media was changed every other day, with ES+FBS media (15%FBS**,** non-essential amino acids, L-glutamine, Penn/Strep, 55uM beta-mercaptoethanol and Lif) days 2 to 6 post-infection and ES+KSR media [Knock-out DMEM (Invitrogen), 15% Knock-out Serum Replacement (Invitrogen), non-essential amino acids, L-glutamine, Pen/Strep, 55uM beta-mercaptoethanol and Lif] days 6 to 16 post-infection. Supplements were added at indicated final concentrations: Tamoxifen (Sigma, 1nM), recombinant Wnt3a (R&D Biosystems, 50ng/mL). E-616452 (BioVision, TgfbR inhibitor, "RepSox", 1uM). Oct4-GFP expression and colony formation was assessed on days indicated, usually day 16 post-infection. High throughput imaging and high content analysis were conducted with the InCell Analyzer 2000 imaging station and software suit (GE). Independent experiments are defined as independent MEF lots infected with independent virus preparations. To validate pluripotency, day 16 iPSC colonies were disassociated with trypsin and plated onto irradiated MEF feeder layers (P1) and expanded. Passage 3 colonies were harvested for RT-qPCR and fixed for immunohistochemistry. Passage 5 colonies were injected into blastocysts.

Teratoma formation

iPS lines were grown on irradiated MEFs or gelatin, trypsinized, and resuspended in PBS. One million iPS cells were injected subcutaneously per side in severe combined immunodeficient (SCID) mice (NCI-Frederick). Tumors were removed when they reached a size of 1-1.5cm in long diameter, fixed in 10% formalin, embedded in paraffin, sectioned, and H&E stained.

Karyotyping

iPS cells were karyotyped as previously described[39].

Blastocyst Injection

Blastocyst injections to assay for chimeric contribution were performed as previously described[39,98]. Blastocysts were obtained from E2.5 super-ovulated and fertilized C57BL/6 females (Taconic). Blastocysts were washed in M2 media (Specialty Media) and grown in KSOM media (Specialty Media) for 16h. 16h after blastocyst collection, 10-15 iPS cells were injected into cultured blastocysts, which were then transplanted into the uteri of E2.5 pseudo-pregnant Swiss-Webster females (Taconic). For analysis of tissue contribution, embryos were collected on E13, and stained for B-gal activity. For analysis of germ line contribution, embryos were collected on E13 and gonads were isolated and imaged under fluorescence. 80% of implanted blastocysts demonstrated high-grade chimeric contribution of iPS lines.

Small RNA Transfections

MicroRNA mimics (MIRIDIAN) and siRNA pools (On-TargetPlus and siGenome) were both generous gifts from Dharmacon. Transfections followed the Dharmafect manufacturer's protocol. DMEM containing 1uM RNA and DMEM containing 6:1000 (v/v) Dharmafect 1 were pre-incubated at room temperture for 5min, then mixed 1:1. After 20min of room temperature incubation, transfection mixture was added to fresh media on cells for a final RNA concentration of 100nM. MTT assays to monitor cell viability post-transfection were performed as previously described[108].

Mir-290 promoter analysis

Previously published ChIP-seq data for c-Myc, n-Myc[135], H3K4me3, and H3K27me3[110] were downloaded as fastq files and aligned to the mm9 (NCBI Build 37.1) assembly of the mouse genome using Eland (GA Pipeline 1.0, Illumina). The mm9 assembly contains the mir-290 locus[38], which was missing from previous assemblies. Following alignment, peak scores were assigned using the Findpeaks 3.1.9.2 algorithm[136]. The peak scores were normalized to the number of genome-mapping sequence reads.

RT-qPCR

Total RNA was collected using either Trizol (manufacture's protocol) or RNeasy spin columns (Qiagen, manufacture's protocol). For mRNA amplification,

RNA (1-5ug) was treated with DNAse I (Invitrogen) and reverse transcribed using the Superscriptase III kit (Invitrogen, manufacture's protocol) with polyT primers. For miRNAs, qRT-PCR was performed either by using TaqMan® miRNA assays (Applied Biosystems) or by polyadenylating the miRNAs and then using a modified oligodT reverse transcription primer as described previously[137] or using TaqMan® miRNA assays (Applied Biosystems). Total cDNA was diluted 1:5 and 1uL per reaction was amplified using gene specific primer sets (500nM**)** and Power SYBR Green PCR master mix (ABI). Endogenous and exogenous Oct4, Sox2, and Klf4 primers were previously described[98]. New primer sets are listed in **(Table 3)**. Specificity of all primer sets was verified through analysis of disassociation curves in experimental, no RT, and water only samples.

Table 3: Primers used for qPCR amplification

Gene	Forward Primer	Reverse Primer
9530068E07Rik	CTGTGCAGCTGTTGTGTATG	ACGCCACTTTCTGCTTTG
Bclaf1	GAGATGGGATTGTTGAAGATG	CAAGTTCTGCTCCCTGTTG
Bptf	AGTGTGCAGAAGTTCTTGAATC	TTGCTGTGTCTAGCTTAGGTG
Brp44l	GCTATCAATGACATGAAGAAATC	TTGTACCTTGTAGGCAAATCTC
Cdh1	AACAACTGCATGAAGGCGGGAATC	CCTGTGCAGCTGGCTCAAATCAAA
cdkn1a	CTGTCTTGCACTCTGGTGTC	TTCTCTTGCAGAAGACCAATC
Cdyl	GGAACCTTCACACAGGAAGTC	TCAGCACTTCACATTCTCTCTC
Cfl2	TCTGGGCTCCTGAAAGTG	TTAATATCGTCCAAGCCATTTAC
Cpsf6	TGACCGAGAGCGAGAATAC	TCAATCACAAGAAGCAAACTG
Ddhd1	CCTTCAGCTTCACCCTCTAC	ATAGCGGCTCTCCACAAG
Dnajc13	ATCCTCTCTCGGTCTTCAGTC	ATGGCTGATGAGGATGTACC
Dnmt3L	GGCCCTTCTTCTGGATATTC	CGCATGCTTGCTCTTCAGCC
Dpysl2	AGCAGGCACCACCTGTTC	CCCAGGCTGGTGATGTTG
Hipk3	CAGCAGCGTTCCCTCAGC	CGATGCCCAAAGTTTCCATTC
Hivep2	CCAGAAGGGCTTTATATCCTAAC	CATAAGCACCTTCTTGGTCTC
Igf2bp2	CTACGCCTTCGTGGACTAC	CTGGATTCTTCTGCTCCTTAG
Lin7c	CCTCAAACGAGGAGATCAG	CAGTAGCTCTACCGCTTTCTC
Marcks	GTGCCCAGTTCTCCAAGAC	GTTGGCTTGCAGCTCCTC
Nol8	TGGAGTGGTATGGGAGGTAG	GACAATGGTTAATTTGCTTTCAC
Nr2c2	GACTTAACTGCTTTGGCTCAG	TGCTGGGACTTCTTTGCTAC
Pten	ACAAAGCAAACAAAGACAAGG	GATTTGATGGCTCCTCTACTG
Rex1	GAAAGTGAGATTAGCCCCGAG	GTCCCCTTTGTCATGTACTCC
Slug	CACATTCGAACCCACACATTGCCT	TGTGCCCTCAGGTTTGATCTGTCT
Tox	TTTCGAACGCAATCACTATC	CATTTCCTGTATTTGGCTCTC
Ywhag	AGCCCTGTGAAGATGGTG	CCGTTCCTCATTGGACAG
Zfp148	AAAGAACTCACAGTGGAGAGAAG	TTTGATGGCACATCTGTTTAG
Lin28	AGTCTGCCAAGGGTCTGGAA	CGCTCACTCCCAATACAGAACA
cMyc	CAGAGGAGGAACGAGCTGAAGCGC	TTATGCACCAGAGTTTCGAAGCTGTTCG
Rpl7	GATTGTGGAGCCATACATTGCA	TGCCGTAGCCTCGCTTGT
U6	CGCTTCGGCAGCACATATA	TTCACGAATTTGCGTGTCAT
microRNAs	MATURE SEQUENCE	GCGAGCACAGAATTAATACGACT

Immunohistochemistry

Cells were fixed for 15 minutes in 4% PFA, washed in PBT (PBS + 0.1% Triton x-100), incubated for one hour at room tempurature with blocking buffer (PBT+1% goat serum+2% BSA), then incubated overnight at 4 degrees in primary antibody in blocking buffer as follows: Nanog 1:50 (Abcam ab21603), SSEA1 1:100 (Univ of Iowa MC-480), Ecad 1:120 (BD Transduction Laboratories 610181), beta-Catenin 1:100 (Cell Signaling 9587) and, JAM-1 (Santa Cruz Biotechnology, sc-25629). For Nanog antibody, cells were also fixed with methanol at -20 degrees C for 5 min, prior to block. Cells were then washed in PBT, incubated for one hour at room temperature in secondary antibody in blocking buffer (Alexa Fluor 1:1000 Invitrogen), and in some cases, rhodamine-conjugated phalloidin (Invitrogen), washed in PBT with Hoechst 33342 1:10000 (Invitrogen), and stored in PBS before imaging. Slides were mounted using ProLong Gold antifade reagent with DAPI (Invitrogen).

HaCaT cell culture and EMT

HaCaT cells were cultured in DMEM with glucose (4.5 g/l) and 10% FBS. Cells were plated at 100,000 cells per 6-well well and transfected the next day with miRNA mimics (ThermoFisher) using Dharmafect 1 (ThermoFisher) according to manufacturer's protocol. Mimics were transfected at a final concentration of 40nM. On day 2 post-transfection, cells were analyzed as follows. For signaling pathway protein quantification, cells were serum starved with 0.5% FBS overnight, treated with 2 ng/ml of TGF-b1 (HumanZyme) for indicated times and lysed. For RNA

quantification, cells were split, then treated with TGF- b1 for 24 hours and lysed. For morphological and immunocytochemical analysis of EMT, cells were split into chamber slides, and treated with TGF- b1 for 72 hours before fixing and imaging. To view the cell morphology or to monitor TGF- b-induced EMT, cells were observed using a Leica DMI 4000B microscope, and bright field pictures were taken using a Leica DFC 350FX camera. Images were analyzed using the Leica Application Suite and Photoshop (Adobe) software. The TbRI kinase inhibitor SB431542 (Sigma-Aldrich) was used at 5 mM in conditions without TGF- b treatment to inhibit secreted autocrine TGF- b.

Statistical Analysis

For small scale experiments performed in three or more independent experiments p-values were calculated using a student's t-Test.

For large-scale siRNA screens, strictly standardized mean difference (SSMD) was calculated to compare single experimental wells to either i) sets of four matched scrambled siRNA transfected wells (Fig 2a and Fig 3b), ii) sets of sixteen matched mock transfection wells (Fig 1b) or iii) sets of individual siRNA (Fig 3c and Supp Fig 9) as outlined previously[121].

Generation of High Confidence Target Lists

Lists of genes significantly down-regulated by either miR-294 or miR-181 were obtained from previous publications. Specifically, for miR-294, microarrays

were used to measure mRNA down-regulation upon addition of miR-294 to *DGCR8-/-* mESCs[97]. For miR-181, SILAC analysis was used to measure protein down-regulation upon addition of miR-181 to HeLa cells[114]. In both cases, authors' cut-offs for significant down regulation were used. To these lists, known miR-294 family or miR-181 family targets were added[120,123–126]. Genes were then required to have miR-294 or miR-181 binding sites in mouse, and to be expressed during the course of MEF to iPSC reprogramming[73].

Luciferase Assays

All experiments were performed using the Dual-Luciferase Reporter Assay System (Promega) on a dual-injecting SpectraMax L (Molecular Devices) luminometer according to the manufacturer's protocol. Ratios of Renilla luciferase readings to firefly luciferase readings were averaged for each experiment. Replicates performed on separate days were mean centered with the readings from the individual days.

B-catenin reporter assay: Topflash reporter plasmid was obtained from Addgene (plasmid 12456)[138]. Mouse embryonic fibroblasts were cultured in Oct4, Sox2 and Klf4 reprogramming conditions described above. 24h post retroviral infection, cells were transfected with miRIDIAN miRNA mimics (Dharmacon) using Dharmafect1 (Dharmacon) as described above. 72h post retroviral infection, cells were transfected with TOPFlash reporter plasmid (final concentration 1ng/μl) and TK-renilla transfection control plasmid (Promega) (final concentration 0.33ng/μl)

using Promega Fugene6 transfection reagent according to manufacturer's protocol. Recombinant murine Wnt3a (R&D biosystems) was added to the transfection mix at a final concentration of 25ng/ml in ESC media. The cells were lysed 24h after TOPFlash transfection/Wnt3a stimulation, and the luciferase assay was performed.

Target verification reporter assay: 3′UTRs of indicated genes were amplified from the mouse genomic DNA cells using the Zero Blunt TOPO (Invitrogen) vector and subcloned into psiCHECK™-2 vector (Promega) using the Cold Fusion Cloning Kit (System Biosciences). 3′UTR seed sequences were mutated using the Quickchange Lightning kit (Agilent). For transfection, 8,000 miRNA-deficient Dgcr8−/− mouse ESCs were plated in ESC media onto a 96-well plate pretreated with 0.2% gelatin. The subsequent day, the cells were transfected with miRIDIAN miRNA mimics (Dharmacon) using Dharmafect1 (Dharmacon) at the manufacturer's recommended concentration of 100 nM. Simultaneously, 200 ng of the psiCHECK-2 construct was transfected into the ESCs using Fugene6 (Roche) transfection reagent according to the manufacturer's protocol. Transfection of each construct was performed in triplicate in each assay. The cells were lysed 24h after transfection, and the luciferase assay was performed.

Western Blot Analysis

MEFs were cultured in Oct4 Sox2 Klf4 reprogramming conditions as described above. 24h post retroviral infection, cells were transfected with miRIDIAN miRNA mimics (Dharmacon) with Dharmafect1 (Dharmacon) as described above.

72h post infection, cells were either serum starved (high glucose (H-21) DMEM, 0.5% FBS**,** non-essential amino acids, L-glutamine, Penn/Strep, 55uM beta-mercaptoethanol) or media was changed to regular ESC media. For some assays, 16hrs after serum starvation / media change, serum starved cells were stimulated with IGF1 protein (Abcam) for five minutes at a concentration of 6nM in serum starvation media. Lysates were collected in lysis buffer (25 mM Tris-HCl, pH 7.9, 150 mM NaCl, 0.1% Nonidet P-40, 0.1 mM EDTA, 10% Glycerol, 1mM DTT) containing 1× protease inhibitor cocktail (Roche) and 1xPhosSTOP Phosphatase Inhibitor Cocktail (Roche). Lysates were incubated at 4 °C for 10 min rocking then collected by scraping. After three snap freeze-thaw cycles, lysate was spun at 4 °C and approximately 20,000*g* in a table-top centrifuge. Protein was quantified using a Bio-Rad protein assay (Bio-Rad). Five micrograms of protein was resolved on a 10% SDS–PAGE gel. Proteins were transferred to Immobilon-FL (Millipore) and processed for immunodetection. Blots were scanned on a Licor Odyssey Scanner (Licor). Antibodies were diluted as follows: GAPDH 1:5,000 (Santa Cruz, sc-25778), Phospho-Akt (Ser473) 1:2000 (Cell Signaling, #4060), Phospho-Akt (Thr308) 1:1000 (Cell Signaling, #2965), Akt (pan) 1:1000 (Cell Signaling, #2920), PTEN 1:2000 (Cell Signaling, #9552), Dpysl2/Crmp2 1:1000 (Cell Signaling, #9393) Phospho-Smad2 (Ser465/467) 1:1000 (Cell Signaling, #3108), Smad2 1:1000 (Cell Signaling, #3103). Secondary infrared-dye antibodies from Licor were used at 1:25,000. Images were quantified using Odyssey Software.

MicroRNA mimic stability assays

The miR-302 sponge consists of complementary sequences to mature miR-302b miRNA with mismatches corresponding to basepairs 9-12 of the mature miRNA. miR-302b sponge sequence corresponding to basepairs 9-11 of the mature miRNA sequence were designed to be identical and a basepair corresponding to 12 was removed from the sponge. The intentional mismatches and deleted basepair in the sponge sequence were designed to induce a bulge in the basepairing between the mature miRNA and the sponge sequence to prevent endonucleolytic cleavage such as those occurring from exact basepairing siRNAs. The sponge sequence is CTACTAAAACACCTAGCACTTA. This sequence was repeated seven times with random 8 bp sequences between each repeated sponge site. The 7X miR-302b sponge fragment was cloned downstream of GFP in the pSIN construct using MluI and NsiI restriction sites.

NIH 3T3 fibroblasts were infected with GFP-302-sponge-puro lentivirus supernatant with 4ug/mL polybrene. After 24h, media was replaced by MEF media. After 48h, cells were split to 40% confluency and puromycin (1µg/ml) was added to this and subsequent media changes. After 10 days, foci of puromycin resistant fibroblast colonies became visible. Cells were grown to high confluency and frozen for subsequent experiments. GFP-302-sponge stably expressing fibroblasts were plated at a confluency of 300,000 cells per 6-well dish in MEF media and puromycin (1µg/ml). The subsequent day, the cells were transfected with miRIDIAN miRNA mimics (Dharmacon) with Dharmafect1 (Dharmacon) at the manufacturer's recommended concentration of 100 nM. For 10 days following transfection, GFP

expression was assessed using FITC-Intensity measurement by flow cytometry (LSRII) and fluorescence microscopy. Cells were kept at constant confluency by 1:3 split every 24h.

Animal Use

All animal experiments described in this article were approved by the Institutional Animal Care and Use Committee of the University of California San Francisco.

Contributions

During the course of these studies, I have been extremely fortunate to be part of an incredible team of researchers, and have received invaluable intellectual contribution, advice, mentorship, and technical training from both within the Blelloch lab and from the greater UCSF community. I have acknowledged those individuals who provided the most pivotal support as a preface to this manuscript. In this section, I would like additionally identify and thank those researchers who directly performed experiments depicted and discussed in the previous chapters.

Joshua E Babiarz conducted the miR-290 promoter ChIP analysis (Fig. 10a&c).

Monica Venere helped significantly with the analysis of the miR-294+OSK chimeras (Fig. 9g).

Collin Melton and I worked very closely for all of the Let7 experiments (Fig. 12&13).

Tobias Greve was a student of mine who was an instrumental collaborator in several of the experiments discussed above. These include miR-294 and miR-181 target verification (Fig. 16b-d & Fig. 24), miR-181+OSK chimera generation (Fig. 20d-e), miRNA activity reporter assays (Fig. 21 c-d), and the miR-294 & miR-181 pathway regulation assays (Fig. 26 b,c&e-g).

Ronald Parchem both designed and developed the miRNA reporter construct (Fig.

21b) and provided significant aid with the miR-181+OSK chimera generation and analysis (Fig. 20d-e).

Samy Lamouille and **Deepa Subramanyam** were both long-term collaborators on many projects, including the EMT studies discussed here (Fig. 27a-c).

Jason Liu performed constructed and performed the luciferase assays confirming TGFBRII as an ESCC target (Fig. 27d)

It has been both an honor and a joy working with each of you. I hope that you learned as much from our projects and interactions as I have. Thank you for your dedication, quality work and friendship.

References

1. Iden, S. & Collard, J. G. Crosstalk between small GTPases and polarity proteins in cell polarization. *Nature reviews. Molecular cell biology* **9**, 846–59 (2008).
2. Walters, D. K. & Jelinek, D. F. A role for Janus kinases in crosstalk between ErbB3 and the interferon-alpha signaling complex in myeloma cells. *Oncogene* **23**, 1197–205 (2004).
3. Cancro, M. P. Signalling crosstalk in B cells: managing worth and need. *Nature reviews. Immunology* **9**, 657–61 (2009).
4. Tian, H., Mythreye, K., Golzio, C., Katsanis, N. & Blobe, G. C. Endoglin mediates fibronectin/α5β1 integrin and TGF-β pathway crosstalk in endothelial cells. *The EMBO journal* **31**, 3885–900 (2012).
5. Kim, D., Rath, O., Kolch, W. & Cho, K.-H. A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. *Oncogene* **26**, 4571–9 (2007).
6. Maston, G. a, Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* **7**, 29–59 (2006).
7. Smith, E. & Shilatifard, A. The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Molecular cell* **40**, 689–701 (2010).
8. O'Loghlen, A. *et al*. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell stem cell* **10**, 33–46 (2012).
9. Schwanhäusser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337–42 (2011).
10. Pucéat, M. Embryological origin of the endocardium and derived valve progenitor cells: From developmental biology to stem cell-based valve repair. *Biochimica et biophysica acta* (2012).doi:10.1016/j.bbamcr.2012.09.013
11. Riggs, J. W. *et al*. Induced pluripotency and oncogenic transformation are related processes. *Stem cells and development* **00**, (2012).

12. Estrov, Z. Stem cells and somatic cells: reprogramming and plasticity. *Clinical lymphoma & myeloma* **9 Suppl 3**, S319–28 (2009).

13. Hochedlinger, K. *et al*. Reprogramming of a melanoma genome by nuclear transplantation. *Genes & development* **18**, 1875–85 (2004).

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

15. Lujan, E., Chanda, S., Ahlenius, H., Sudhof, T. C. & Wernig, M. Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. *Proceedings of the National Academy of Sciences* **109**, 2527–2532 (2012).

16. Inagawa, K. & Ieda, M. Direct Reprogramming of Mouse Inagawa, K., & Ieda, M. (2012). Direct Reprogramming of Mouse Fibroblasts into Cardiac Myocytes. Journal of cardiovascular translational research. doi:10.1007/s12265-012-9412-5Fibroblasts into Cardiac Myocytes. *Journal of cardiovascular translational research* (2012).doi:10.1007/s12265-012-9412-5

17. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. a In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627–32 (2008).

18. Kalisky, T., Blainey, P. & Quake, S. R. Genomic analysis at the single-cell level. *Annual review of genetics* **45**, 431–45 (2011).

19. Miyanari, Y. & Torres-Padilla, M.-E. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* **483**, 470–3 (2012).

20. Nazor, K. L. *et al*. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell stem cell* **10**, 620–34 (2012).

21. Fabian, M. R. & Sonenberg, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature Structural Molecular Biology* **19**, 586–593 (2012).

22. Bartel, D. P. Review MicroRNAs : Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).

23. Farh, K. K.-H. *et al*. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).

24. Vogel, C. & Chothia, C. Protein Family Expansions and Biological Complexity. *PLoS Computational Biology* **2**, 13 (2006).

25. Erwin, D. H. *et al*. The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals. *Science* **334**, 1091–1097 (2011).

26. Peterson, K. J., Dietrich, M. R. & McPeek, M. A. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *BioEssays news and reviews in molecular cellular and developmental biology* **31**, 736–47 (2009).

27. Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. & Peterson, K. J. MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2946–2950 (2008).

28. Sempere, L. F., Cole, C. N., McPeek, M. A. & Peterson, K. J. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *Journal of experimental zoology Part B, Molecular and developmental evolution* **306**, 575–588 (2006).

29. Christodoulou, F. *et al*. Ancient animal microRNAs and the evolution of tissue identity. *Nature* **463**, 1084–1088 (2010).

30. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genetics* **39**, 380–385 (2007).

31. Suh, N. *et al*. MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Current biology : CB* **20**, 271–277 (2010).

32. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7634–7638 (1981).

33. Melton, C., Judson, R. L. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 126 (2010).

34. Motti, D., Bixby, J. L. & Lemmon, V. P. MicroRNAs and neuronal development. *Seminars in fetal & neonatal medicine* 6–11 (2012).doi:10.1016/j.siny.2012.07.008

35. Ivey, K. N. & Srivastava, D. MicroRNAs as regulators of differentiation and cell fate decisions. *Cell stem cell* **7**, 36–41 (2010).

36. Davalos, V. & Esteller, M. MicroRNAs and cancer epigenetics: a macrorevolution. *Current opinion in oncology* **22**, 35–45 (2010).

37. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell stem cell* **2**, 230–240 (2008).

38. Marson, A. *et al*. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).

39. Blelloch, R., Venere, M., Yen, J. & Ramalho-Santos, M. Generation of induced pluripotent stem cells in the absence of drug selection. *Cell stem cell* **1**, 245–247 (2007).

40. Aoi, T. *et al*. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science (New York, NY)* **321**, 699–702 (2008).

41. Hanna, J. *et al*. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).

42. Tsai, S.-Y. *et al*. Single Transcription Factor Reprogramming of Hair Follicle Dermal Papilla Cells to Induced Pluripotent Stem Cells. *Stem cells Dayton Ohio* **1**, 1–16 (2011).

43. Markoulaki, S. *et al*. Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nature Biotechnology* **27**, 169–171 (2009).

44. Nichols, J. *et al*. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).

45. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics* **24**, 372–376 (2000).

46. Boyer, L. A. *et al*. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).

47. Loh, Y.-H. *et al*. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics* **38**, 431–440 (2006).

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

49. Kuroda, T. *et al*. Octamer and Sox Elements Are Required for Transcriptional cis Regulation of Nanog Gene Expression. *Society* **25**, 2475–2485 (2005).

50. Heng, J.-C. D. *et al*. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell stem cell* **6**, 167–174 (2010).

51. Avilion, A. *et al*. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development* **17**, 126–40 (2003).

52. Ichida, J. K. *et al*. A Small-Molecule Inhibitor of Tgf-β Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. *Stem Cell* **5**, 491–503 (2009).

53. Maherali, N. & Hochedlinger, K. Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Current biology : CB* **19**, 1718–1723 (2009).

54. Jiang, J. *et al*. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature Cell Biology* **10**, 353–360 (2008).

55. Guo, G. *et al*. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development Cambridge England* **136**, 1063–1069 (2009).

56. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).

57. Feng, B. *et al*. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nature Cell Biology* **11**, 197–203 (2009).

58. Chen, J. *et al*. BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone. *Cell Research* **21**, 205–212 (2011).

59. Lyssiotis, C. A. *et al*. Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 8912–8917 (2009).

60. Huangfu, D. *et al*. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature biotechnology* **26**, 1269–1275 (2008).

61. Dang, C. V. MYC on the Path to Cancer. *Cell* **149**, 22–35 (2012).

62. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).

63. Nakagawa, M. *et al*. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology* **26**, 101–106 (2008).

64. Yu, J. *et al*. Induced pluripotent stem cell lines derived from human somatic cells. *Science (New York, NY)* **318**, 1917–1920 (2007).

65. Tsubooka, N. *et al*. Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. *Genes to cells devoted to molecular cellular mechanisms* **14**, 683–94 (2009).

66. Huangfu, D. *et al*. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnology* **26**, 795–7 (2008).

67. Kawamura, T. *et al*. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**, 1140–1144 (2009).

68. Vazquez-martin, A. *et al*. Activation of AMP-activated protein kinase (AMPK) provides a metabolic barrier to reprogramming somatic cells into stem cells © 2012 Landes Bioscience . © 2012 Landes Bioscience . *Cell cycle Georgetown Tex* **11**, 974–989 (2012).

69. Shi, Y. *et al*. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell stem cell* **2**, 525–8 (2008).

70. Samavarchi-Tehrani, P. *et al*. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell stem cell* **7**, 64–77 (2010).

71. Li, R. *et al*. A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. *Stem Cell* **7**, 51–63 (2010).

72. Menendez, S., Camus, S. & Belmonte, J. C. I. p53: Guardian of reprogramming. *Cell Cycle* **9**, 3887–3891 (2010).

73. Mikkelsen, T. *et al*. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).

74. Hanna, J. *et al*. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).

75. Chin, M. H. *et al*. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell stem cell* **5**, 111–123 (2009).

76. Prigione, A., Fauler, B., Lurz, R., Lehrach, H. & Adjaye, J. The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. *Stem Cells* **28**, 721–733 (2010).

77. Marson, A. *et al*. Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell stem cell* **3**, 132–5 (2008).

78. Anton, R., Kestler, H. A. & Kühl, M. Beta-catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. *FEBS Letters* **581**, 5247–5254 (2007).

79. Miki, T., Yasuda, S.-Y. & Kahn, M. Wnt/β-catenin Signaling in Embryonic Stem Cell Self-renewal and Somatic Cell Reprogramming. *Stem Cell Reviews* **7**, 836–46 (2011).

80. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine* **10**, 55–63 (2004).

81. Ying, Q.-L. *et al*. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–23 (2008).

82. Lluis, F., Pedone, E., Pepe, S. & Cosma, M. P. Periodic activation of Wnt/beta-catenin signaling enhances somatic cell reprogramming mediated by cell fusion. *Cell stem cell* **3**, 493–507 (2008).

83. Xu, R.-H. *et al*. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell stem cell* **3**, 196–206 (2008).

84. Shin, M., Alev, C., Wu, Y., Nagai, H. & Sheng, G. Activin/TGF-beta signaling regulates Nanog expression in the epiblast during gastrulation. *Mechanisms of Development* **128**, 268–278 (2011).

85. Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).

86. Sun, H. *et al*. PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6199–6204 (1999).

87. Watanabe, S. *et al*. Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene* **25**, 2697–2707 (2006).

88. Liu, N. *et al*. Exogenous Nanog alleviates but is insufficient to reverse embryonic stem cells differentiation induced by PI3K signaling inhibition. *Journal of Cellular Biochemistry* **106**, 1041–1047 (2009).

89. Nakamura, T. *et al*. Effects of Akt signaling on nuclear reprogramming. *Genes to cells devoted to molecular cellular mechanisms* **13**, 1269–1277 (2008).

90. Smith, A. G. *et al*. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).

91. Williams, R. L. *et al*. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687 (1988).

92. Matsuda, T. *et al*. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *the The European Molecular Biology Organization Journal* **18**, 4261–4269 (1999).

93. Smith, A. G. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17**, 435–462 (2001).

94. Yang, J. *et al*. Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell stem cell* **7**, 319–328 (2010).

95. Silva, J. *et al*. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS biology* **6**, e253 (2008).

96. Viswanathan, S. R. & Daley, G. Q. Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445–449 (2010).

97. Melton, C., Judson, R. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621–626 (2010).

98. Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature Biotechnology* **27**, 459–461 (2009).

99. Subramanyam, D. *et al*. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature Biotechnology* **29**, 443–448 (2011).

100. Anokye-Danso, F. *et al*. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell stem cell* **8**, 376–388 (2011).

101. Pfaff, N. *et al*. miRNA screening reveals a new miRNA family stimulating iPS cell generation via regulation of Meox2. *EMBO Reports* **12**, 1154–1160 (2011).

102. Li, Z., Yang, C., Nakashima, K. & Rana, T. M. Small RNA-mediated regulation of iPS cell generation. *The EMBO Journal* (2011).

103. Yang, C., Li, Z. & Rana, T. M. microRNAs modulate iPS cell generation microRNAs modulate iPS cell generation. *Spring* **17**, 1451–1460 (2011).

104. Miyoshi, N. *et al*. Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell stem cell* **8**, 633–638 (2011).

105. Choi, Y. J. *et al*. miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nature Cell Biology* **13**, 1353–1360 (2011).

106. Jayawardena, T. M. *et al*. MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes. *Circulation research* **110**, 1465–73 (2012).

107. Houbaviy, H. B., Murray, M. F. & Sharp, P. A. Embryonic stem cell-specific MicroRNAs. *Developmental cell* **5**, 351–358 (2003).

108. Wang, Y. *et al*. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nature Genetics* **40**, 1478–1483 (2008).

109. Chen, X. *et al*. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).

110. Mikkelsen, T. S. *et al*. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

111. Brambrink, T. *et al*. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell stem cell* **2**, 151–159 (2008).

112. Melton, C., Judson, R. L. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621–626 (2010).

113. Selbach, M. *et al*. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).

114. Baek, D. *et al*. The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).

115. Leung, A. K. L. *et al*. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nature Structural & Molecular Biology* **18**, 237–244 (2011).

116. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).

117. Orom, U. A. & Lund, A. H. Isolation of microRNA targets using biotinylated synthetic microRNAs. *Methods San Diego Calif* **43**, 162–165 (2007).

118. Leung, A. K. *et al*. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nature structural & molecular biology* (2011).

119. Zheng, G. X. *et al*. A latent pro-survival function for the mir-290-295 cluster in mouse embryonic stem cells. *PLoS genetics* **7**, e1002054 (2011).

120. Poliseno, L. *et al*. Identification of the miR-106b~25 microRNA cluster as a proto-oncogenic PTEN-targeting intron that cooperates with its host gene MCM7 in transformation. *Science signaling* **3**, ra29 (2010).

121. Zhang, X. D. *et al*. The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *Journal of biomolecular screening the official journal of the Society for Biomolecular Screening* **12**, 497–509 (2007).

122. O'loghlen, A. *et al*. MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation. *Stem Cell* **10**, 33–46 (2012).

123. Cichocki, F. *et al*. Cutting Edge: MicroRNA-181 Promotes Human NK Cell Development by Regulating Notch Signaling. *The Journal of Immunology* **187**, 6171–6175 (2011).

124. Wang, B. *et al*. TGFbeta-mediated upregulation of hepatic miR-181b promotes hepatocarcinogenesis by targeting TIMP3. *Oncogene* **29**, 1787–1797 (2010).

125. Wang, Y. *et al*. Transforming growth factor-β regulates the sphere-initiating stem cell-like feature in breast cancer through miRNA-181 and ATM. *Oncogene* (2010).

126. Ji, J. *et al*. Identification of microRNA-181 by genome-wide screening as a critical player in EpCAM-positive hepatic cancer stem cells. *Hepatology (Baltimore, Md)* **50**, 472–480 (2009).

127. Redmer, T. *et al*. E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Reports* **12**, 720–726 (2011).

128. Mirza, A. M. *et al*. Cooperative regulation of the cell division cycle by the protein kinases RAF and AKT. *Molecular and cellular biology* **24**, 10868–10881 (2004).

129. Kohn, A. D. *et al*. Construction and characterization of a conditionally active version of the serine/threonine kinase Akt. *The Journal of biological chemistry* **273**, 11937–11943 (1998).

130. Xu, J., Lamouille, S. & Derynck, R. TGF-beta-induced epithelial to mesenchymal transition. *Cell Research* **19**, 156–172 (2009).

131. Ebert, M. S. & Sharp, P. A. Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell* **149**, 515–524 (2012).

132. Abe, M. & Sato, Y. Puromycin insensitive leucyl-specific aminopeptidase (PILSAP) is required for the development of vascular as well as hematopoietic system in embryoid bodies. *Genes to cells devoted to molecular cellular mechanisms* **11**, 719–729 (2006).

133. Fei, T. *et al*. Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. *Genome Research* **20**, 36–44 (2010).

134. Witke, W., Sutherland, J. D., Sharpe, A., Arai, M. & Kwiatkowski, D. J. Profilin I is essential for cell survival and cell division in early mouse development. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3832–3836 (2001).

135. Chen, X. *et al*. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).

136. Fejes, A. P. *et al*. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**, 1729–1730 (2008).

137. Shi, R. & Chiang, V. Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* **39**, 519–525 (2005).

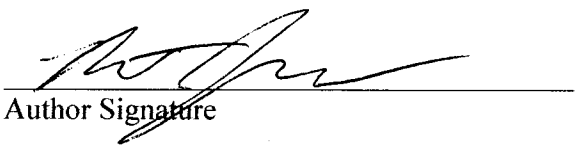
138. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish Prickle , a Modulator of Noncanonical Wnt / Fz Signaling , Regulates Gastrulation Movements. *Current Biology* **13**, 680–685 (2003).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

12/16/12

Date